

DOI: 10.13376/j.cblls/2018116

文章编号: 1004-0374(2018)09-0967-13



张博, 北京大学生命科学学院教授。多年来以斑马鱼为主要模式动物, 致力于应用多种功能基因组学技术, 研究脊椎动物早期胚胎发育与器官形成的基因表达调控机制, 建立人类疾病的斑马鱼模型并进行发病机制研究。同时, 还致力于开发并完善斑马鱼基因组靶向修饰技术, 先后创建了一系列以 TALEN 和 CRISPR/Cas 为基础的斑马鱼基因组定点突变技术。作为通讯作者在 *Nature Biotechnology*、*Nature Methods*、*PLoS Biology*、*PNAS* 等期刊发表学术论文七十多篇。

基因组靶向敲入技术在模式动物中的 发展与应用: 以斑马鱼为例

韩冰舟, 张亚鸽, 张 博*

(北京大学生命科学学院细胞增殖与分化教育部重点实验室, 北京大学基因组编辑研究中心, 北京 100871)

摘 要: 基因组靶向敲入是指将特定的外源核酸序列转入细胞或个体基因组中的特定位点, 以实现条件性基因敲除、单碱基或序列替换、细胞或基因标记等多种精确和(或)复杂的基因组靶向修饰。通过长同源臂介导的同源重组(HR)和微同源序列介导的同源修复两种途径可实现精确的靶向敲入, 非同源末端连接(NHEJ)途径则可介导非精确的靶向敲入, 或称靶向插入。一般而言, 基于同源序列的精确敲入的效率低于 NHEJ 诱导的靶向插入。受益于以 TALEN、CRISPR/Cas 系统为代表的人工核酸内切酶介导的基因组编辑技术的飞速发展, 基因组靶向敲入得以在多个物种中越来越广泛地应用, 并极大地推动了基因功能与疾病模型研究。不过, 在个体水平上, 相比于简单的 indel 突变, 基因组靶向敲入仍较难实现, 效率仍然偏低。现介绍基因组靶向敲入的基本原理, 并以斑马鱼为例, 介绍基因组靶向敲入在模式动物中的应用与相关技术的发展历程, 并着重强调实验设计与操作的要点, 同时对基因组靶向敲入技术的发展进行了展望。

关键词: 基因组编辑; 基因敲入; 模式动物; 斑马鱼; CRISPR/Cas 系统; 非同源末端连接; 同源重组
中图分类号: Q78; Q959.468 **文献标志码:** A

Development and applications of targeted knock-in techniques in model animals: zebrafish as an example

HAN Bing-Zhou, ZHANG Ya-Ge, ZHANG Bo*

(Key Laboratory of Cell Proliferation and Differentiation of Ministry of Education, Peking University Genome Editing Research Center, College of Life Sciences, Peking University, Beijing 100871, China)

收稿日期: 2018-07-24; 修回日期: 2018-08-02

基金项目: 国家重点研发计划干细胞及转化研究专项(2016YFA0100500); 国家重大科学研究计划项目(2015CB942803); 国家自然科学基金项目(31671500, 81371264); 北大生科-启东创新基金

*通信作者: E-mail: bzhang@pku.edu.cn

Abstract: Targeted knock-in (KI) refers to the experimental methods to introduce exogenous DNA sequence into a specific site to achieve genome modification such as conditional knock-out (KO), nucleotide or fragment substitution, cell or gene labeling, etc. Precise KI could be achieved via traditional long-homology arm-directed homologous recombination (HR) or micro-homology based DNA repair, while imprecise KI could be generated through non-homologous end joining (NHEJ)-mediated targeted insertion, which is generally higher in efficiency than the homology-based precise KI. In recent years, following the revolution of genome editing technology brought by the fast development of engineered endonucleases including TALEN, CRISPR/Cas system, KI strategy becomes more and more popular in many model organisms, which greatly facilitated the study of gene function and disease modeling. However, the efficiency of knock-in is still much lower than knock-out at individual level, and the technique practice is still challenging for most researchers. Here, using zebrafish as a major model, we briefly summarized the basic principle and progress of knock-in approaches, as well as experimental design and technical considerations. We also discussed the future directions of this important yet challenging technology.

Key words: genome editing; knock-in; model animal; zebrafish; CRISPR/Cas system; NHEJ; HR

近年来,人们巧妙地利用细菌的一些特殊的基因表达调控机制或免疫机制,以及多年来对真核生物转录因子作用机制的研究成果,逐步创建出若干种能够特异切割DNA靶序列的人工核酸内切酶(engineered endonuclease, EEN),为基因组靶向修饰技术带来了一场又一场的革命。目前,人工核酸内切酶主要包括锌指核酸酶(zinc finger nuclease, ZFN)、类转录激活因子效应物核酸酶(transcription activator-like effector nuclease, TALEN)、规律性成簇间隔的短回文重复序列(clustered regularly interspaced short palindromic repeats, CRISPR)及相关蛋白(CRISPR-associated, Cas)系统(CRISPR/Cas system)等3种类型^[1]。其中,ZFN是最先被发明的一种EEN,但由于其DNA结合能力不稳定、与靶序列的对应性较差^[2-3]以及总体效率低等原因^[4],目前已经很少有人使用。TALEN是基于植物病原菌编码的转录激活因子开发的一种特异且高效的EEN^[5-6],然而,由于TALE序列高度重复,并且不同的靶点需要不同的TALE单元组合,导致TALEN的构建和使用均比较复杂^[7],通用性低,也不便于改造,目前使用的频率也已不高。CRISPR/Cas系统于2013年开始用于靶向突变基因组,是借鉴原核生物免疫系统而开发的一种最新的EEN技术^[8-9]。该系统分为多种类型,其中最为人熟知和广泛使用的是由Cas9核酸酶及其对应的gRNA构成的Cas9/gRNA系统。由于CRISPR/Cas系统使用统一的核酸酶,其靶点识别的特异性通过gRNA的碱基配对实现^[10],针对不同的DNA靶序列只需构建新的gRNA,因此,该系统便于设计和使用,并且成本低廉,一经问世便成为最受欢迎的基因组编辑技术^[11-12]。

人工核酸内切酶的主要作用是特异性识别并结合DNA靶序列、对其或其邻近的序列进行切割,造成DNA双链断裂(double strand break, DSB),从而激活细胞自身的DNA损伤修复机制^[13],利用该机制的容错性改变靶基因的序列,最终达到基因敲除(knock-out, KO)的效果。也就是说,EEN只需要在特定的DNA序列上造成一个DSB,便足以诱导该位点产生突变。这是由于细胞通常会通过非同源末端连接(non-homologous end joining, NHEJ)的途径修复DNA双链断裂,而NHEJ是一种快捷但并不精确的修复方式,往往会在切割位点处造成若干个碱基的插入和(或)缺失(insertion/deletion, indel)突变;对于编码基因而言,如果靶点位于其编码区的外显子上,并且产生了非3的整倍数的indel,则会导致移码突变与蛋白质翻译的提前终止,有时还会触发NMD(nonsense-mediated decay)机制,诱导突变的mRNA产生降解,从而破坏靶基因的表达与功能。不过,在某些情况下,indel突变不足以完全破坏靶基因或其他功能性元件(如增强子)的功能,或者有时可能会引起所谓的“遗传补偿”(genetic compensation)效应,从而掩盖应有的表型或干扰表型分析。大片段基因组删除(genomic deletion)策略理论上可以更有效地造成靶基因失活,其原理是在靶基因/靶元件内部或者两侧分别设计相距较远的两个甚至多个靶位点,以便在同一条DNA分子上造成两个或多个DSB,这样就有机会使两个不同靶点的DSB断点之间通过NHEJ修复连接起来,从而丢掉两个靶点/DSB断点之间的序列,造成靶基因/靶元件的大部分甚至全部删除/缺失,彻底破坏靶基因/靶元件的结构与功能^[14-15]。广义而言,

