

特约评述

DOI: 10.12211/2096-8280.2024-082

活细胞记录器在细胞谱系追踪中的应用和前景

姜百翼, 钱珑

(北京大学定量生物学中心, 北京 100871)

摘要: 基于DNA的活细胞分子记录器技术, 通过诱导可遗传的DNA变异, 为细胞历史的追溯提供了一种创新手段。作为新一代细胞谱系追踪方法的代表, 该技术能够与单细胞测序、多组学测序等技术相集成, 帮助科研人员重构细胞发育分化路径及肿瘤起源的谱系发生树, 是探究这些核心生物学议题的有效平台。本综述系统性回顾了自2016年以来基于Cas9的分子记录器在谱系追踪领域的技术演变轨迹与应用进展, 同时综合分析了一些新型分子记录器的研究动态, 并对其优势与局限性进行了评估。自2016年以来, 以CRISPR-Cas9系统为核心的分子记录器取得了显著进展, 并逐渐成为该领域的主流技术, 研究人员在优化编辑效率和增加记录位点等方面进行了充分的探索。尽管如此, 以Cas9为基础的分子记录器仍面临CRISPR-Cas9系统固有限制与挑战, 例如DNA双链断裂带来碱基缺失, 进而引起记录信息丢失。这促使研究者们探索开发新型分子记录器, 以期作为谱系追踪的更高效精准的工具。先导编辑器、DNA结合蛋白融合碱基编辑器以及T7转录聚合酶融合碱基编辑器等基于新原理的分子记录器能够避免DNA双链断裂, 以碱基替换而非碱基缺失的形式写入信息。相较于Cas9系统, 它们展现出独特优势, 同时也伴随着潜在的风险与挑战。先导编辑器可以以时间顺序的方式记录信息, 但脱靶效应仍然是一个问题。DNA结合蛋白融合碱基编辑器提高了编辑效率和特异性, 但它们在细胞类型中的有效性需要进一步探索。T7 RNA聚合酶融合碱基编辑器已经在体内定向进化系统中取得了成功, 但它们在哺乳动物系统中的应用仍然有限。未来, 基于DNA的分子记录器的研究应着重于优化编辑效率、降低信息丢失率、提高谱系恢复效率, 并探索其在复杂生物系统中的应用潜力。

关键词: 分子记录器; 细胞谱系追踪; CRISPR相关蛋白; 基因突变; 碱基编辑

中图分类号: Q816 文献标志码: A

Application and prospect of live cell DNA-based molecular recorders in cell lineage tracing

JIANG Baiyi, QIAN Long

(Center for Quantitative Biology, Peking University, Beijing 100871, China)

Abstract: Tracing the division and differentiation history of cells is a critical issue in organismal development and

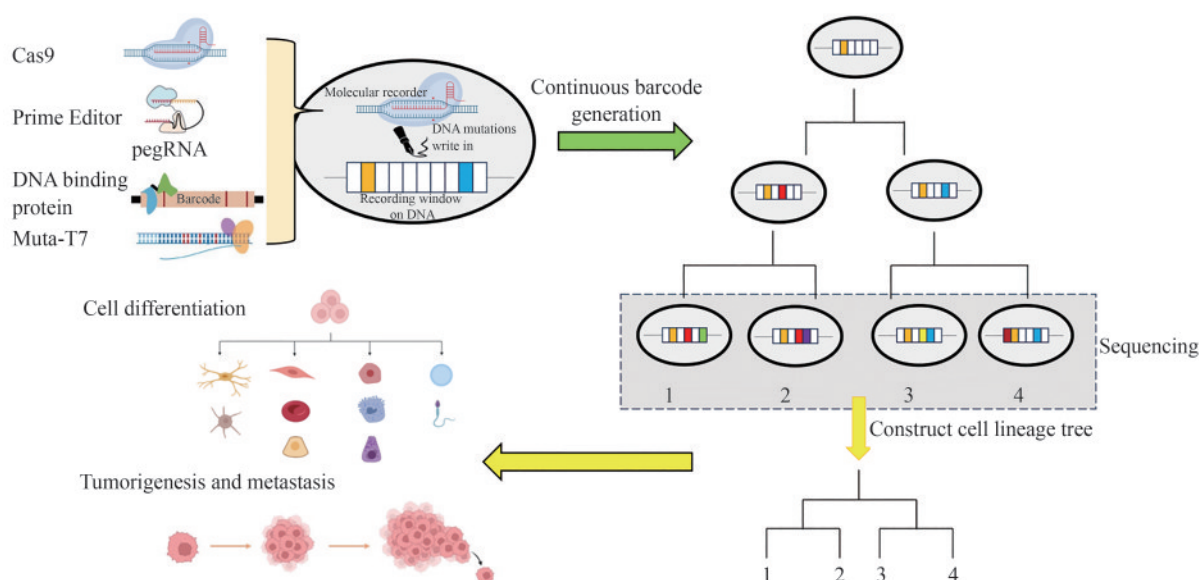
收稿日期: 2024-11-27 修回日期: 2025-03-03

基金项目: 国家重点研发计划 (2020YFA0906900)

引用本文: 姜百翼, 钱珑. 活细胞记录器在细胞谱系追踪中的应用和前景[J]. 合成生物学, 2025, 6(3): 651-668

Citation: JIANG Baiyi, QIAN Long. Application and prospect of live cell DNA-based molecular recorders in cell lineage tracing[J]. Synthetic Biology Journal, 2025, 6(3): 651-668

cancer research. Live cell DNA-based molecular recorders, a synthetic system that induces heritable DNA variations, offers an innovative approach for reconstructing cell lineage histories. As a representative for the new generation of cell lineage tracing method, this system can be integrated with high-throughput single-cell sequencing and multi-omics analysis, enabling the reconstruction of developmental differentiation pathways of cells and the phylogenetic trees of tumorigenesis as well. Live cell DNA-based molecular recorders serve as an effective platform for exploring these core biological processes. This review systematically analyzes the technological evolution of Cas9-based molecular recorders in lineage tracing since 2016 and its applications, while also analyzing the research trends of some novel molecular recorders and evaluating their advantages and limitations. Since 2016, molecular recorders based on the CRISPR-Cas9 system have made significant progress and gradually become the mainstream technology in this field. However, Cas9-based molecular recorders still suffer from several inherent limitations, such as the low lineage resolution due to insufficient editing efficiencies, the loss of recorded information caused by DNA double-strand breaks, and potential lineage merging due to barcode homoplasmy. These limitations pose challenges for researchers to explore and develop new types of molecular recorders as more efficient and precise tools for cell lineage tracing. Novel molecular recorders based on new principles, such as prime editors, DNA-binding protein-fused base editors, and T7 RNA polymerase-fused base editors, can avoid DNA double-strand breaks and record information through base substitutions rather than deletions. Compared to the Cas9 system, they exhibit unique advantages but also come with potential risks and challenges. Prime editors can record information in a temporal sequential manner, but off-target effects remain a concern. DNA-binding protein-fused base editors offer high editing efficiencies and specificities, but their effectiveness across different cell types requires further exploration. T7 RNA polymerase-fused base editors have achieved success in *in vivo* directed evolution systems, but their application in mammalian systems is still limited. In the future, the research of DNA-based molecular recorders should focus on optimizing editing efficiency, reducing information loss, improving lineage recovery efficiency, and exploring their application potentials in complicated biological systems.



