

综 述

新型靶向基因组编辑技术研究进展*

杨发誉 葛香连 谷 峰**

(温州医科大学眼视光学院 眼视光学和视觉科学国家重点实验室培育基地 卫生部视觉科学研究重点实验室 温州 325000)

摘要 传统的靶向基因组编辑技术改造基因效率非常低,严重制约了基础研究和临床应用。因此,新的靶向基因组编辑工具的研究显得非常重要,以此来提高基因原位修复、定点整合及高通量基因敲除的效率。主要论述了近年来发现的新型靶向基因组编辑技术即锌指核酸酶(ZFN)、转录激活子样效应因子核酸酶(TALENs)、规律成簇间隔短回文重复(CRISPR)/Cas系统。从它们的发现、结构和研究进展及应用前景等方面进行了总结;通过比较三者的优缺点,发现规律成簇间隔短回文重复(CRISPRs)具有明显的优点。

关键词 靶向基因编辑 锌指核酸酶(ZFN) 转录激活子样效应因子核酸酶(TALENs) 规律成簇间隔短回文重复(CRISPRs)

中图分类号 Q753

靶向基因组编辑技术是指通过定点改造基因组DNA来改变基因组中的特定基因,从而定点修复致病基因(gene correction)^[1],同样的策略可以实现定点(*in situ*)插入一个基因或DNA元件(knock-in)^[2-3],同时该技术也可将正常基因进行基因敲除(gene knock-out)来研究其在体内的功能。早期的技术主要依赖于基因同源重组(基因打靶, gene targeting)及体细胞核移植技术来完成对特定基因的改造及个体的发育。在这一时期的重大进展包括: Recombineering^[4], single-stranded oligonucleotides^[5], AAV-mediated gene target^[6]的突破,是靶向技术上的一大变革,但这些方法总体而言存在效率低,技术要求高而且成本高的缺点,严重制约了基础研究和临床应用。近年来,新的靶向基因组编辑工具不断被发现,其中包括锌指核酸酶(zinc finger nucleases, ZFN)、转录激活子样效应因子核酸酶[transcription activator-like (TAL) effector nucleases, TALENs]和最新发现的规律成簇间隔短回文重复

(clustered regularly interspaced short palindromic repeats, CRISPRs)/Cas系统。这些新技术的应用,大大提高了基因定点编辑效率,使高通量的靶向基因编辑成为可能,这些为靶向基因组编辑提供了更为广阔的平台。

1 新型靶向基因编辑技术简介

1.1 锌指核酸酶(ZFN)

早在1983年锌指结构首次在研究非洲爪蟾转录因子中被发现,它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^[7]。而锌指核酸酶(ZFN)是人工设计的并具有锌指结构的蛋白质,由一个DNA识别结构域和一个非特异性核酸内切酶的剪切结构域融合而成。其N末端为锌指蛋白DNA结合域,由一系列Cys₂-His₂锌指蛋白串联组成,每个锌指蛋白识别并结合一个特异的三联体碱基。C末端为非特异性核酸酶Fok I剪切结构域, Fok I是海床黄杆菌(*Flavobacterium okeanokoites*)表达的一种限制性内切酶,只有当两个单体形成二聚体时才具有活性^[8]。

目前锌指核酸酶技术已经成功应用于很多物种的遗传研究,例如拟南芥、烟草、线虫、斑马鱼、非洲爪蟾、果蝇、大鼠、小鼠、猪、人、黑麦金斑蝶^[9]等。通过使用

收稿日期:2013-10-18 修回日期:2013-12-24

* 国家科技部“973”计划(2013CB967502)、国家自然科学基金(81201181/H1818)、浙江省卫生厅省部共建项目(201339279)资助项目

**通讯作者,电子邮箱:gufengw@gmail.com

ZFN 来编辑非模式生物——黑麦金斑蝶的基因组,这为剖析其多种生物过程的分子机制奠定了基础^[9]。

1.2 转录激活子样效应因子核酸酶(TALEN)