基因组序列的大片段删除可视为基因敲除的一种形式。由于 indel 突变和大片段删除只涉及内源基因组 DNA 序列之间的简单连接, 并且 NHEJ 是大多数真核生物细胞中主导的 DSB 修复途径, 因而基因敲除的技术难度相对而言并不大, 属于 EEN 技术中最为简单的一类应用。

除了基因敲除, EEN 还可用于对基因组进行更为复杂或精细的靶向修饰, 如在靶基因中引入外源序列以实现条件性基因敲除或基因/细胞标记, 或对靶基因进行单碱基或序列替换等精确修饰。这些应用一般都需要提供一个外源的 DNA 作为供体 (donor) 或序列模板, 在 EEN 切割基因组形成 DSB 后, 通过同源重组 (homologous recombination, HR) 或微同源末端连接 (micro-homology mediated end joining, MMEJ) 介导的同源修复 (homology-directed repair, HDR) 途径, 或者较为高效的 NHEJ 途径, 将供体序列整合到靶点中, 从而达到在基因组中插入供体序列或利用供体序列替换基因组内源序列的效果^[16]。这类靶向修饰方式一般称为基因组靶向敲入 (knock-in, KI)、基因组定点敲入, 或者简单地称为基因敲入。通过长同源臂 (homology arm) 介导的同源重组、微同源序列介导的同源修复等两种途径可实现精确的靶向敲入, 最终的结果可以视为由供体同源臂之间的序列精确替换了基因组上同源臂之间的内源序列, 因此也可以称为靶向替换; 如果只需要插入外源序列, 而并不需要替换基因组上的任何序列, 则可以以待插入位点两侧的基因组序列作为同源臂设计供体, 通过同源修复途径将供体同源臂之间的序列精确地插入到基因组中。靶向插入 (targeted insertion) 也可以通过 NHEJ 途径实现, 其原理可以简单地视为基因组上的 DSB 断点两侧分别跟供体的一端发生了 NHEJ 连接。跟同源修复途径相比, NHEJ 介导的靶向插入具有如下几个特点: (1) 供体无需同源臂序列, 构建相对简单; (2) 所有的供体序列整体插入靶位点的 DSB 之处; (3) 是一种不精确的靶向敲入方式, 这是因为 NHEJ 是一种不精确的修复途径, 供体序列与靶点连接的同时, 往往会产生碱基的插入和 (或) 缺失; (4) 效率一般大大高于由同源序列介导的精确敲入。相比于基因敲除, 基因敲入由于依赖外源供体在靶点位置的整合, EEN 切割靶点后发生 KI 的效率通常远远低于 indel 突变效率, 并且会随着需要插入或替换序列的长度增加而进一步下降。因此, 基因敲入虽然为各种科研工作所需要, 但这项技术始终具有很高的技

术门槛, 成为开展基因功能与机制深入与精细研究的拦路虎。在不同的实验模式中, 简便、高效地实现各种形式的基因组靶向敲入一直是人们不断努力尝试解决但是仍未完全攻克的难题。本文以斑马鱼 (*Danio rerio*) 中基因组靶向敲入技术的研究进展为例, 简要介绍在模式动物中实现基因组靶向敲入的方式与原理, 简析这些技术能够达成的实验目的, 并简单展望这一技术未来的发展方向。

1 在模式动物中实现基因敲入的意义与挑战

1.1 基因敲入技术对于模式动物研究的意义

以 TALEN、CRISPR/Cas 系统为代表的 EEN 技术在制备 indel 等简单突变体方面的应用已经较为成熟, 然而, 想要达到一些复杂的实验目的, 如基因表达标记或特殊突变体的制备, 通常需要通过基因敲入对靶基因进行外源 DNA 序列的插入或替换。目前, 基因功能与分子机制研究越来越深入, 在个体水平解读基因组的整体功能与基因互作及表达调控网络也日趋复杂、精细, 简单的基因敲除在很多情况下已经无法满足人们复杂多样的研究需要, 而是需要通过基因敲入技术实现条件性基因敲除、对基因或细胞进行特异标记 (示踪)、对基因组进行碱基或序列替换等精确修饰等复杂应用^[1]。例如, 很多重要的发育调控基因往往具有复杂的时空特异性表达图谱, 在胚胎发育的不同时期和 (或) 多个不同的组织中发挥不同的作用, 全身性的基因敲除往往不足以全面、深入、细致地解读该基因的功能, 而条件性基因敲除则能够弥补这方面的不足。常用的条件性基因敲除策略是基于 Cre/loxP 重组系统, 需要将两个长度为 30 多个碱基对的 loxP 序列分别定点插入到靶基因一个或多个外显子的两侧 (内含子中)。又如, 在 TALEN^[17-18] 和 CRISPR/Cas 系统^[19-22] 等广泛应用于小鼠和斑马鱼之后, 人们制备了许多基因的 indel 突变, 然而, 其中有相当多的突变体并未观察到明显表型, 或是泛表达的基因只在某些组织器官中出现可见的表型。2015 年, Kok 等^[23] 发现, 对于多个斑马鱼基因, 使用吗啉环反义核酸类似物 morpholino (MO) 进行基因敲低得到的表型与该基因 indel 突变体的表型呈现出明显的差异, 突变体的表型往往弱于注射了 MO 的胚胎 (morphant), 有的甚至无法检测到可见表型。2015 年, Rossi 等^[24] 报道认为, 某些基因发生 indel 突变后会导致一些跟靶基因结构或功能类似的其他基因的表达上调, 从而产生基因补偿效应, 造成突变体不

出现应有的表型或表型减弱。此外，在小鼠中曾有多例报道，靶基因整体敲除后并未造成明显的异常表型，或仅导致比较轻微的表型，而组织特异性敲除这些基因后，反而出现了更为严重的发育异常^[25]。这些现象说明，只有利用基因敲入技术对基因组进行更为复杂和精确的修饰，进而实现对基因结构更加精细的改变和调整，对基因表达进行更加精准的追踪和调控，才能更为完整、准确地解读基因的功能。因此，建立并优化简单、高效的基因敲入技术很有必要。事实上，基因敲入技术不仅有助于基因功能与分子机制的深入与精细研究，而且在构建人类疾病的动物模型、基因治疗、农作物与家禽家畜水产动植物以及宠物的品种改良等领域也具有重要的应用前景。

1.2 在个体水平进行基因敲入的难点

在个体水平，如模式动物中，进行基因组编辑的难度往往高于体外培养细胞，进行基因敲入实验更是具有挑战性。这是由多种因素造成的，而本质上则是由单细胞克隆培养与多细胞个体发育之间的差异决定的，或者更确切地说，是由于基因组编辑诱导的突变事件发生在细胞水平，而最终的目的却是获取携带均一突变的多细胞个体，这两者之间的矛盾所造成的。细胞在体外培养时可以进行克隆繁殖，每个细胞理论上都可以跟其他细胞分开，通过单克隆培养，单独形成遗传背景均一的后代类群。而多细胞个体在正常情况下是由一个单细胞受精卵通过严格有序的细胞增殖与分化逐步发育而来，很难进行克隆繁殖。不论是类型单一的外体培养的单细胞克隆群体，还是由多种不同类型的细胞构成的多细胞有机体，EEN在诱导突变时都是以每个DNA靶点为单位进行的，突变在每个细胞内独立发生，细胞之间彼此互不干扰。因此，对于体外培养细胞而言，获取突变体的途径相对简单，只要得到靶基因被成功编辑的单个细胞，如果该基因突变不影响细胞的存活和增殖，即可通过简单的克隆培养和有丝分裂将该突变传递至子代细胞，稳定保留下来，从而达成实验目的。对于由多细胞构成的模式动物而言，由于需要获取每个细胞都带有相同突变类型、基因型均一的多细胞有机体，因而创建突变体的途径较为复杂。在进行模式动物的基因组编辑时，一般采用显微注射的方法向雌性生殖细胞或单细胞受精卵中注射基因打靶体系，这就造成靶基因突变跟胚胎发育相伴而生，细胞边分裂/分化边突变，每个细胞中靶基因是否被编辑或者被编辑的