Keywords: molecular recorders; cell lineage tracing; CRISPR-related protein; gene mutation; base editing

在生命科学研究中,理解细胞的发育、分化及其在疾病(尤其是肿瘤)发生和进展中的行为,始终是人们关注的重点。每一个有性生殖的生物,最初都是由一个合子(受精卵)发育而来。受精卵通过细胞分裂和分化,形成多种不同类型的子代细胞,这些细胞在特定的位置执行特定功能,共同构成完整的生物个体^[1-2]。因此,探索从受精卵到成熟个体的发育过程,以及在这一过程中细胞如何进行自我更新和分化,是生物学研究的基本任务之一^[3-4]。对于肿瘤学家来说,肿瘤内部的癌细胞可能属于不同的亚克隆群体(subclones)^[5-6]。在肿瘤发展的过程中,研究体细胞获得的癌症驱动突变、肿瘤的治疗耐药性和转移行为是如何随着肿瘤的发展而变化的,具有重要意义^[7-8]。近年来,人们关注到细胞亚克隆群体的动态交互与逐渐获得的遗传和表观遗传改变是推动其进展的关键^[9],越来越多的研究关注于监测肿瘤从单个细胞发展为侵袭性肿瘤的过程,识别驱动肿瘤扩展的稀有亚克隆及其转录程序,进而揭示肿瘤细胞可塑性的变化和肿瘤进化的多条路径,最终有可能对临床实践产生直接影响^[10-12]。

为实现上述目标,细胞谱系追踪(cell lineage tracing)技术应运而生。该技术旨在标记和追踪细胞在发育过程中的历史,揭示其演化和分化路径^[1-4]。在基于动态的分子记录器的谱系追踪技术之前,也有一些其他的谱系追踪方法。在发育生物学领域,早期的异体移植和嵌合体技术^[13-14]以及21世纪初的利用报告基因进行静态标记^[15-17]等方法也可以在一定程度上重建细胞谱系。这种方法被统称为前瞻性(prospective)谱系追踪,意为在细胞尚未分化前就引入对其谱系的标记^[5],但是对于哺乳动物的复杂发育进程而言,其时间分辨率处于一个非常低的水平,并且难以兼容高通量测序技术^[1-2]。在肿瘤生物学领域,有一些研究通过静态条形码来追踪肿瘤发生过程中原癌基因的突变路径^[18-20]。还有研究开发了带有静态条形码的肿瘤谱系,研究肿瘤在获得耐药性突变时的进化路径^[21-23]。另外一些传统的谱系追踪依赖自然发生的基因组变异,通过一些基于体细胞突变的分析可以初步回溯肿瘤的谱系^[24-25]。这也被称为回溯性(retrospective)谱系追踪,与前瞻性谱系追踪

相对,意为在细胞分化结束后回溯推断细胞的谱系发生树^[5]。然而,尽管这些技术提供了一定的视角,但受限于突变频率过低、测序技术的成熟度和样本完整性问题,使得对亚克隆层次的解析变得困难^[26-27]。

随着科学技术的进步,传统的细胞谱系追踪方法逐渐演变,尤其是近十年来,基于DNA的活细胞分子记录器技术的出现,为我们提供了全新的视角和工具^[28-29]。基于DNA的活细胞分子记录器,尤其是CRISPR-Cas9系统的应用,标志着细胞谱系追踪技术的一次重大飞跃。这类新技术不仅提高了谱系追踪和记录的精确性和时效性,也在更大范围内推动了个体发育和肿瘤生物学的研究。通过对细胞分裂和分化过程的高通量监测,研究者们能够深入了解细胞的谱系关系及其在各种生理和病理状态下的动态变化^[30-32]。随着对这项技术的不断优化和发展,我们期待它在未来的科学研究中发挥更加重要的作用。

本综述总结了活细胞分子记录器技术的一般性设计原理和设计需求,对传统的基于Cas9的分子记录器在谱系追踪领域的技术演变轨迹与应用进展进行了系统性的回顾。由于基于Cas9的分子记录器有一定的固有缺陷,因此也催生了一些新型分子记录器的研究,本综述也对这些新型分子记录器在谱系追踪领域的应用进行了详细的探讨,并对其优势与局限性进行了评估。

1 活细胞分子记录器应用于细胞谱系追踪研究的基本原理和设计需求

基于DNA的活细胞分子记录器技术,作为细胞谱系追踪的创新工具,旨在通过引入DNA序列上的可遗传标记来记录细胞的历史。其核心理念是利用基因编辑技术在特定的基因位点上产生突变,这些突变在细胞分裂和发育过程中被传递到子代细胞,从而形成细胞谱系的记录。这一技术的关键在于如何有效地实现标记的多样性和可测量性,以便在后续分析中揭示细胞的发育轨迹。其本质是利用某种转基因方法(例如转染/显微注射)将能够产生细胞DNA多样性的外源蛋白或其

他基因工程材料,例如重组酶系统^[33-35]、转座酶系统^[36]、CRISPR-Cas9系统^[37-38]等,递送进入培养皿中的细胞或动物胚胎细胞,进而自发地或在外界人为控制下随着时间的推移,在成功转基因的细胞内逐渐积累DNA的多样性,并且在不同的子代细胞中能够随机产生不同的DNA特征。最后,通过单细胞测序,读取每个细胞的DNA信息,利用序列分析等生物信息学的手段,建立细胞发育过程的谱系发生树。其中,主流的手段是利用CRISPR-Cas9系统引起靶向区域DNA的持续突变的方法^[39-40]。

2016年,用于谱系追踪的合成靶标阵列基因组编辑技术(genome editing of synthetic target arrays for lineage tracing, GESTALT)的发明,可以说是新一代细胞谱系追踪技术的起点。研究人员通过显微注射的方法,将Cas9蛋白和10种不同的sgRNA导入斑马鱼的胚胎。这10个sgRNA靶向斑马鱼基因组中的10个预先编辑好的位点(来自上一代的成鱼),这些位点串联为长度约为250 bp的基因座,被插入一个编码绿色荧光蛋白的EGFP基因的3'端的非编码区^[41]。由于Cas9蛋白会引起DNA双链断裂,进而通过内源的DNA修复机制在断裂位点附近产生随机的碱基缺失,因此这10个sgRNA的靶点附近序列会随着细胞的分裂被编辑。由于每次编辑的发生都是随机的,因此,一旦一个细胞分裂成两个细胞,这两个子代细胞会共享在祖代细胞中产生的编辑^[41]。然而在分裂之后,子代细胞中分别发生的后续编辑却不一定相同。因此,共享越近的祖代细胞的细胞之间的共享编辑会越多,据此原理,结合一定的算法,就能够建立细胞的谱系发生树^[41]。

基于上述基本原理,一个理想的、适合于细胞谱系追踪的分子记录器,理论上应该具有如下特点^[42-44](图1):第一,分子记录器应该无明显细胞毒性,并且具有简洁性,不应占用太多细胞资源,对细胞的基因表达和调控程序以及细胞生理状态无影响。同时,分子记录器应该无脱靶效应,不应对其他区域的DNA进行突变。第二,分子记录器应该具有合适的、可调的记录效率,因为过高的突变率会导致编辑序列的突变饱和,过低的突变率会导致具有区分度的突变过少,这两