TALE 蛋白家族来自一类特殊的植物病原体,即黄单胞杆菌(*Xanthomonas* sp.),是一种天然蛋白质。它是由黄单胞杆菌通过 III 型分泌系统注入到宿主细胞内的一类蛋白效应子^[10]。早在 1989 年,Bonas 等^[11]就从辣椒斑点病细菌 *Xanthomonas campestris* pv. *vesicatoria* (xcv) 中发现了 TALE 蛋白家族的第一个成员 AvrBs3。TALE 蛋白包括三个组成部分,第一个组成部分来自于天然 TALE 蛋白,其 N 端含有 III 型分泌系统所需的分泌和易位信号,即易位结构域;第二部分是 DNA 结合结构域,是一段由 1.5 ~ 33.5 不等的 TALE 单元组成的重复氨基酸序列,每个单元又由 34 个氨基酸组成,其中 32 个氨基酸是高度保守的,只有第 12 位和 13 位氨基酸是可以变化的,能够特异地结合一个碱基序列,因此这两个氨基酸又被称为重复变异双残基(repeat variant diresidue, RVD),最后的 0.5 个单元只含有前面的 20 个氨基酸;第三部分位于天然 TALE 蛋白的 C 端,含有一个核定位信号(nuclear localization signals, NLS)和转录激活域(transcriptional activation domain),此部分能帮助 TALE 蛋白从细胞质进入细胞核同时发挥转录激活作用^[11-13]。此后科学家发现,TALE 蛋白的核酸结合域的氨基酸序列与其靶位点的核酸序列有恒定的对应关系。因此,利用 TALE 的序列模块,可组装成特异结合任意 DNA 序列的模块化蛋白,仿照 ZFN 的模式,把 TALE 中的转录激活结构域(AD)替换成核酸内切酶的切割结构域,构建成 TALE 核酸酶,对基因组的特定靶位点进行定向切割,从而达到靶向编辑内源性基因的目的^[14]。

作为一种新兴的基因组定点编辑工具,TALEN 技术正受到越来越多的关注。自 2009 年起,基于 TALE 结构的人工核酸内切酶已经成功地应用于包括芽殖酵母、果蝇、斑马鱼、线虫、小鼠^[15]、水稻、蟋蟀、家蚕、非洲爪蟾、猪、牛和拟南芥在内的多个物种的遗传研究,以及体外培养的哺乳动物细胞,包括胚胎干细胞(ES)^[16]和诱导性多能干细胞(iPS)^[17]。值得一提的是,在哺乳动物的遗传操作中,Y 染色体一直被认为是操作的禁区,没有成功改动过,Wang 等^[18]成功利用 TALEN 技术做到了这一点。随着研究的不断推进,TALE 核酸酶(即 TALEN)已成为 TALE 应用中最引人注目的研究。

1.3 规律成簇间隔短回文重复(CRISPR)/Cas 核酸酶

CRISPR 重复结构最早于 1987 年在大肠杆菌(*Escherichia coli*) K12 的 iap 基因侧翼序列中被发现^[19]。2000 年,Mojica 等^[20]提出把 CRISPR 作为重复序列家族的一类成员,并将其命名为 Short Regularly Spaced Repeats (SRSRs)。2002 年,Jansen 等^[21]最终将该类重复序列命名为 Clustered regularly interspaced short palindromic repeats (CRISPR),以反映该类重复序列的结构特点。CRISPR 基因结构的主体通常由高度保守的同向重复序列(repeat)与间隔序列(spacer)构成的多段 R-S 结构组成,称为基因座(CRISPR locus)。其中重复序列的长度一般为 21 ~ 48 bp,由于具有回文序列,可以形成发卡结构;重复序列之间的间隔序列一般为 26 ~ 72 bp,而间隔序列长度与细菌种类和 CRISPR 位点有关。在基因座之前,是前导序列(leader)与一系列的 CRISPR 相关蛋白基因座(CRISPR-associated),即 Cas 蛋白基因。总之,由 Cas 蛋白基因、前导序列和基因座共同构成了 CRISPR 基因序列。CRISPR 基因座与其相关基因能够针对噬菌体感染质粒接合和转化所造成的基因导入而形成特异性的防御机制,为原核生物提供对抗噬菌体等外源基因的获得性免疫能力,被称为 CRISPR 干扰(CRISPR interfering, CRISPRi)^[22-24],这种结构的作用机理可能与真核生物的 RNA 干扰过程类似^[25]。