情况有可能不同，从而形成基因型不同的嵌合体胚胎。由这样的胚胎形成的亲代个体称为奠基者(Founder)。这样的实验策略就带来了体外单克隆培养细胞无需面对的三个棘手的难题。第一个难题是需要通过种系传递(germline transmission)才能获得基因型一致、可稳定遗传的突变个体。嵌合体胚胎中，只有发生在生殖细胞/种系中的突变才有可能稳定地遗传给后代，因此，获得可稳定遗传的多细胞突变体的关键在于种系传递的效率。由于生殖细胞在胚胎中所占的比例很低，若靶基因的突变效率不够高，则生殖细胞的突变嵌合率就会比较低，从而对在子代中遴选突变携带者造成很大的困难，甚至会导致实验失败。第二个难题是获得突变体的实验周期较长。体外培养细胞通过有丝分裂进行克隆扩增，实验周期较短，而多细胞个体往往需要较长的发育与世代交替的时间，实验周期相对较长，斑马鱼、小鼠等脊椎动物的世代周期就更长。第三个难题是无法通过对嵌合体进行药物或遗传标记来筛选、富集所需的突变。体外培养细胞可快速、大量地克隆培养，并通过简便的化学试剂转染或电穿孔等方法将基因打靶体系引入细胞，然后利用抗性基因或流式细胞分选术结合荧光标记进行高通量筛选，富集打靶成功的细胞，然后通过单克隆培养挑选并迅速扩增所需的突变细胞。因此，对体外培养细胞进行基因打靶不仅实验周期较短，而且对靶点突变效率的要求也不高。而在模式动物实验中，由多个细胞构成的嵌合体Founder胚胎中往往含有不少未被成功编辑的细胞，只有一部分细胞的靶基因会产生突变，因而不能通过药物或者其他遗传标记筛选的方式富集打靶成功的胚胎，而只能在其子代中人工逐一遴选阳性实验结果，因此，整个实验流程的操作难度较大，时间、人力和物力成本均较高；更重要的是，靶点的突变效率对实验的成功率具有决定性的影响，这一点在基因敲入实验中体现得尤为明显。此外，一些模式动物在胚胎发育早期细胞分裂的速度较快，染色质经常处于凝缩状态，不利于EEN的识别、结合以及切割，也不利于外源序列的整合。因此，在模式动物中实现高效、便捷、普适、稳定的基因组靶向敲入，一方面具有多种重要的应用需求，另一方面则仍然是一项十分具有挑战性的工作。

2 基于同源修复途径的基因组靶向敲入策略

2.1 利用传统长同源臂实现基因敲入

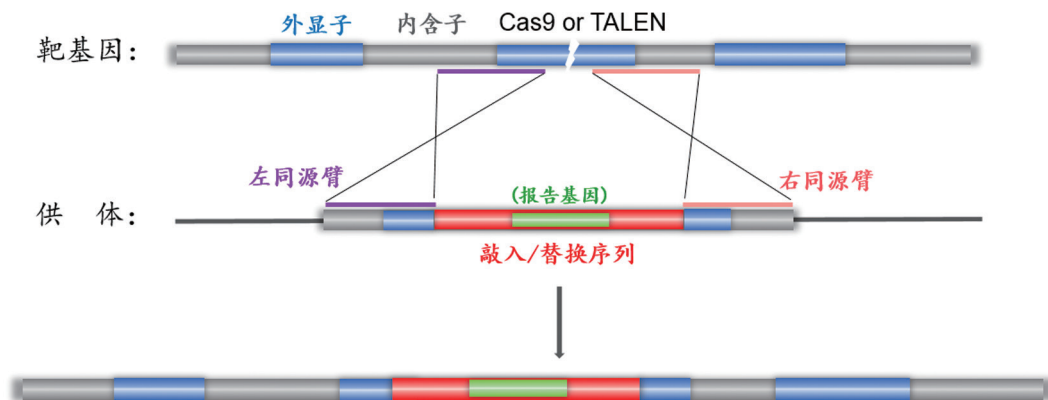
构建条件性基因敲除和基因标记的经典方法主

要是利用同源重组介导的基因敲入将外源 DNA 序列精确引入靶基因中。模式动物中最早成功应用的基因组定点修饰技术便是利用了基于同源重组的 DNA 损伤修复途径, 20 世纪 80 年代诞生于对小鼠的研究中。将两侧各带有一段和靶位点附近的序列一致的同源臂的外源供体 DNA 转入小鼠胚胎干细胞, 细胞自身的染色体能够以很低的效率 (约 10^{-6} 量级) 与供体 DNA 发生同组, 从而将外源序列精确置换到内源基因组中^[26]。由于自然情况下发生同源重组的效率极低, 因此, 往往需要在供体 DNA 上添加正、负筛选标记并优化条件, 排除掉未发生重组或供体随机整合进入基因组的细胞, 进行大量筛选, 才有可能在众多细胞株中找到成功发生同源重组的细胞。然后, 将筛选到的胚胎干细胞移植入小鼠囊胚中, 如果这些细胞发育成生殖细胞, 就有可能将突变遗传至下一代, 从而完成基因编辑小鼠的构建。如果不经过上述胚胎干细胞途径, 而是直接向早期胚胎中注射供体序列, 则由于自发同源重组的效率太低, 只有极少数细胞能被成功编辑, 种系传递效率会非常低; 同时, 由于注射后的胚胎均为同源重组突变的嵌合体, 无法对整个胚胎进行药物筛选, 理论上只能通过繁养大量的 Founder 及其后代对突变体进行遴选, 而这样的实验成本是不可承受的。因此, 长期以来只有小鼠能够通过体外培养胚胎干细胞筛选得到带有突变的细胞, 再通过囊胚移植获得同源重组诱导靶向突变的个体。所以, 这一基于胚胎干细胞的靶向敲入技术迄今仅限于应用于小鼠, 很难扩展到其他模式生物中。有研究表明, 靶位点发生 DNA 双链断裂能够提高同源重组

的效率。人工核酸内切酶技术出现后, 人们便开始尝试利用它来提高同源重组的效率, 希望藉此能够无需通过胚胎干细胞和药物筛选, 直接在个体水平实现对基因组进行精确修饰并稳定遗传, 也就是说, 直接在生物体内对生殖细胞进行精确编辑。Urnov 等^[27]报道, EEN 能将靶位点的同源重组效率提高 3~4 个数量级, 极大地方便了筛选工作。目前, 在斑马鱼等很多传统上无法进行同源重组介导的基因组编辑的物种中, 借助于 TALEN 或 CRISPR/Cas 系统, 结合带有同源臂的供体 DNA, 有不少都顺利实现了外源序列的精确整合或替换。

同源重组介导的 DNA 修复需要提供两侧含有同源序列的外源 DNA 供体, 最终使两个同源臂之间的序列经重组进入内源基因组中 (图 1: 相当于“替代”了基因组上两个同源臂之间的内源序列, 或者说供体上两个同源臂之间的序列跟内源基因组中两个同源臂之间的序列发生了互换)。可以通过设计具有不同序列与功能的供体, 精确地操控和编辑基因组, 借此实现精确的点突变、基因修复、基因敲入、基因替换和条件性等位基因构建等, 从而在疾病治疗、基因标记、特殊突变体构建等研究中发挥重要作用。

利用 EEN 结合 HR 介导的基于传统长同源臂的基因敲入实验方案类似于前述小鼠胚胎干细胞的方法, 需要克隆靶点两侧的同源臂, 同时, 根据实验需要设计同源臂之间的供体序列, 以便达到单碱基替换^[28-29]或引入短肽序列标签^[30]、重组酶识别位点 (*loxP*)^[30-31]或荧光蛋白报告基因^[32-33]等实验目的。本实验室和他人的研究结果均表明, 同源臂越



图中的红色序列为外源功能序列(插入/替换序列), 可以是任何序列, 例如含有点突变、序列标签或结构域缺失/改变的靶基因序列, 或用于跟内源基因融合或替换内源基因的报告基因的序列。