种情况都会导致条形码同质性(barcode homoplasmy),无法有效地构建谱系发生树。因此,对于不同的应用问题,其记录速率应该符合细胞分裂的时间尺度。第三,分子记录器应该具有长时间稳定性,其突变活性应该能够维持数周甚至数月的时间,能够在足够多的细胞分裂代数内保持稳定。第四,分子记录器应该具有足够的记录空间,其靶向区域应该允许足够多的突变产生并被记录,因为过短的靶向区域也会导致条形码同质性,缩短分子记录器的使用寿命。

基于DNA的分子记录器在发育生物学和肿瘤生物学的研究中都具有不可替代的重要作用和广阔的应用前景。这些记录器利用CRISPR-Cas9技术引入多样化且不可逆的基因组编辑,能够在细胞分裂过程中持续标记细胞及其后代。严格意义上说,基于分子记录器的谱系追踪技术属于前瞻性谱系追踪,因为其核心方式是在一开始就引入分子记录器,随着细胞分裂过程进行记录,可以随时通过测序进行信息捕捉和谱系重构,理论上,其谱系重构效率和时间精度远超传统的基于静态条形码或是体细胞突变的分析。这种高覆盖率的研究方式使得研究者能够在发育生物学和肿瘤生物学等领域实现细致分析,为相关领域的基础研究和临床应用提供了新的视角和工具,也使得研究者能够对复杂的发育过程进行系统的分析,从而更好地理解细胞在不同阶段如何与环境相互作用。

2 基于CRISPR-Cas9系统的活细胞分子记录器的发展脉络和代表性研究

2016年,CRISPR-Cas9系统被引入细胞谱系追踪领域,推动了这一领域的迅速发展。本节中,我们将对近年来的代表性研究进行系统性的介绍(图1、表1),重点关注Cas9分子记录器的技术发展和遇到的问题,同时,也对每项研究的体系和结论进行简单的介绍。

2.1 技术发展脉络

在上文提到的2016年的GESTALT方法中,研究人员在细胞培养和活体斑马鱼中生成了成千上

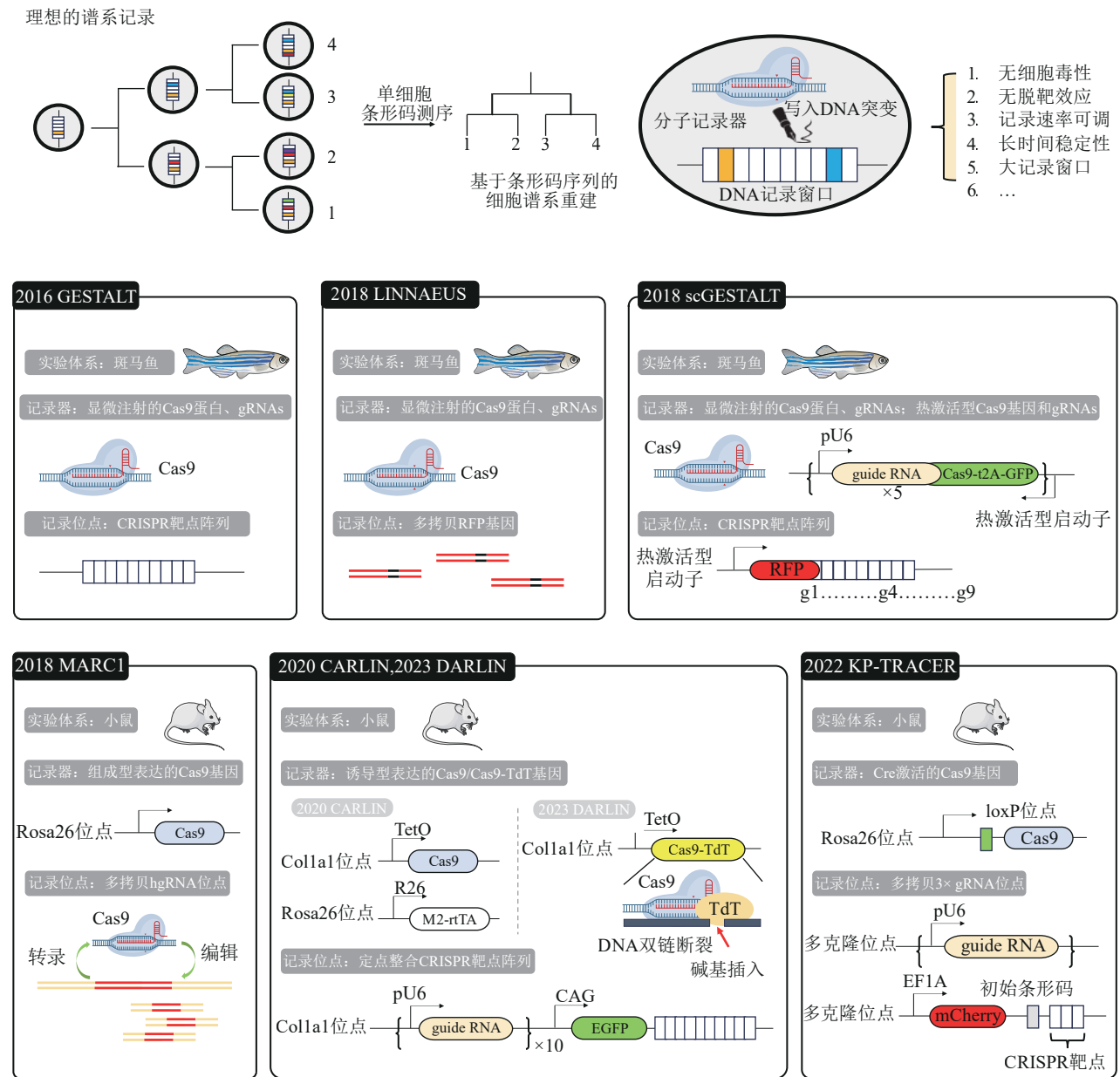


图1 理想的分子记录器^[42-44]和Cas9分子记录器的代表性研究^[41-42, 45-49]

[Rosa26 Coll1a1: 基因组上 landing pad 位点, 用于外源基因的定点整合。R26, pU6, CAG, EF1A: 组成型启动子。TetO: 诱导型启动子, 由rtTA激活。M2-rtTA: 诱导系统, 由诱导剂 doxycycline(dox) 激活]

Fig. 1 Ideal molecular recorders^[42-44] and representative studies of Cas9-based molecular recorders^[41-42, 45-49]

[Rosa26, Coll1a1: landing pad sites on genome, for site-specific integration of exogenous genes. R26, pU6, CAG, EF1A: constitutive promoters. TetO: inducible promoter, activated by rtTA. M2-rtTA: inducible system, activated by doxycycline(dox) inducer]

万的谱系信息条形码等位基因, 通过对单个斑马鱼进行数十万个细胞的采样, 发现大多数成体器官中的细胞来源于相对较少的胚胎前体。需要注意的是, 这项工作并没有采用单细胞测序的技术, 而是采取组织二代测序的方法, 在成鱼不同的组织处取样, 以此推断这些组织从胚胎发育而来的