2013 年初,来自麻省理工学院和哈佛大学的研究团队利用产脓链球菌(*Streptococcus pyogenes*)和嗜热链球菌(*Streptococcus thermophiles*)中的 CRISPR 酶和 RNA,在小鼠和人类细胞的 DNA 上进行了插入、切割、修复和编辑^[26]。这一突破性的研究,谱写了新型靶向基因组编辑技术新篇章。研究者认为 CRISPR/Cas9 系统由一个可诱导的 Cas9 蛋白和一个人工设计的 sgRNA 发夹复合体组成。后者由三个结构域组成,分别是一个长 20 nt 的与 DNA 特异性结合的互补区域、一个长 42 nt 的与 Cas9 蛋白结合的发夹结构域和一个长 40 nt 来源于脓链球菌的转录终止子。在 CRISPR/Cas9 系统的设计中,要特别注意 PAM 的原则问题,人工设计的 sgRNA 链对应的 DNA 链的一边需为 NGG 序列。有催化活性的 Cas9 蛋白在 NGG 前完成切割,这为其介导的基因组编辑提供了基础^[27](图 1)。为了使 CRISPR/Cas9 系统起到基因调控的功能而非基因编辑功能的目的,研究者使用无催化活性的 Cas9 蛋白,用于 RNA 指导的转录调控。任何基因的转录都能通过改造的 Cas9

基于最新研究成果 CRISPR/Cas9 技术比锌指核酸酶 (ZFNs) 和转录激活因子样效应物核酸酶 (TALENs), 具有更大的优势, CRISPR 更易于操作, 也具有更强的扩展性。CRISPR 还可以同时针对同一细胞中的多个位点进行研究^[27], Rudolf 等^[30] 利用 CRISPR/Cas9 技术一次性构建多突变转基因鼠, 这对于 ZFNs 和 TALENs 两项技术而言是难以实现的。这种功能在针对多基因关联突变而引起的疾病研究中将发挥重要作用, 这类疾病通常是由于多个基因发生突变后共同作用而引发的。为了进一步比较 TALENs 和 CRISPRs 的相对作用效率, 研究人员利用同一平台验证这两种方法, 对同一 hPSC 细胞系的同一基因组位点进行了分析, 结果表明 CRISPR/Cas9 方法的效率更高。TALENs 对一个等位基因产生克隆的效率是 0% ~ 34%, 而 CRISPR/Cas9 方法的效率则为 51% ~ 79%, 并且后者还能较容易地生成纯合子突变克隆 (总克隆的 7% ~ 25%)。此外这种方法也比 TALENs 方法敲入克隆的比例高出近十倍^[35]。研究人员推测 CRISPR/Cas9 的这种突出性能是由 Cas9 蛋白能更高量表达所致, 在 hPSCs 中的耐受性也比 TALENs 更好。

此外, 在目前的模式动物生产技术上, 胚胎干细胞是一个绕不开的过程。而 CRISPR/Cas9 技术则无需胚胎干细胞, 因此遗传学研究可能将不再局限于有限数量的模式生物了, 这打破了模式生物的定义。现在, 任何可以进行胚胎操作的动物都可以成为基因组工程的研究目标。由于许多动物的基因组已经完成测序, 因此, 利用这一技术我们可以在更多的物种中进行高效的遗传操作。但是 CRISPR/Cas9 也存在一些缺点, 其中尤其值得注意的是三点: 首先 G(N)19NGG 的要求有时会出现限制性; 其次 CRISPR/Cas9 方法的脱靶率还有待进一步确定^[36]; 最后 CRISPR/Cas9 方法会在预期靶点以外的位点上生成多余的 DNA 突变^[29]。

通过对三种新型靶向基因编辑技术的比较, CRISPR/Cas9 技术的优越性是显而易见的, 它的操作更简便, 应用范围更广, 靶向效率更高, 而且可以在同一细胞中同时进行多个靶位点的操作。更有可能打破传统模式生物的定义, 这将使遗传学研究不再局限于有限数量的模式生物了。此外, CRISPR/Cas9 技术还可以进行基因沉默等方面的操作 (比如使 Cas9 蛋白失活), 或者赋予 Cas9 蛋白更多新的功能 (比如使其具有转录因子样的转录活性等)。

3 新型靶向基因编辑技术应用前景

目前, 锌指核酸酶 (ZFN) 已成功应用于源自病人的诱导多能干细胞 (iPS), 来遗传修复帕金森综合症相关的突变基因 SNCA^[37]。而类转录激活因子效应物核酸酶 (TALEN) 自 2010 年底开始成功应用于基因打靶, 很快成为一种比锌指核酸酶 (ZFN) 更容易设计、特异性更高和毒性更低的人工核酸内切酶。虽然所有的位点特异性核酸酶在当前的应用都被限制在体细胞中, 但在干细胞中的研究正在不断进步, 包括诱导多能干细胞 (iPS) 的生产和操作, 这些进展将最终开辟出无数基因治疗的新方向, 包括源于自体干细胞移植的疗法^[38]。