图1 同源重组介导的基因组靶向修饰示意图

长，敲入的效率越高^[28,32-33]。综合考虑效率和操作的便利性，在斑马鱼中同源臂的长度一般采用0.7~1.6 kb^[34]。需要注意的是，供体同源臂的序列必须跟内源基因完全一致，否则会影响同源重组的效率；同时，也要求用来注射的受精卵中来自雌雄双亲的内源基因的同源臂区域的序列也要完全一致（即同源臂序列应纯合）。不过，由于成为模式动物的历史较短，斑马鱼基因组的多态性/杂合度较高，SNP频率较高，寻找同源臂序列纯合的亲鱼并不是一件轻松的工作。表1列举了部分利用HR在斑马鱼中进行基因敲入的研究成果（表1），可以看到，目前在斑马鱼中已报道的利用CRISPR/Cas系统、通过HR途径实现的可稳定遗传的靶向敲入均为单碱基替换；通过HR途径在靶点精确敲入较长片段的供体序列，如荧光蛋白报告基因等，目前在斑马鱼中仅报道过用TALEN实现的例子^[32-34]。值得注意的是，在上述通过HR途径进行单碱基替换的实验中，为了避免重组成功后靶点被再次切割，作者

在供体中还设计了靶点序列的突变；这样，在最终获得的“单碱基突变体”中，除了目的碱基发生了点突变，在该碱基之外的靶点序列中某些碱基也同时带有突变，因此，严格而言这并非真正的“单碱基”突变，只是目的碱基之外的靶点突变如果位于外显子上，一般均设计成同义突变，理论上不会影响所编码的氨基酸。换句话说，如果希望通过EEN和HR途径得到真正意义上的单碱基突变，就要求待突变的目标碱基位于靶点之内，并且应该是靶点中的关键碱基，从而保证该碱基突变后靶点不会被二次切割，这样才能做到只突变该碱基即可达到目的，而无需额外设计其他靶点序列的突变。实际上，不仅单碱基突变实验一般需要额外突变靶点序列，在HR介导的其他基因敲入实验中一般也需要在供体中引入靶点的突变，以保证重组成功、供体序列替换了原有的基因组序列之后靶点区域不会被再次切割。只有在某些特殊情况下，才无需在靶点中额外设计突变^[33]。此外，目前已报道的案例中，同源臂

表1 在斑马鱼中利用传统长同源臂与HR实现基因敲入的代表性案例

靶基因	EEN	靶基因修饰类型	实验策略		F ₀ 效率 [#]	种系传递效率	参考文献
			同源臂大小	被替换的基因组序列(同源臂与靶点之间的距离: bp)*			
<i>th</i>	TALENs	荧光报告基因敲入(~800 bp)	800~900 bp	50 (31, 19)	~50%	1.5% (4/275)	[32]
<i>sox2</i>	TALENs	荧光报告基因敲入(~800 bp; in-frame插入到靶基因的终止密码子之前、TALENs的间隔序列中)	610 bp~3.7 kb	0 (0, 0)	最高92%	8.0% ^{&} (29/363)	[33]
<i>gfap</i>	TALENs	(同上)	>1 kb	0 (0, 0)	不详	11.3% (5/44)	[33]
<i>Alb</i> [@]	Cas9/gRNA	单碱基替换+靶点同义突变(回复点突变)	400~500 bp	70 (18, 52)	46% [§]	10.7% (3/28)	[28]
<i>Gol</i> [%]	TALENs	单碱基替换+靶点突变+ <i>len</i> 报告系统敲入(回复点突变) [^]	>1 kb	~100 (28, 80?)	100%	10.3% (4/39)	[30]
<i>kcnh6a</i>	TALENs	双 <i>loxP</i> + <i>len</i> 报告系统敲入 [^]	>1 kb	~400 (400, 0?)	100%	7.0% (3/43)	[30]

注: * 对于TALEN而言, EEN与靶点之间的距离指的是同源臂靠近靶点的一端距离两个TALEN之间的间隔序列中中心碱基的bp数。

[#] 指对F₀胚胎通过表型(色素恢复)^[28]或荧光报告基因观察^[33], 或者PCR^[30,32-33]方法估算, 得到的携带体细胞敲入的F₀占所有F₀的比例(除非测序验证, 否则无法区分或排除靶点处的非精确整合或靶点外的随机整合)。

[&] 该比率是通过观察注射多种不同大小的供体得到的F₀外交得到的F₁胚胎中荧光报告基因的表达情况估算的平均值; 根据该论文提供的Southern杂交结果推测, 大约有2/3的有荧光报告基因表达的F₁在靶点处有供体插入。

[@] 靶向的是点突变等位基因*alb*^{b4}。

[§] 胚胎存活率很低, 只有10%生长到成体。

[%] 靶向的是点突变等位基因*glo*^{b1}。

[^] *len*报告系统指的是两侧含有同向的FRT序列、跟靶基因转录方向相反的、可在斑马鱼晶状体细胞中特异表达荧光蛋白指示基因的FRT-CV-FRT cassette, 其中CV代表*α-crystallin::Venus*, 大小约为1.8 kb。该系统有助于筛选KI事件的种系传递情况, 不过对于评估F₀胚胎的KI效率应该没有什么优势。筛选完成后, CV指示基因可以用Flipase重组酶去除。

的位置都非常接近于 EEN 位点^[28,30,32-33]。一般认为, 同源重组的效率跟同源臂与 DSB 之间的距离有关, 距离越远效率越低, 且该距离最好能控制在 100 bp 之内。这样就难以通过 HR 置换大段的内源基因组序列, 给一些特殊等位基因的构建造成了一定的困难和限制。例如, 构建以 Cre/*loxP* 为基础的条件性等位基因时, 往往需要在某个重要外显子两侧的内含子中各插入一个同向的 *loxP* 序列, 如果希望通过一次同源重组实现两个 *loxP* 序列的插入, 就需要一次性整体替换该外显子及其两侧的部分内含子序列, 这样的片段一般都会超过 100 bp, 这种较大片段的置换往往会严重影响同源重组的效率, 从而大大增加了实验难度。事实上, 目前在斑马鱼中仅有一例报道利用 TALEN 成功筛选到通过一次同源重组在一个 400 bp 左右的内源外显子的两侧各插入一个 *loxP* 序列的鱼系^[30]。不过, 该实验采用了特殊的供体设计, 敲入基因组的供体序列中还额外带有一个两侧含有同向的 *FRT* 序列、可在晶状体细胞中特异表达的荧光蛋白指示基因 (称为 *len* 报告系统, 表 1), 用于辅助挑选 KI 成功的鱼系。值得一提的是, 选择外显子上的靶点进行 KI 实验时, 如果靶基因具有重要功能, 则有可能由于靶点产生大量 indel 突变而威胁到基因编辑后胚胎的存活, 从而加大筛选所需 KI 鱼系的工作量, 甚至有可能无法筛选到可存活的 KI 鱼系^[28]。因此, HR 的应用仍存在较大的挑战性, 如何提高同源重组介导的基因组精确修饰效率, 尤其是长片段精确插入/替换的效率, 同时避免高效靶点因 NHEJ 造成大量的 indel 所带来的对胚胎存活的负面影响, 依然是本领域亟待解决的一个重要问题。许多研究者探究了供体的构象 (线性化供体或环状质粒供体) 对于同源重组介导的 KI 效率的影响, 遗憾的是目前报道的实验结果并不完全一致, 特别是体外线性化的 DNA 作为供体与环状质粒直接作为供体, 两者孰优孰劣仍存在争议^[28,30,32-33]; 不过, 目前较为一致的结论是, 将环状质粒导入细胞后, 利用 *I-SceI* 归位内切酶或 Cas9/gRNA 系统对其进行体内线性化后作为 KI 的供体所产生的 KI 效率显著高于体外线性化的或环状的供体^[28,30-31], 这一现象的具体分子机理仍有待解析。此外, 在体外培养细胞和小鼠胚胎中, 有报道通过敲低/抑制 NHEJ 途径相关基因 (*Lig4*, *Ku70*, *Xrcc4*), 从而避免 NHEJ 与 HR 的竞争, 可提升同源重组的效率^[35-38]。不过, 抑制 *Lig4* 对提升同源重组的效果仍有争议^[39]。本实验室在斑马