分化发生时间的远近。因此, 这项工作的分辨率远没有达到单细胞的水平^[41]。

自2016年以来, 人们在应用这项新技术的同时, 也认识到其具有明显的缺陷, 其主要是由CRISPR-Cas9系统在细胞内产生DNA双链断裂的固有特点造成的。首先, Cas9引起的DNA条形码

表1 Cas9分子记录器的特点总结

Table 1 Summary of the characteristics of Cas9-based molecular recorders

名称	分子记录器	记录窗口大小/bp	靶点数量/site	主要突变类型	追踪时间 ^① /d	物种	是否稳定整合 ^②	研究领域	参考文献
GESTALT	Cas9	266	10	碱基删除	3	斑马鱼	否	发育	[41]
scGESTALT	Cas9	363	10	碱基删除	23~25	斑马鱼	否	发育	[47]
MARC1	Cas9	240	约20	碱基删除	19~21	小鼠	是	发育	[46]
LINNAEUS	Cas9	75	16~32	碱基删除	5	斑马鱼	否	发育	[45]
CHYRON	Cas9+TdT	100	5	碱基插入	22	HEK293T细胞系	是	—	[50]
CARLIN	Cas9	276	10	碱基删除	7	小鼠	是	发育	[48]
DARLIN	Cas9+TdT	828	30	碱基插入	7	小鼠	是	发育	[42]
2021年的一项研究	Cas9	—	约30	碱基删除	54	小鼠	是(仅肿瘤)	肿瘤	[51]
KP-TRACER	Cas9	—	30~90	碱基删除	150~180	小鼠	是(嵌合体)	肿瘤	[49]

① 追踪时间：参考文献中的最长实验时间。

② 是否稳定整合：是否将分子记录器整合到细胞基因组上。

① Tracing time: maximal experimental time in the cited studies.

② Stable integration: integrated the molecular recorders into the cell genomes.

主要以碱基缺失 (deletions) 形式存在, 缺失可能会导致条形码发生丢失或简并现象, 干扰谱系发生树的建立。其次, Cas9 的编辑效率并不能保证在一次分裂的时间内, 对每个细胞中都至少引入一个编辑, 也就会导致时间分辨率的下降。另外, 编辑窗口窄, 即可被编辑的位点少, 这也会导致条形码发生简并现象, 干扰谱系发生树的建立^[3] (图2)。因此, 在后续的研究中, 不同研究组纷纷开发了各种不同的方法来提高谱系追踪技术的编辑效率、时间分辨率和编辑稳定性。

增加 Cas9 靶点数量是一个较为常用的提高效率的方法, 因为其增加了分子记录器的记录窗口, 扩展了可能的编辑空间, 尽可能地保证分子记录器在每个细胞的每个分裂代时都能产生独特的条形码。例如, 2018 年的 “LINNAEUS” 技术利用预先在斑马鱼基因组中插入的十多个 RFP 基因作为 gRNA 的靶点, 提高了条形码的多样性^[45]。通过将单细胞 RNA 测序与报告基因的谱系条形码的计算分析相结合, 研究人员在斑马鱼幼体及成年鱼的心脏、肝脏、胰腺和大脑皮层中重建了发育谱系树^[45]。

除了增加靶点数量外, 靶点的位置也是一个需要考虑的重要因素, 因为哺乳动物细胞的基因组环境较为复杂, 基因组上下文 (context) 可能对 Cas9 的编辑效率产生潜在的影响。因此, 在 2018 年的 “MARC1” 技术中, 研究人员在小鼠基

因组中随机插入了 60 多个 gRNA 的靶点, 并且提出了 homing guide RNA (hgRNA) 的方法, 也就是让 gRNA 靶向自身序列, 这样就能够实现连续的编辑。研究人员将基因组上的不同位点的编辑效率进行标定和排序, 排除掉那些编辑过快或者过慢的位点, 以便能够筛选出适合于建立谱系发生树的位点^[46]。MARC1 分子记录器在受精时被激活, 并在妊娠期间持续进行诱变, 最终生成的发育条形码小鼠记录了谱系特异性的突变信息。研究人员利用这些记录可靠地重建了发育早期的谱系, 并研究了大脑的轴向发育^[46]。

为了提高时间分辨率, 研究人员采用多通道的分子记录器来标记追踪不同的发育阶段, 即在人为控制的时间点启动不同的分子记录器。例如, 在 2018 年的 “scGESTALT” 技术中, 研究人员提前将 gRNA 的表达框插入基因组, 并且采用热激活启动子控制其转录, 这样, 能够人为控制条形码编辑的时间窗口^[47]。该方法依赖一个可诱导表达系统, 使条形码可以在多个时间点被编辑, 从而捕获后期发育阶段的谱系信息。对约 60 000 个斑马鱼幼体大脑转录组的测序识别出了超过 100 种细胞类型和标记基因。利用这些数据, 生成了包含数百个分支的谱系树, 帮助揭示了细胞类型、大脑区域和基因表达级联反应在分化过程中的限制^[47]。