CRISPR/Cas9 技术自 2013 年初开始成功应用以来, 短短一年时间已应用于许多生物的研究, 其在基因编辑中的优越性显而易见。但要实现 CRISPR/Cas9 技术的广泛应用, 在未来需要进一步改进和完善, 重点应该是解决脱靶和引起多余突变的问题, 建立完善的 CRISPR/Cas9 信息平台, 使其更有利于在哺乳动物细胞中发挥作用。可以预见, ZFN/TALEN 和 CRISPR/Cas9 技术的发展将大大加快基因操作相关领域的研究步伐, 尤其是 CRISPR/Cas9 技术的发展及应用, 这些将为基因治疗和基因功能分析提供强大的工具。目前科学家已找到了影响 CRISPR-Cas9 系统精确度的关键因素^[39], 近期他们又发现采用 Cas9 切口酶 (Cas9 nickase) 来诱导双链断裂能提高靶向特异性^[28]; 利用 CRISPR/Cas9 技术在小鼠和人类干细胞中纠正遗传缺陷, 有效地治疗疾病^[31-32]; 通过构建 lentiCRISPR, 研究人员进行了靶向全基因组范围 CRISPR - Cas9 敲除 (GeCKO), 可以在人类细胞中实现正向和负向的选择性筛选基因^[40]。这些发现将使该系统在人类细胞中的应用更安全, 更方便为各种人类疾病构建出动物模型, 也为改写遗传缺陷, 有效地治疗疾病开辟新篇章。

致谢 本项目受国家科技部 973 项目 (2013CB967502), 国家自然科学基金 (81201181/H1818), 浙江省卫生厅省部共建项目 (201339279), 温州医科大学人才启动项目 (QTJ 12011) 资助, 特此致谢。

参考文献

- [1] Chamberlain J R, Schwarze U, Russell D W, et al. Gene targeting in stem cells from individuals with osteogenesis

- imperfecta. Science, 2004, 303(5661): 1198-1201.
- [2] Johnson L, Mercer K, Jacks T, et al. Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. Nature, 2001, 410(6832): 1111-1116.
- [3] Beard C, Hochedlinger K, Jaenisch R, et al. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. Genesis, 2006, 44(1): 23-28.
- [4] Pentao Liu, Nancy A J, Neal G C. A highly efficient recombineering-based method for generating conditional knockout mutations. Genome Res, 2003, 13(3): 476-484.
- [5] Ellis H M, Yu D, Court D L, et al. High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. Proc Natl Acad Sci U S A, 2001, 98(12): 6742-6746.
- [6] Miller D G, Wang P R, Russell D W, et al. Gene targeting *in vivo* by adeno-associated virus vectors. Nat Biotechnol, 2006, 24(8): 1022-1026.
- [7] Miller J, McLaehlan A D, Klug A, et al. Repetitive zinc binding domains in the protein transcription factor III A from *Xenopus oocytes*. EMBO J, 1985, 4(6): 1609-1614.
- [8] Kim Y G, Chandrasegaran S. Chimeric restriction endonuclease. PNAS, 1994, 91(3): 883-887.
- [9] Christine Merlin, Lauren E Beaver, Orley R Taylor, et al. Efficient targeted mutagenesis in the monarch butterfly using zinc-finger nucleases. Genome Research, 2013, 23: 159-168.
- [10] Boch J, Bonas U. *Xanthomonas* AvrBs3 family-type III effectors: discovery and function. Annu Rev Phytopathol, 2010, 48: 419-436.
- [11] Bonas U, Stall R E, Staskawicz B, et al. Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. Mol Gen Genet, 1989, 218(1): 127-136.
- [12] Kay S, Hahn S, Marois E, et al. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. Science, 2007, 318(5850): 648-651.
- [13] Moscou M J, Bogdanove A J. A simple cipher governs DNA recognition by TAL effectors. Science, 2009, 326(5959): 1501.
- [14] Cong L, Zhou R H, Kuo Y C, et al. Comprehensive interrogation of natural TALE DNA-binding modules and transcriptional repressor domains. Nat Commun, 2012, 3(7): 968.
- [15] Wang H, Hu Y C, Markoulaki S, et al. TALEN-mediated editing of the mouse Y chromosome. Nat Biotechnol, 2013, 31(6): 530-532.
- [16] Lombardo A, Genovese P, Beausejour C M, et al. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. Nat Biotechnol, 2007, 25(11): 1298-1306.
- [17] Hockemeyer D, Wang H, Kiani S, et al. Genetic engineering of human pluripotent cells using TALE nucleases. Nat Biotechnol, 2011, 29(8): 731-734.
- [18] Wang H, Hu Y C, Jaenisch R. TALEN-mediated editing of the mouse Y chromosome. Nat Biotechnol, 2013, 31(6): 530-532.
- [19] 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(12): 5429-5433.
- [20] Mojica F J, Diez-Villasenor C, Soria E, et al. Biological significance of a family of regularly spaced repeats in the genomes of Archaea, bacteria and mitochondria. Mol Microbiol, 2000, 36(1): 244-246.
- [21] Jansen R, Embden J D, Gaastra W, et al. Identification of genes that are associated with DNA repeats in prokaryotes. Mol Microbiol, 2002, 43(6): 1565-1575.
- [22] Grissa I, Vergnaud G, Pourcel C, et al. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. BMC Bioinformatics, 2007, 8: 172.
- [23] Karginov F V, Hannon G J. The CRISPR system: small RNA-guided defense in bacteria and archaea. Mol Cell, 2010, 37(1): 7-19.
- [24] Deveau H, Garneau J E, Moineau S, et al. CRISPR/Cas system and its role in phage-bacteria interactions. Ann Rev Microbiol, 2010, 64: 475-493.
- [25] Marraffini L A, Sontheimer E J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet, 2010, 11(3): 181-190.
- [26] Cong L, Ran F A, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. Science, 2013, 339(6121): 819-823.
- [27] Lei S Q, Matthew H Larson, Luke A, et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell, 2013, 152: 1173-1183.
- [28] Ran F A, Hsu P D, Zhang F, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell, 2013, 154(6): 1380-1389.
- [29] 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(9): 822-826.
- [30] Rudolf Jaenisch, Haoyi Wang, Hui Yang, et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. Cell, 2013, 153: 1-9.
- [31] Wu Y, Liang D, Li J, et al. Correction of a genetic disease in mouse *via* use of CRISPR-Cas9. Cell Stem Cell, 2013, 13(6): 1-9.