鱼中尝试过类似的方法, 遗憾的是效果并不明显。

2.2 利用短同源臂实现基因敲入

除了利用上述传统的长同源臂 HR 进行基因敲入外, 最近有报道利用较短的同源序列 (通常为 30~50 bp, 可以短至 5 bp) 同样可以使同源臂之间的序列整合进入靶点, 有研究者将这一机制称为微同源末端连接^[40]。在小鼠或大鼠中, 短同源序列介导的基因敲入通常采用双链 DNA 供体^[41], 而根据已发表的论文, 斑马鱼中则通常使用单链寡聚核苷酸 (single-stranded oligodeoxynucleotides, ssODNs) 作为供体; 目前仅有一例用的是双链 DNA 供体^[42], 其余均为 ssODN。ssODN 一般采用化学反应进行商业化合成, 因此, 供体的长度受限; 目前的合成极限为 150 nt, 但是超过一定长度后序列质量没有保证, 费用也很高, 通常在 120 nt 之内质量比较有保证。因此, 基于 ssODN 的短同源臂途径的方法比较适用于敲入短序列。表 2 列举了一些利用短同源臂在斑马鱼中实现基因敲入的研究成果。在单碱基突变、敲入标签蛋白或 *loxP* 等短序列时, 相比传统的 HR, 短同源臂途径供体构建简便, 敲入效率也比较接近。然而, 短同源臂途径相对而言较难实现长片段的敲入, 目前仅有一例报道成功利用短同源臂的载体敲入荧光蛋白报告基因, 采用的是双链 DNA 供体^[42]。不过, 该实验结果仍存在一些问题, 似乎未能确保荧光蛋白报告基因的正常表达。例如, 测序结果显示 *tyr* 靶基因中精确敲入了 in-frame 的 *mCherry* 报告基因, 但是作者并未观察到荧光表达, 原因不明。需要注意的是, 带有短同源臂的供体也有可能通过 NHEJ 途径整合进入靶点, 从而造成非精确敲入 (在接口处形成 indel); 如果 indel 没有造成移码突变, 则通过 indel 插入的报告基因也有可能正确表达, 就像发生了精确敲入一样。这种情况一般难以通过 DNA 测序之外的方法与精确敲入进行区分。而理论上, 这种非精确整合的效率通常会高于基于同源修复的精确整合, 这显然不利于筛选精确敲入的阳性个体。由于利用短同源臂途径进行 DSB 修复, 尤其是采用 ssODN 作为供体修复的分子机理尚不明确, 因而难以从机制出发提升该方法的效率, 和 (或) 避免 NHEJ 的干扰/竞争。因此, 采用短同源臂途径进行靶向敲入时, 测序确认最终的结果显得格外重要。此外, 作为一种 DNA 序列, 除了整合到靶点中, ssODN 还有可能在靶点之外的基因组区域产生随机整合或脱靶整合; 并且有研究表明, 同源臂序列越短, 随机整合发生的频率相对

表2 在斑马鱼中利用短同源臂途径实现基因敲入的部分案例

靶基因	EEN	实验策略*			F ₀ 效率 ^{&}	种系传递效率	参考文献
		靶基因修饰类型	供体大小	供体组成 [#]			
<i>ponzr1</i>	TALENs	<i>EcoRV</i> 酶切位点敲入(插入TALENs的间隔区)	46 nt	20+7+19	~57%	约2.2% (2/4 x 8/186)	[43]
<i>crhr2</i>	TALENs	<i>mloxP</i> 敲入(替换了38 bp的内含子, 含部分间隔序列和全部右侧TALEN)	88 nt	27+34+27	~38%	5.2% (3/58)	[43]
<i>C9orf72</i>	Cas9/gRNA	HA标签敲入(in-frame插入靶点内的起始密码子之后)	120 nt	41+30+49	45%	未筛选到精确敲入	[44]
<i>tardbp</i>	Cas9/gRNA	(同上)	120 nt	33+54+33	70%	不详	[44]
<i>tyr</i>	Cas9/gRNA	荧光蛋白报告基因+质粒骨架敲入(in-frame插入靶点, 未替换内源序列)	(体内线性化的质粒)	(0~40 bp同源臂)	53%~85% [§]	不详	[42]
<i>krttlc19e</i>	Cas9/gRNA	荧光蛋白报告基因敲入(in-frame插入靶点的终止密码子之前, 未替换内源序列)	(同上)	(40 bp同源臂)	38%	15% [@]	[42]
<i>tardbp</i>	Cas9/gRNA	单碱基替换(造成靶点突变)	23 nt 42 nt	15+1+7 26+2+12 [%]	不详 不详	2.2% (1/46) 3.9% (3/77)	[45]
<i>fus</i>	Cas9/gRNA	单碱基替换(造成靶点突变)	33 nt	14+1+18	不详	2.1% (1/47)	[45]

注: * Hisano等^[42]采用的是通过Cas9/gRNA系统体内线性化双链供体质粒, 除此之外, 其他论文均采用单链寡聚核苷酸ssODN作为供体。

[#] 供体组成按照“左同源臂+敲入序列+右同源臂”的顺序依次排列。

[&] 指对F₀胚胎通过PCR/酶切^[42-44]、测序^[42]或者荧光报告基因观察^[42]估算得到的携带体细胞敲入的F₀占所有F₀的比例(除非测序验证, 否则无法排除靶点处的非精确整合或靶点外的随机整合)。

[§] 随着供体同源臂序列长度的增加(0~40 bp), 通过PCR检测到供体质粒整合的效率逐渐增加。不过, 此时并未区分精确整合与非精确整合; 根据该论文提供的数据推断, 供体同源臂序列长度为0时检测到的F₀胚胎的供体敲入事件(53%)均为非精确整合。

[@] 针对胚胎期有荧光表达的F₀进行筛选的结果。

[%] 供体右侧同源序列上携带同义突变, 该实验是为了证明确实是由于ssODN的重组导致*tardbp*靶点位置产生了点突变, 而非NHEJ导致的。

于同源整合就会越高^[42]。由于ssODN这种供体的序列较短, 鉴定随机插入的情况、追踪随机插入的位点无疑会比较困难; 如何确保最终获得的鱼系中ssODN只插入到了靶位点, 如何鉴别并排除其在基因组其他位点的随机整合, 这也是一个需要格外注意并有待解决的问题。

总之, 以同源修复为基础的基因组精确敲入, 无论是基于长同源臂的传统HR途径, 还是基于微同源序列的修复途径, 是目前最为精确、潜在用途最广泛, 因而最被模式动物科研工作者期待的基因敲入技术。然而, 即使在应用人工核酸内切酶诱导产生DSB后, 同源重组的效率仍然很低, 并且极大地依赖于高效的靶点^[34]。此外, 这一方法的效率还很容易受到其他因素的影响, 如同源臂中若存在错配碱基便会大大降低重组效率, 因此, 往往需要

谨慎挑选同源臂区域序列一致(两条同源染色体序列纯合)的亲本进行实验。这对于SNP较多的斑马鱼而言无疑意味着相当大的工作量。因此, 在致力于探索提高同源重组效率的方法的同时, 人们还设法寻求更容易实现且可满足研究需求的替代方法, 如利用占主导地位的李氏修复途径, 将线性化的外源DNA供体整体直接插入靶点的DSB处, 从而实现DNA大片的靶向敲入。

3 基于非同源末端连接途径的基因组靶向插入

前面已经介绍, 同源重组事件在细胞中发生的概率很低, 因此, HR介导的基因敲入效率整体偏低。对感兴趣的基因进行精确靶向敲入, 需要进行大量筛选, 并且不能保证成功。人工核酸酶诱导DNA发生双链断裂后, 细胞主要是通过NHEJ的方式进