为了解决 Cas9 引起的过多缺失的问题, 一些研

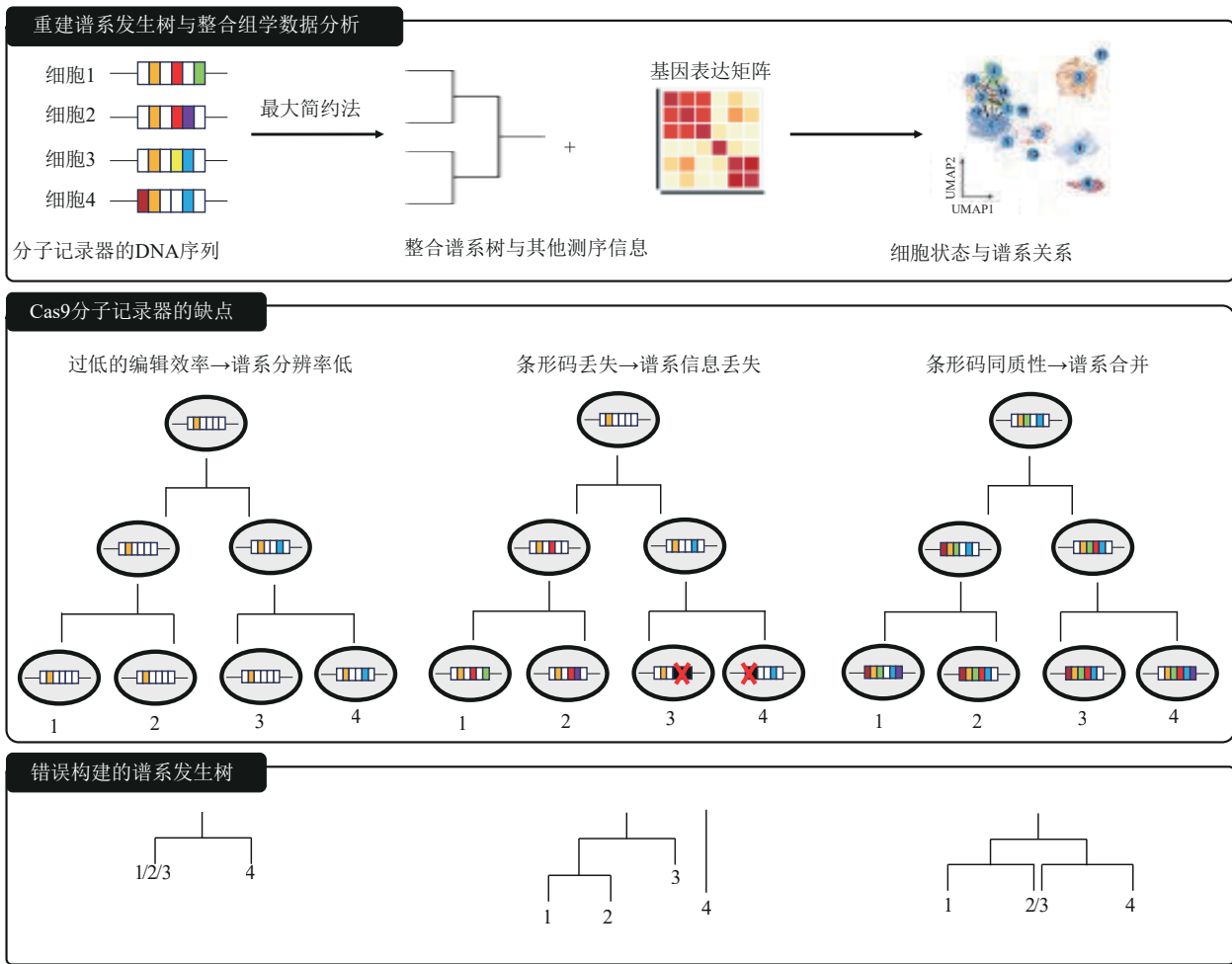


图2 重建谱系发生树与整合组学数据分析^[43, 49], Cas9分子记录器的缺点与错误构建的谱系发生树^[3, 42, 49]
Fig. 2 Reconstructed phylogeny trees and integrated omics data analysis^[43,49] as well as shortcomings of Cas9-based molecular recorders and errors in reconstructing phylogeny trees^[3,42,49]

究尝试改造原始的Cas9蛋白,降低缺失突变的发生率,并引入碱基插入(insertions)或替换(substitutions)突变。例如,在2021年报道的“CHYRON”技术中,为了实现随机DNA条形码的体内写入,研究人员将Cas9与非模板依赖的DNA聚合酶末端脱氧核苷酸转移酶(terminal deoxy nucleotidyl transferase, TdT)融合表达^[50]。在gRNA的引导下,TdT能够在双链断裂处随机插入碱基,降低了编辑区的缺失突变率。研究人员成功地将CHYRON应用于进化谱系追踪和记录用户选择的细胞刺激。然而,该研究只在HEK293T细胞中进行了测试,并没有进行活体动物的测试^[50]。

在应用基于Cas9的分子记录器进行活体动物的发育谱系追踪方面,已经有基于转基因鼠的研

究,并且综合了上述的各种改进策略。2020年,“CARLIN”小鼠诞生,研究人员在小鼠基因组中插入了Cas9的表达位点、gRNA的表达位点以及gRNA的靶点序列^[48]。其中,Cas9的表达由Tet-On系统控制,在诱导剂的作用下才会进行表达。gRNA的靶点沿用了编码绿色荧光蛋白的EGFP基因的3'端非编码区,也是10个不同的gRNA的靶点串联。该模型利用CRISPR技术以可诱导的方式生成多达44 000个转录条形码,适用于在发育或成体的任何时间点进行顺序条形码记录。研究人员利用CARLIN识别了胎肝造血干细胞(HSC)克隆活动的内在偏倚,并揭示了HSC在应对损伤时存在一个先前未被重视的克隆瓶颈(clonal bottleneck)。然而,“CARLIN”小鼠中的

碱基缺失编辑仍然占大多数^[48]。

到了2023年,研究人员改进了“CARLIN”小鼠,制作了新一代的转基因小鼠,即“DARLIN”小鼠。与“CARLIN”小鼠相比,“DARLIN”小鼠采用了Cas9融合TdT的策略,并在基因组上引入了3个gRNA靶点阵列,即30个gRNA靶点。从单细胞测序的结果来看,召回率已经达到了60%,即60%的单细胞数据具有能够进行分析的有效编辑,并且编辑的缺失率下降,平均每个细胞有3个缺失突变和2个插入突变^[42]。通过该模型,研究人员探讨了发育过程中造血干细胞的命运偏倚和迁移特征,并建立了联合转录组学和表观基因组学的单细胞测量方法,发现细胞克隆记忆(clonal memory)与全基因组DNA甲基化相关,而非由基因表达或染色质可及性所控制^[42]。DARLIN技术在一定程度上克服了现有谱系记录工具的局限性,能够在70%的单细胞中检测到编辑后的条形码,但是对于需要长期追踪的生物学问题仍有不足,例如肿瘤发生和转移的谱系追踪问题^[42]。

除了上述的转基因小鼠策略,在肿瘤研究中,一些情况下需要进行肿瘤移植或是诱导小鼠产生肿瘤,也有研究人员对Cas9分子记录器在肿瘤中的应用进行了优化,优化策略包括通过转染癌细胞或干细胞的实验来筛选合适的靶点数量的细胞,之后进行肿瘤移植或是嵌合体小鼠的培养^[49, 51]。

在肿瘤研究领域,Jonathan S. Weissman课题组有一系列的代表性研究,将基于Cas9的谱系追踪系统应用于肿瘤细胞的谱系分析,提供了关于肿瘤演变的新见解。在2021年,研究团队应用了一种基于Cas9的单细胞谱系追踪器,在小鼠肺癌异种移植模型中研究了转移的速率、路径和驱动因素^[51]。在该研究中,他们成功追踪并深度解析了数万癌细胞在数月的生长和扩散过程中的系统发育关系,揭示了转移能力的显著异质性。研究还发现,这些相关基因能够驱动癌细胞的侵袭性,揭示了KRT17在这一过程中的抑制作用^[51]。在2022年,他们研究了肿瘤进化的机制,特别是在小鼠模型中分析了Kras和Trp53驱动的肺腺癌的肿瘤演化过程。研究者通过在体外编辑了一种既能够被诱导成肿瘤、也能够被诱导激活基于Cas9的

分子记录器的mESC细胞,进而通过形成嵌合体小鼠,期望能够跟踪从单细胞发展成肿瘤乃至肿瘤转移的进程^[49]。研究人员观察到不同的肿瘤中可能有不同的亚克隆构成,并且联合单细胞转录组测序分析,发现细胞的异质性在部分肿瘤形成的过程中增加。此外,通过敲低抑癌基因,研究人员发现肿瘤的演化路径也会随之发生改变。最后,研究人员还通过空间转录组测序,验证了转移瘤可能起源于初始肿瘤的某个亚克隆部分^[49]。然而,在这项研究中,作者提到仍然有许多可能的肿瘤的中间状态无法捕捉到,对于每个肿瘤,他们建立的谱系追踪系统也仅仅回溯了平均数百个细胞。尤其是对于转移瘤来说,由于记录时间已经超过5个月,其内部的记录空间几乎已全部饱和,难以推断转移瘤内部的谱系结构^[49]。

2.2 谱系发生树的重建与多组学联合细胞谱系分析

细胞谱系数据解析是一个高度复杂且多层次的流程,涵盖了从数据预处理、亚克隆识别、谱系发生树构建到验证和解释的多个步骤,我们在这部分详细介绍如何用生物信息学手段从分子记录器产生的数据中重建谱系发生树,以及多组学数据(尤其是单细胞组学数据)如何与细胞谱系分析相结合,获得对生物学问题的更深层次理解(图2)。