- 659-662.
- [32] Schwank G, Koo B K, Clevers H, et al. Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell*, 2013, 13(6): 653-658.
- [33] Method of the Year 2011. *Nat Methods*, 2012, 9(1): 1.
- [34] Breakthrough of the year: The runners-up. *Science*, 2012, 338(6114): 1525-1532.
- [35] 张金脉, 任兆瑞. TALENs: 一种新的基因定点修饰技术. *生命科学*, 2013, 25(1): 54-59.
- Jinmai Zhang, Zhaorui Ren. TALENs: A new genome site - specific modification technology. *Chinese Bulletin of Life Sciences*, 2013, 25(1): 54-59.
- [36] Qirong Ding, Stephanie N Regan, Yulei Xia, et al. Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell*, 2013, 12(4): 393-394.
- [37] Soldner F, Laganère J, Cheng A W, et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset parkinson point mutations. *Cell*, 2011, 146(2): 318-331.
- [38] Gaj T, Gersbach C A, Barbas C F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*, 2013, 31(7): 397-405.
- [39] Hsu P D, Scott D A, Zhang F, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, 2013, 31(9): 827-832.
- [40] Shalem O, Sanjana N E, Zhang F, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, 2014, 343(6166): 84-87.

Progress of Next-generation Targeted Gene-editing Techniques

YANG Fa-yu GE Xiang-lian GU Feng

(School of Ophthalmology and Optometry, Wenzhou Medical University, State Key Laboratory Cultivation Base and Key Laboratory of Vision Science, Ministry of Health and Zhejiang Provincial Key Laboratory of Ophthalmology and Optometry, Wenzhou 325000, China)

Abstract Manipulating genomes by traditional targeted genome editing technique (gene targeting) is inefficient, making it impractical or difficult to use the technique as a gene-therapy approach to cure diseases and decipher gene functions. To overcome this problem, next-generation targeted gene-editing techniques were developed to achieve higher efficiency for gene correction, specific locus integration or knock-in and high throughput gene knock-out. The progress of new techniques for targeted genome-editing tools were reviewed, including zinc finger nucleases (ZFN), transcription activator-like effector nucleases (TALENs), and a clustered regularly interspaced short palindromic repeats (CRISPR)/Cas system. A brief summary of the history, recent structure, progress, and future prospects was presented. After comparing these tools, it was found that CRISPR systems offer an advantage over ZFN and TALEN.

Key words Targeted gene editing tool Zinc Finger nucleases(ZFN) Transcription activator-like effector nucleases (TALENs) Clustered regularly interspaced short palindromic repeats(CRISPRs)