行损伤修复, 发生 NHEJ 的几率要远高于 HR, 因此, 利用 NHEJ 途径介导基因敲入, 或称靶向插入、定点插入, 其效率理论上会高于同源重组, 只是供体 DNA 会作为一个整体全部插入靶点 (含有不需要的载体骨架), 并且插入位点/靶点的基因组内源序列与供体序列的连接之处往往会带有 indel, 造成不精确的插入。目前 NHEJ 介导的 DNA 靶向插入效率较高的做法, 一是针对内源靶基因筛选高效的 EEN 靶位点, 以确保在基因组上高效产生 DSB; 二是在供体 DNA 上引入高效的 EEN 靶点 (跟内源靶点相同或不同均可), 将供体以环状质粒的方式跟 EEN 系统共同注射到胚胎中, 使其在体内被线性化, 这样在 NHEJ 修复过程中就有可能将供体 DNA 整体插入到基因组的 DSB 位点。表 3 总结了目前已发表的利用 NHEJ 在斑马鱼中实现基因组靶向插入的实验案例 (表 3), 可见该方法目前主要用于向靶位点引入 *Gal4*、荧光蛋白等长片段外源基因。在最初发表的靶向插入研究中, 来自法国 Del Bene 研究组的人员将靶点设计在转基因斑马鱼的 *EGFP* 编码区, 供体中含有 *Gal4* 编码序列, 供体成功插入后 *EGFP* 被破坏, *Gal4* 表达, 从而引起 *Tg (UAS:mCherry)* 转基因鱼表达红色荧光, 出现荧光信号的转换。通过这样的靶向插入, 能够将转基因动物的绿色荧光标记转变为红色荧光, 并且可以遗传给后代, 其效率高于同源重组介导的长片段精确敲入, 种系传递的效率可达 10%^[46]。利用上述方法能够有效地对靶基因进行遗传标记, 只不过这并不是一个精确的标记。不过, 以编码区/外显子作为插入位

点还会额外带来两个问题。(1) 由于插入位点位于 *EGFP* 基因的编码区, 且供体仅提供了 *Gal4* 的编码序列而自身并不具有启动子和增强子, 因此, *Gal4* 需要以正确的方向插入 *EGFP* 基因, 并且跟 *EGFP* 基因处于相同的读码框 (in-frame) 才能正确表达。由于 NHEJ 修复并不精确, 同时再考虑到插入方向的问题, 这就导致供体插入后只有六分之一的可能性满足上述要求, 成为有效的插入; 这无疑大大降低了实验效率, 增加了工作量。(2) 这一策略往往会导致内源基因被破坏。在某些情况下, 人们希望在标记内源基因的同时, 并不破坏内源基因的表达和功能。为此, 有研究组改变策略, 通过靶向目的基因的上游序列, 并在供体基因的编码区前添加基本启动子序列的方式, 在不破坏内源基因的情况下实现了基因标记, 同时也避免了读码框的问题^[47]。这实际上相当于达到了类似于增强子诱捕 (enhancer trap) 的效果。但是, 靶向内源基因的上游序列也有可能影响该基因的调控序列, 并且报告基因的启动子是外源提供的, 这种遗传标记是否能够完全真实地模拟内源基因的表达情况依然存疑。

为了避免上述策略的缺陷和不足, 中国科学院上海神经科学研究所杜久林研究组对上述方案进行了进一步改进, 靶点选择在最后一个内含子的区域, 在供体中引入靶点序列 (用于供体的体内线性化) 以及靶点下游的内含子 (包括剪接受体序列), 同时也将其余的外显子序列 (包括编码序列和 3'-UTR) 补充完整 (图 2)^[48]。这样, 通过 mRNA 的剪接, 靶点下游剩余的内源基因编码序列和外源报告基因能

表3 在斑马鱼中利用NHEJ途径实现基因敲入的案例

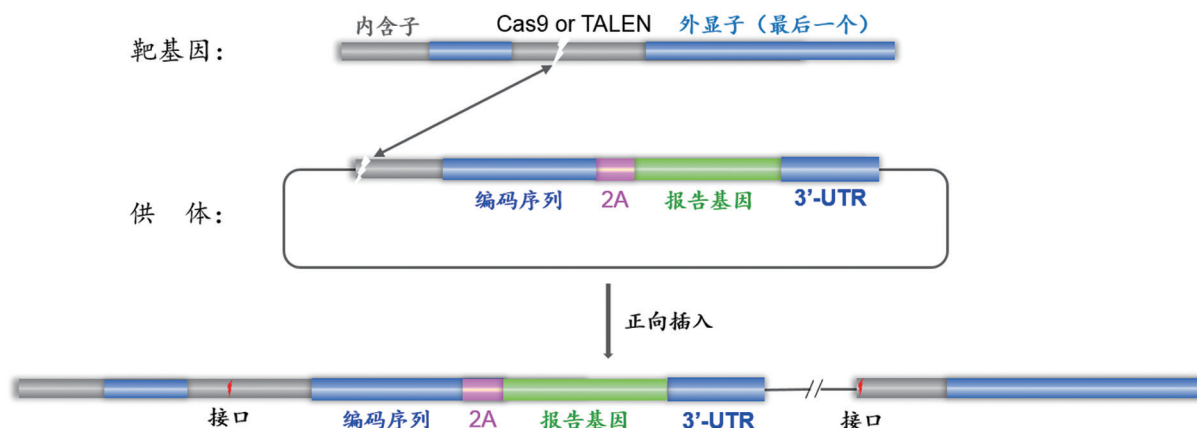
靶基因	EEN及靶点位置	实验策略*	F ₀ 效率	种系传递效率	参考文献
<i>neurod:eGFP</i>	Cas9/gRNA, TALENs; 编码区	<i>Gal4 (E2A-KalTA4)</i> 插入	~76%	10.3% [#] (3/29)	[46]
<i>kif5aa</i>	(同上)	(同上)	~10%	不详	[46]
<i>glyt2</i>	Cas9/gRNA; 启动子上游序列	最小启动子+荧光标记可转换 的报告基因插入(通过 <i>loxP</i>)	(5%~10%) ^{&}	60.0% [§] (6/10)	[47]
<i>eng1b</i>	(同上)	(同上)	(5%~10%) ^{&}	30.0% [§] (3/10)	[47]
<i>th</i>	Cas9/gRNA; 内含子	荧光报告基因插入	~24%	12.0% [§] (3/25)	[48]
<i>gfap</i> 等	(同上)	<i>Gal4</i> 插入 荧光报告基因插入	不详 12%~36%	7.1% (2/28) 不详	[48]

注: * 供体均采用了体内线性化。

[#] 仅统计了阅读框正确的个体; 如果计入阅读框改变的个体, 则为9/29(31.0%)。

[&] 仅统计了荧光表达广泛的F₀。

[§] 仅筛选了有荧光表达的F₀。



在靶基因的内含子和供体中分别产生DSB后，线性化的DNA供体就有可能正向插入靶点区域。报告基因一般编码荧光蛋白，跟靶基因共同转录、共同翻译，然后通过2A肽段的切割跟靶基因产物分开，成为独立的荧光蛋白，一方面可反映靶基因的时空表达情况，同时也标记了靶基因表达的细胞。也可以去掉靶基因编码序列与报告基因之间的2A序列，这样内源靶基因与报告基因就会作为一个融合蛋白进行共同表达。这样做的好处是相当于使内源蛋白带上了一个序列标签，除了仍然可以在细胞水平标记靶基因的时空表达谱，还可以实现更精细的标记，如标记靶蛋白的亚细胞定位。然而，需要注意的是，报告基因的融合有可能对内源蛋白三维结构、定位和(或)功能造成一定的影响。