首先,研究人员从实验样本中通过单细胞测序技术获得转录组数据、分子记录器编辑过的条形码DNA序列、DNA甲基化等数据,从条形码DNA序列中提取包含对Cas9靶点目标阵列序列的读取,识别出所有不同的编辑事件(称为等位基因)。研究人员使用最大简约法(maximum parsimony)构建谱系树^[43]。最大简约法是一种基于进化事件最小化的树构建方法,研究人员通过比较等位基因的编辑模式,寻找反映不同编辑事件的谱系发生树拓扑结构。从中,选择编辑事件最少的树作为最优树,这种树拓扑结构最有可能解释观察到的等位基因模式的条形码群体演化历史。生成的树包含由早期编辑事件定义的主要分支(clades),能够清晰地地区分大多数克隆群体^[43]。

在理论方面,一些计算工作探讨了靶点数(记录空间)、突变效率、传代时间和次数以及最终的细胞数等多种因素对最终重建谱系发生树的质量的影响^[52-54]。首先,增加靶点数是提高谱系重建精度的有效策略,更多的靶点能够提供更大的突变组合空间,从而支持更复杂的谱系树重建。有研究通过理论分析证明,记录位点数量需要与细胞分裂的最小次数相关,以确保在每次细胞分裂中积累足够的突变信息^[52]。其次,适当的突变效率可以确保在细胞分裂过程中积累足够的突变信息,而不会过早耗尽记录位点或导致突变信息不足^[53]。在此基础上,减少数据缺失和降低条形码同质性的影响也是优化谱系记录效果的重要方面^[53]。此外,传代时间、传代次数越多,对于突变效率的要求越高。最终细胞数越多,所需的记录位点数量也会越多,以支持更复杂的谱系树重建^[54]。在一项研究中,研究人员通过模拟验证了在细胞数约为65 000的情况下,记录位点数量与突变效率、传代时间等因素的关系^[52]。

在初步建立谱系树后,可以根据每个细胞携带的等位基因组合确定其克隆身份,并通过将单细胞RNA测序数据分配到其对应的克隆中,可以构建一个图结构,其中每个节点代表一个细胞,而边则连接共享等位基因的细胞^[43]。每个孤立的子图代表一个独立的克隆,这种图结构能够直观地展示细胞之间的谱系关系。为了验证树的准确性,研究人员还可以结合克隆内和克隆间细胞在基因表达、染色质可及性和DNA甲基化方面的相似程度,比较克隆内细胞的相似性与随机化克隆结构的相似性,进一步推断细胞之间的谱系关系,揭示克隆在分化过程中的命运偏倚和共享特征^[55-56]。

基于DNA的分子记录器与其他组学测序结合,允许研究者同时获取细胞的转录组信息和谱

系历史。这种多重数据的整合提供了细胞状态变化与谱系关系之间的深度洞察,使得科学家能够识别早期的命运偏向、细胞分化的关键时刻以及这些变化如何影响整个生物体的发育过程^[57-58],或揭示肿瘤细胞状态在进化过程中的具体路径。此外,这种大规模数据整合也为计算方法提供了便利,目前已经有一些计算研究在不断优化从多重数据重建谱系发生树的算法^[59-60]。

3 非Cas9依赖的新型分子记录器在细胞谱系追踪中的进展

基于Cas9分子记录器的细胞谱系追踪技术证明了进行高通量的细胞谱系的记录和回溯的可能性,同时,也能够高通量采集细胞的转录组等信息,将单细胞转录组数据与细胞的谱系数据联系起来,揭示更深层次的细胞分裂、分化的基因表达特征。然而,自2016年以来,尽管有许多研究对基于Cas9的分子记录器进行了各种优化,尝试克服其固有缺陷,但是无论在迄今为止最先进的“DARLIN”小鼠中,还是在肿瘤研究的嵌合体小鼠中,仍然存在着编辑效率低、条形码丢失和条形码同质性的问题。

进行体内的DNA分子记录的能力并不是CRISPR-Cas系统所独有的。自2020年左右开始,新的基因编辑技术得到广泛关注,目前已经产生了一些不同原理的基于DNA的分子记录器,都具有应用于细胞谱系追踪的前景,并有潜力克服基于Cas9的分子记录器的一些缺点。在本章中,我们将介绍三种有潜力的、基于不同于CRISPR-Cas原理的新型分子记录器(图3),并且将其优势、劣势与基于Cas9的分子记录器进行对比(表2)。

表2 新型分子记录器与Cas9分子记录器对比分析

Table 2 Comparative analysis of novel molecular recorders and the Cas9-based molecular recorder

分子记录器	稳定性	时序写入	特异性	编辑窗口大小	编辑效率	技术发展程度	参考文献
Cas9	易丢失	否	易脱靶	窄	—	高	[41-42,45-51]
先导编辑	高	是	易脱靶	窄	低于Cas9	中	[61,65-66]
DNA结合蛋白	高	否	强	宽	高于Cas9	中	[62,67-68]
Muta-T7	高	否	强	可调节	未知	低	[63-64,69-71]