图2 靶向内含子的基因插入示意图

够跟内源的上游 mRNA 无缝连接，既避免了阅读框的问题，又可避免对内源基因的破坏，是目前最灵活的一种靶向插入方法。

总之，NHEJ 介导的靶向插入是一种较为高效的基因敲入方式，目前主要用于在基因组的特定位点插入外源基因，以实现对内源基因或特定的细胞(组织)进行遗传标记，或者在特定的细胞/组织中表达外源基因。理论上通过设计更加复杂、含有更多元件的供体，应该有可能达成其他实验目的。不过，这一技术同样依赖于高效率的靶点，并且在序列敲入的精确性以及对内源序列的还原程度方面，跟同源重组介导的精确基因敲入存在一定差距；同时，还引入了载体骨架等并不需要的外源序列，有可能会对最终的基因敲入效果造成干扰^[49]。可以预见，这一技术的进一步改进与完善，将和同源重组介导的基因修饰一起，帮助人们在未来实现更多种多样的复杂的基因组修饰。

4 总结与展望

综合前文的介绍，目前已经有多种方法能够在模式动物斑马鱼中靶向敲入外源 DNA 序列，包括碱基的精确替换、短肽标签序列或重组酶识别位点(如 *loxP*)敲入、荧光蛋白报告基因或其他功能元件的敲入等。在这些方法中，一部分是由体外培养细胞或小鼠等模式系统中的方法借鉴和改进而

来；而另一些则是首先运用于斑马鱼中，理论上可以推广到其他模式动物中。正如之前所总结的，同源重组介导的靶向敲入与非同源末端连接介导的靶向插入技术各有优劣，研究者可根据自身研究工作的需要，选择最为合适的方案进行实验，从而用最简便的方式达成实验目的(表4)。如果要求对靶基因进行精确修饰(包括碱基或序列的精确替换)，一般需要采用同源序列介导的靶向敲入；如果插入或替换的片段较小(例如碱基或短序列替换，插入 *loxP*、酶切位点或短序列标签)，则可采用基于短同源臂的靶向敲入策略。需要注意的是，如果要制备严格意义上的单碱基替换突变体(仅突变目标碱基，其他碱基不能变)，通过 EEN 和 HR 途径进行这一实验时对靶点的选择就会有比较严格的限制，要求待突变的碱基位于靶点内比较重要的位置才有可能达到目的。如果不要求精确修饰，则效率更高的 NHEJ 介导的靶向插入应该是更好的选择。跟同源序列介导的靶向敲入相比，该策略不仅效率较高，可以在靶点引入较大的片段(例如荧光报告基因)，而且在靶点的选择和供体的构建上也更加灵活。此外，在对靶基因进行的修饰的类型与效果上，基于 NHEJ 的靶向插入策略也比较灵活，理论上可以通过在供体上添加一种或多种功能元件，实现更加复杂多样的基因组修饰，如模拟点突变或任意其他类型的突变、基因或细胞的报告基因标记、条件性等

表4 三种基因组靶向敲入方法的比较

敲入方法	效率	供体		敲入效果	用途		
		供体构成 ^d	供体构建 ^e		单碱基替换/序列替换 ^a	短片段插入 ^b	长序列插入 ^c
长同源臂介导的HR	很低	长同源臂 ^d	复杂 ^e	精确	√	√	√
微同源臂途径	低	短同源臂 ^f	简单 ^g	精确 ^h	√	√	? ⁱ
NHEJ介导的靶向插入	高	无需同源臂	灵活	不精确	?	?	√ ^j

注: a. 一般只能置换较短的内源序列(例如100 bp以内的内源序列)。

b. 适用于敲入酶切位点、短肽序列标签、重组酶识别位点(如*loxP*, 以实现条件性基因敲除)等短序列。

c. 可敲入荧光蛋白报告基因、*Gal4*基因等大片段。

d. 长同源臂介导的HR途径的供体一般采用的是长dsDNA; 同源臂越长越好, 大小一般为0.7~1.6 kb; 供体体内线性化敲入效率较高。

e. 需要克隆较长的同源臂。如果供体中包含靶点序列(位于同源臂中或插入/替换序列中), 则一般需要对该序列进行适当突变, 以避免重组成功后靶点被再次或反复切割^[28,30]。需要注意的是, 如果靶点序列位于同源臂上, 对靶点进行突变会影响同源重组的效率。此外, 如果是制备单碱基突变/替换, 还要考虑在目的碱基之外对靶点进行额外的突变是否对实验目的有影响。

f. 目前斑马鱼中微同源臂途径的供体采用的主要是ssODN; 同源臂较短, 大小通常为30~50 bp, 也可以短至5 bp。

g. ssODN一般直接从公司订购; 双链供体如果序列较短, 也可直接从公司订购。如果供体中包含靶点序列(位于同源臂中或插入/替换序列中), 则一般需要对该序列进行适当突变, 以避免重组成功后靶点被再次或反复切割^[28,30]。需要注意的是, 如果靶点序列位于同源臂上, 对靶点进行突变会影响同源重组的效率。此外, 如果是制备单碱基突变/替换, 还要考虑在目的碱基之外对靶点进行额外的突变是否对实验目的有影响。

h. 由于存在NHEJ机制的竞争, 理论上会导致较多的不精确插入, 除非测序, 否则难以跟精确敲入相区分。此外, ssODN在靶点之外的随机插入较难鉴定与追踪。

i. 难以实现大片段的敲入, 目前仅有一例报道。

j. 存在无用的载体骨架的插入。

位基因的高效一步构建, 以及同步实现条件性基因敲除与等位基因标记或基因型标记等。

目前看来, 制约基因敲入实验成功与否的最大影响因素是靶点的切割效率, 即单独使用 EEN 时造成 indel 的效率。据笔者不完全统计, 在目前报道成功的案例中, 所使用的 EEN 的靶点效率均在 70% 以上(包含不同的检测方案, 如二代测序、T7E1 酶切、限制性内切酶法等)。可以说, 筛选得到高效率的 EEN 靶点是成功实现基因敲入的第一步, 也是最为关键的步骤。根据笔者实验室使用 Cas9/gRNA 系统进行靶向诱变的经验, 跟外显子相比, 内含子靶点效率普遍偏低, 筛选到高效靶点的难度较大。因此, 迫切需要能够高效评估靶点效率或筛选高效靶点的方法和技巧。有研究表明, 在模式动物中, 使用 Cas9 蛋白对同一靶点的 indel 效率一般高于使用 mRNA; 不过, Del Bene 研究组曾报道, 能够使用 Cas9 mRNA 实现 NHEJ 靶向插入的位点, 在使用 Cas9 蛋白后, 无论是 F₀ 体细胞还是种系传递均未检测到基因敲入事件, 原因不明^[34]。最后, 由于基因敲入的效率相比基因敲除仍然较低, 所以在设计供体时, 最好能引入一些便于筛选阳性结果的元件, 如报告基因和酶切位点等, 以便简单

估计敲入效率, 并对阳性个体进行准确、简便的筛选。此外, 基于 NHEJ 的靶向插入虽然效率较高, 但在具体应用上目前仍较为单调, 有待进一步开发和优化, 特别是通过优化供体设计实现更多的基因组修饰效果。

虽然人们现在已经拥有相当丰富的基因敲入工具箱, 然而, 事实上在模式动物中实现基因敲入仍然比较困难, 这不仅仅是因为效率普遍不高, 还由于技术细节过多、影响因素复杂、基因敲入的分子机制尚不完全明确等因素, 导致该技术难以学习和推广。要解决这一问题, 需要多个领域的研究者从不同角度共同努力, 改进和优化现有的基因敲入实验条件和供体形式, 进一步解析基因敲入过程中 DNA 损伤修复的精细分子机制, 以及发展更为高效、精确、普适的基因组编辑工具等。如果有一天这些问题都得到了彻底的解决, 基因敲入技术极大发展, 让人们能够随心所欲地、如同在磁盘上随意写入和删除信息一般任意操作基因组, 那么不仅能够极大地推进生命科学基础理论研究, 也必将有利于构建疾病的动物模型、模拟并活体追踪细胞的癌化与转移、方便药物筛选等多个医疗健康相关领域的研究与应用, 并且还能推广到家禽家畜和农作物、

宠物等的优化育种，甚至临床治疗中，产生不可估量的深远影响，造福人类与大自然。

[参 考 文 献]

- [1] 肖安, 张博. 人工核酸内切酶介导的新一代基因组编辑技术进展. 生物工程学报, 2015, 31: 917-28
- [2] Gabriel R, Lombardo A, Arens A, et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat Biotechnol, 2011, 29: 816-23
- [3] Pattanayak V, Ramirez CL, Joung JK, et al. Revealing off-target cleavage specificities of zinc-finger nucleases by *in vitro* selection. Nat Methods, 2011, 8: 765-70
- [4] Xiao A, Wu Y, Yang Z, et al. EENdb: a database and knowledge base of ZFNs and TALENs for endonuclease engineering. Nucleic Acids Res, 2013, 41: D415-22
- [5] Moore FE, Reyon D, Sander JD, et al. Improved somatic mutagenesis in zebrafish using transcription activator-like effector nucleases (TALENs). PLoS One, 2012, 7: e37877
- [6] Mussolino C, Morbitzer R, Lutge F, et al. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res, 2011, 39: 9283-93
- [7] 沈延, 肖安, 黄鹏, 等. 类转录激活因子效应物核酸酶 (TALEN) 介导的基因组定点修饰技术. 遗传, 2013, 35: 395-409
- [8] 李君, 张毅, 陈坤玲, 等. CRISPR/Cas系统: RNA靶向的基因组定向编辑新技术. 遗传, 2013, 35: 1265-73
- [9] 郑武, 谷峰. CRISPR/Cas9的应用及脱靶效应研究进展. 遗传, 2015, 37: 1003-10
- [10] Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science, 2012, 337: 816-21
- [11] Slaymaker IM, Gao L, Zetsche B, et al. Rationally engineered Cas9 nucleases with improved specificity. Science, 2016, 351: 84-8
- [12] Kulcsar PI, Talas A, Huszar K, et al. Crossing enhanced and high fidelity SpCas9 nucleases to optimize specificity and cleavage. Genome Biol, 2017, 18: 190
- [13] 周金伟, 徐绮斌, 姚婧, 等. CRISPR/Cas9基因组编辑技术及其在动物基因组定点修饰中的应用. 遗传, 2015, 37: 1011-20
- [14] Lee HJ, Kim E, Kim JS. Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome Res, 2010, 20: 81-9
- [15] Xiao A, Wang Z, Hu Y, et al. Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. Nucleic Acids Res, 2013, 41: e141
- [16] Nemudryi AA, Valetdinova KR, Medvedev SP, et al. TALEN and CRISPR/Cas genome editing systems: tools of discovery. Acta Naturae, 2014, 6: 19-40
- [17] Huang P, Xiao A, Zhou M, et al. Heritable gene targeting in zebrafish using customized TALENs. Nat Biotechnol, 2011, 29: 699-700
- [18] Panda SK, Wefers B, Ortiz O, et al. Highly efficient targeted mutagenesis in mice using TALENs. Genetics, 2013, 195: 703-13
- [19] Chang N, Sun C, Gao L, et al. Genome editing with RNA-guided Cas9 nuclease in zebrafish embryos. Cell Res, 2013, 23: 465-72
- [20] Hwang WY, Fu Y, Reyon D, et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol, 2013, 31: 227-9
- [21] Liu D, Wang Z, Xiao A, et al. Efficient gene targeting in zebrafish mediated by a zebrafish-codon-optimized cas9 and evaluation of off-targeting effect. J Genet Genomics, 2014, 41: 43-6
- [22] Sung YH, Kim JM, Kim HT, et al. Highly efficient gene knockout in mice and zebrafish with RNA-guided endonucleases. Genome Res, 2014, 24: 125-31
- [23] Kok FO, Shin M, Ni CW, et al. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. Dev Cell, 2015, 32: 97-108
- [24] Rossi A, Kontarakis Z, Gerri C, et al. Genetic compensation induced by deleterious mutations but not gene knockdowns. Nature, 2015, 524: 230-3
- [25] El-Brolosy MA, Stainier DYS. Genetic compensation: a phenomenon in search of mechanisms. PLoS Genet, 2017, 13: e1006780
- [26] Thomas KR, Capecchi MR. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell, 1987, 51: 503-12
- [27] Urnov FD, Miller JC, Lee YL, et al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature, 2005, 435: 646-51
- [28] Irion U, Krauss J, Nusslein-Volhard C. Precise and efficient genome editing in zebrafish using the CRISPR/Cas9 system. Development, 2014, 141: 4827-30
- [29] Zhang Y, Huang H, Zhang B, et al. TALEN- and CRISPR-enhanced DNA homologous recombination for gene editing in zebrafish. Methods Cell Biol, 2016, 135: 107-20
- [30] Hoshijima K, Juryneć MJ, Grunwald DJ. Precise editing of the zebrafish genome made simple and efficient. Dev Cell, 2016, 36: 654-67
- [31] Hoshijima K, Juryneć MJ, Grunwald DJ. Precise genome editing by homologous recombination. Methods Cell Biol, 2016, 135: 121-47
- [32] Zu Y, Tong X, Wang Z, et al. TALEN-mediated precise genome modification by homologous recombination in zebrafish. Nat Methods, 2013, 10: 329-31
- [33] Shin J, Chen J, Solnica-Krezel L. Efficient homologous recombination-mediated genome engineering in zebrafish using TALE nucleases. Development, 2014, 141: 3807-18
- [34] Albadri S, Del Bene F, Revenu C. Genome editing using CRISPR/Cas9-based knock-in approaches in zebrafish. Methods, 2017, 121-122: 77-85
- [35] Chu VT, Weber T, Wefers B, et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol, 2015, 33: 543-8
- [36] Maruyama T, Dougan SK, Truttmann MC, et al. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat

- Biotechnol, 2015, 33: 538-42
- [37] Bertolini LR, Bertolini M, Maga EA, et al. Increased gene targeting in Ku70 and Xrcc4 transiently deficient human somatic cells. *Mol Biotechnol*, 2009, 41: 106-14
- [38] 李国玲, 钟翠丽, 莫健新, 等. 动物基因组定点整合转基因技术研究进展. *遗传*, 2017, 39: 98-109
- [39] Zhang JP, Li XL, Li GH, et al. Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome Biol*, 2017, 18: 35
- [40] Sfeir A, Symington LS. Microhomology-mediated end joining: A back-up survival mechanism or dedicated pathway? *Trends Biochem Sci*, 2015, 40: 701-14
- [41] Nakade S, Tsubota T, Sakane Y, et al. Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat Commun*, 2014, 5: 5560
- [42] Hisano Y, Sakuma T, Nakade S, et al. Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Sci Rep*, 2015, 5: 8841
- [43] Bedell VM, Wang Y, Campbell JM, et al. *In vivo* genome editing using a high-efficiency TALEN system. *Nature*, 2012, 491: 114-8
- [44] Hruscha A, Krawitz P, Rechenberg A, et al. Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development*, 2013, 140: 4982-7
- [45] Armstrong GA, Liao M, You Z, et al. Homology directed knockin of point mutations in the zebrafish *tardbp* and *fus* genes in ALS using the CRISPR/Cas9 system. *PLoS One*, 2016, 11: e0150188
- [46] Auer TO, Durore K, De Cian A, et al. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res*, 2014, 24: 142-53
- [47] Kimura Y, Hisano Y, Kawahara A, et al. Efficient generation of knock-in transgenic zebrafish carrying reporter/driver genes by CRISPR/Cas9-mediated genome engineering. *Sci Rep*, 2014, 4: 6545
- [48] Li J, Zhang BB, Ren YG, et al. Intron targeting-mediated and endogenous gene integrity-maintaining knockin in zebrafish using the CRISPR/Cas9 system. *Cell Res*, 2015, 25: 634-7
- [49] Li J, Zhang B, Bu J, et al. Intron-based genomic editing: a highly efficient method for generating knockin zebrafish. *Oncotarget*, 2015, 6: 17891-4