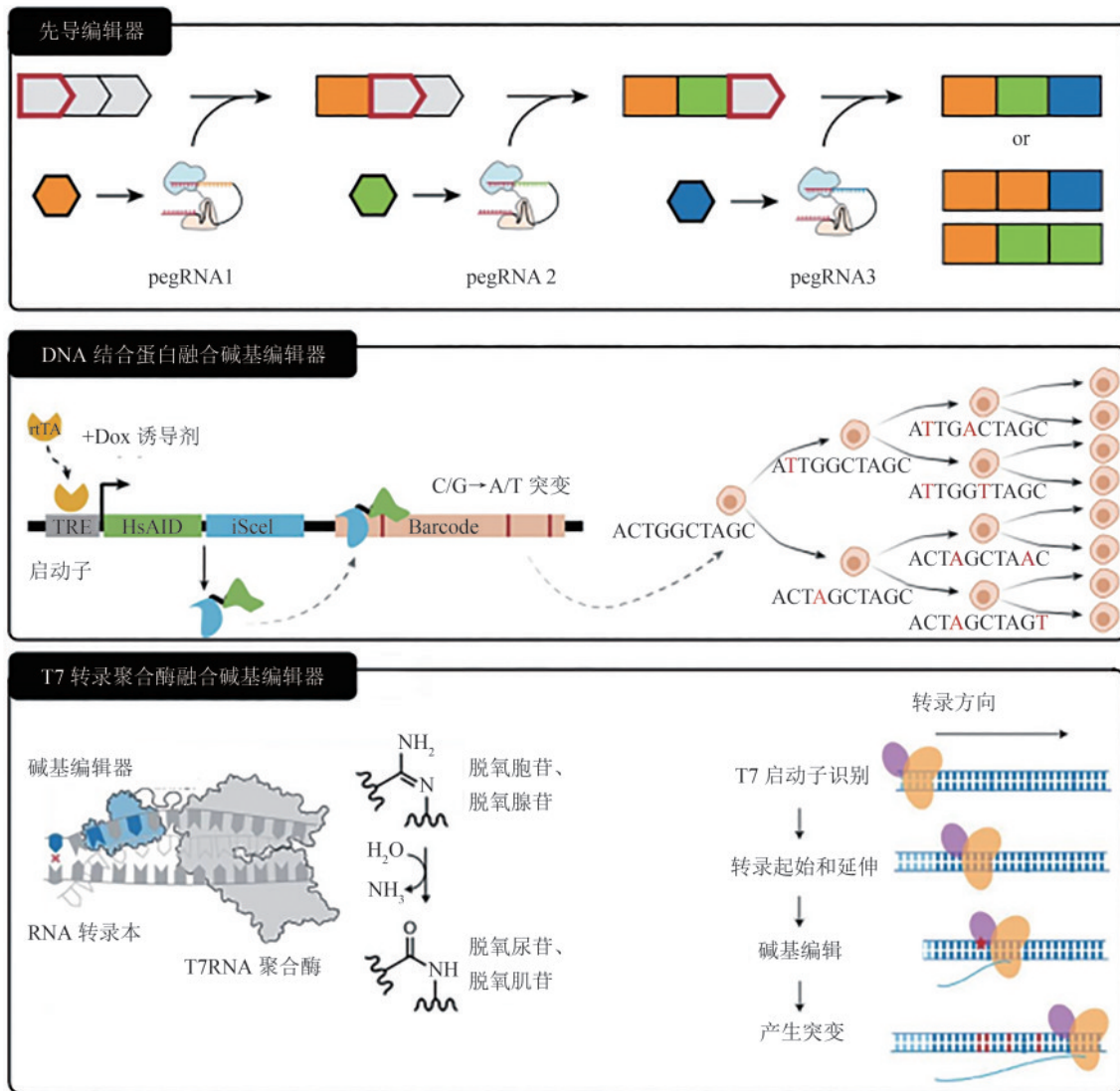


图3 三种新型分子记录器的工作原理^[61-64]

(pegRNA—先导编辑引导RNA; TRE—可在DOX存在时被rtTA激活的启动子;
HsAID—碱基编辑器; iSceI—DNA结合蛋白; rtTA—诱导系统)

Fig. 3 Working principles for three novel molecular recorders^[61-64]

(pegRNA—prime editing guide RNA; TRE—a promoter that can be activated by rtTA in the presence of DOX;
HsAID—base-editor; iSceI—DNA binding protein; rtTA—inducible system)

3.1 先导编辑器

先导编辑 (prime editing) 是一种新型基因组编辑技术, 通过替换目标DNA中的特定序列, 实现精准的基因修改。其基本原理是利用融合了反转录酶的nCas9 (nickng Cas9, 在靶位点引起单链DNA切口的Cas9变体蛋白), 结合特定的先导编辑引导RNA (prime editing guide RNA, pegRNA), 在指定位置诱导单链断裂。pegRNA不仅包含用于

反转录的编辑序列, 还包括结合序列, 从而实现精准的DNA替换、插入和删除, 无需引入双链断裂或供体DNA模板。这种机制确保了高效的基因组编辑, 克服了传统基因编辑中常见的旁观者编辑和脱靶效应等问题, 同时最大限度地减少了不必要的基因组干扰^[72-74]。

Jay Shendure 课题组将先导编辑用于分子记录器的开发和谱系追踪研究。该方法包括一个由部分CRISPR-Cas9靶位点组成的串联阵列, 通过短

序列插入编辑记录每次编辑的 pegRNA 身份, 同时将引导编辑位置向前移动。这一系统能够保持多达 20 个连续的序列事件的记录^[61] (图 3)。按序插入式先导编辑细胞历史记录 (prime editing cell history recording by ordered insertion, peCHYRON) 技术通过逐个插入带有独特条形码的 pegRNA, 在特定基因组位点插入 20 bp 的序列, 其中包含 3 bp 的突变作为条形码和 17 bp 的恒定序列。每次插入都使先前的结合序列失活, 从而实现事件的顺序记录^[65]。为进一步增强多重记录能力, 增强子驱动基因组多重转录活动系统 (enhancer-derived genomic recording of transcriptional activity in multiplex, ENGRAM) 方法将多个信号和特异性条形码整合进 pegRNA 中, 从而能够同时捕获多个基因的转录活性^[66]。

与原始的 Cas9 系统相比, 基于先导编辑的新型分子记录器具有一定的优势。首先, 先导编辑能够实现精准的基因编辑, 减少旁观者编辑和非靶向效应。与传统 CRISPR 技术相比, 先导编辑只需在目标位点产生单链断裂, 有效避免了双链断裂带来的不确定性, 使得细胞谱系的记录更加可靠。其次, 先导编辑在理论上具有更高的时间分辨率和抗干扰能力。在 Shendure 等^[65-66] 的设计中, 先导编辑器 (prime editor) 可以按时间顺序插入条形码, 作为对单纯的序列突变信息的补充, 从而提高时间分辨率, 并在一定程度上避免条形码丢失、测序错误等问题对谱系重建带来的影响。

当然, 作为一项新兴技术, 基于先导编辑的分子记录器在细胞谱系追踪中的应用仍然不够成熟, 可能具有一些潜在的风险。首先, 先导编辑的编辑效率在不同细胞类型中可能不均一, 某些细胞中的效率较低, 这可能导致谱系追踪结果的不可靠性, 尤其是在需要长期跟踪的实验中。其次, 设计和优化多个 pegRNA 是先导编辑应用中的一个重要挑战。每个 pegRNA 需要针对特定靶位点设计, 增加了实验的复杂性和技术要求, 可能导致实验结果不一致, 甚至脱靶效应。由于仍然采用 Cas9 蛋白作为核心元件, Cas9 系统固有的潜在的脱靶风险也是限制先导编辑技术应用的重要因素^[75-77]。

此外, 转基因沉默风险也是对几乎所有分子

记录器的一个重要挑战。转基因沉默是指尽管细胞内存在编码 DNA, 但转基因的表达随时间下调甚至很快丧失, 这为分子记录器在细胞内的长期应用带来了风险^[78]。转基因沉默与表观遗传调控密切相关, 如 DNA 甲基化^[79-81]、组蛋白修饰等^[82]。在活体细胞中, 转基因沉默可能影响先导编辑记录的持久性, 导致记录的信息逐渐丧失, 限制了长期跟踪的能力。这个风险也是绝大部分 DNA 分子记录器都可能会面临的挑战, 因为 DNA 分子记录器必须通过外源基因整合的方法引入, 包括前文所述的 Cas9 系统。目前, 已经有研究表明, 先导编辑系统在哺乳动物细胞基因组上的活性受染色质环境影响, 尤其是转录起始位点可能会较大程度地影响先导编辑效率^[83]。

3.2 DNA 结合蛋白融合碱基编辑器

利用 DNA 结合蛋白融合碱基编辑器进行目标 DNA 序列的编辑, 同时结合了 DNA 结合蛋白对其靶点的特异性和碱基编辑器的精准性。DNA 结合蛋白能够识别特定的 DNA 序列并结合, 从而定位编辑器到目标基因组区域。碱基编辑器 (如 ABE 或 CBE) 被 DNA 结合蛋白引导至目标位点, 改变 DNA 碱基对而不引入双链断裂。这种方法允许精确的碱基替换、插入或删除。

目前, 基于单碱基突变的细胞谱系追踪系统 (substitution mutation-aided lineage-tracing system, SMALT) 作为一种新兴的细胞谱系追踪系统, 在 2021 年和 2024 年的研究中得到了应用和验证。SMALT 系统将碱基编辑器和 iSceI DNA 结合蛋白相融合, 这个蛋白在之前已被证明可以和锌指蛋白联用, 提高基因编辑效率^[84]。

在发育生物学领域, 2021 年, 研究者通过 SMALT 技术成功在果蝇中使用一个 3000 bp 的合成条形码序列, 记录了超过 20 个突变, 显著提升了复杂多细胞生物细胞谱系的重建能力, 并成功构建了包含数千个内部节点的高质量细胞谱系树^[67]。2024 年, SMALT 技术被应用于斑马鱼。研究者重建了多个斑马鱼鳍细胞的细胞谱系树, 结果显示再生鳍细胞主要源自鳍的相同部位, 并通过多次从同一个体中采样生殖细胞, 揭示了生殖细胞群

体的稳定性及其早期分离特征^[68]。

除了上述两项在发育领域的研究外，同一研究团队也将SMALT体系用于肿瘤的谱系追踪。他们利用SMALT技术追踪小鼠的结直肠癌肿瘤谱系，在单细胞水平上探究了肠道癌变的起源和演化路径^[62]（图3）。通过分析条形码中的突变模式重建单细胞谱系树，研究人员揭示了大多数肿瘤（66.7%）具有多克隆起源，表明在炎症驱动的肠道肿瘤发生中，多个独立的细胞谱系的并行扩张是常见的^[62]。研究人员量化了每个肿瘤的创始祖细胞数量（ N_p ），发现单克隆肿瘤的 N_p 约为1，而多克隆肿瘤的 N_p 在2到33不等，表明早期肿瘤发生可能由多个独立的祖细胞驱动。对于肠道息肉的分析同样支持该结论^[68]。研究人员分析了小鼠肠道多个息肉的单细胞谱系，发现所有息肉都表现出多克隆起源，揭示了多克隆起源的复杂性^[62]。此外，研究人员比较了单克隆和多克隆肿瘤的突变负担，发现单克隆肿瘤积累了更多的基因组突变，其可能代表了肿瘤进展的晚期阶段^[62]。最后，通过谱系发生树估计祖细胞的时间，研究人员推断了息肉起始的克隆扩张时间，并结合小鼠肠上皮细胞的增殖速率，以此推测人类家族性腺瘤性息肉病（FAP）患者的肿瘤起始可能早在婴儿期就已发生^[62]。这些重要发现为结直肠癌的早期筛查和干预提供了新视角，指出在多克隆病变阶段介入的潜在机会，可能改善临床结果^[62]。

此外，还有另一项研究利用Cas9切口酶（nCas9）和碱基编辑器融合进行谱系追踪^[85]。这种方法也同样不会引起双链断裂，而是引起单链断裂和碱基替换。但是由于nCas9的切割位点较为固定，因此只能在较窄的范围内引起碱基编辑。

与基于Cas9的分子记录器相比，这项技术也具有独特优势。首先，DNA结合蛋白具有高度特异性，其引导能力增强了碱基编辑器在特定DNA位点的活性，减少了脱靶的风险，提高了编辑的特异性。其次，这项技术也具有灵活性和可扩展性。研究人员可以设计多种DNA结合蛋白-碱基编辑器融合体，以适应不同细胞类型或状态的调控需求，便于追踪多个细胞谱系^[62, 67-68]。目前已经有基于Cas9的碱基编辑器能够同时兼容2种碱基编辑蛋白，因此，将DNA结合蛋白与多种碱基编辑

蛋白融合，开发多样的分子记录器工具，以应对各种可能的复杂环境，将有利于更好地发挥这项技术在谱系追踪中的作用^[86-87]。另外，这项技术具有远超Cas9分子记录器的大范围的编辑窗口。DNA结合蛋白的结合位点长达数十个碱基，在范围内的碱基及其附近的碱基都有可能被编辑，通过串联多个位点，能够显著扩大编辑窗口的范围，提升谱系追踪的可靠性。

这项技术在具有广泛应用前景的同时，也具有潜在的风险。DNA结合蛋白的细胞类型依赖性是不可忽视的因素。在不同细胞类型中，不同的DNA结合蛋白的结合效率存在差异，导致融合蛋白在不同细胞中的活性不一致，需要进行额外优化^[62, 67-68]。此外，尽管提高了特异性和结合效率，作为融合蛋白的另一部分的碱基编辑器的编辑效率仍低于理想水平，可能需要进一步的技术改进。最后，外源性DNA结合蛋白和碱基编辑器可能引发免疫反应，影响其在细胞中的长期应用。因此，需要更多的实验证据来支持其长时间稳定性。

3.3 T7转录聚合酶融合碱基编辑器

T7家族的RNA聚合酶（RNAP），包括T7 RNAP、Sp6 RNAP、K1.5 RNAP、T3 RNAP等，是一类来自于噬菌体的RNA聚合酶^[88-89]。它们是一种长约800~900氨基酸的单体RNA聚合酶，能够在体内或体外以双链DNA为模板合成单链RNA，且具有产生较长的转录本的能力^[90-91]。这一类聚合酶的转录特异性依赖于其长度在20~25 bp左右的特异性启动子^[92-94]。

T7转录聚合酶融合碱基编辑器（Muta-T7）技术是在T7家族RNAP的基础上进行改造，通过将T7 RNAP与碱基编辑器蛋白融合表达，可以实现细胞内基因连续进化^[95-96]。当T7 RNAP对目的基因进行转录时，碱基编辑器可以对DNA进行诱变，达到高突变率^[63, 97-98]。目前，这项技术的研究主要集中在大肠杆菌的连续进化系统上。2018年以来，有课题组对此技术进行了深入研究，不断改进Muta-T7在大肠杆菌中的突变率^[63, 95-98]。该技术已被视为与EvolvR和OrthoRep相似的具有广泛

前景的体内连续进化技术^[69]。Muta-T7系统已经在植物细胞^[70]、酵母细胞^[71]和哺乳动物细胞^[64]中实现了初步应用,并且在哺乳动物细胞中被证实具有可调控的、多通道的转录能力^[99-101]。因此,Muta-T7有希望被用于哺乳动物的谱系追踪问题。

尽管尚无在细胞谱系追踪方面的应用,但通过在体内连续进化问题中的应用来看,Muta-T7系统相比于基于Cas9的分子记录器可能具有一定的优势。首先,Muta-T7系统仅由一个融合蛋白和一个靶向区域组成,仅需2个表达框,工具酶的总长度预计在6~7 kbp,靶向区域的长度可以在几百到几千碱基对的范围内调整^[63, 95-98]。而CRISPR-Cas9系统的工具酶需要表达Cas蛋白和gRNA,其总长度在5~6 kbp,且gRNA的靶向区域一般是10个gRNA靶点串联,长度在200 bp左右。因此,Muta-T7系统对细胞的负担与CRISPR系统相似,但提供了更大的编辑窗口,有助于在一定程度上避免条形码同质性的问题。其次,Muta-T7的突变效率可以通过突变启动子方便地调整,相较于CRISPR系统中依赖于gRNA序列的设计,这种方法显得更加灵活和可控^[62, 67-68, 81]。此外,T7 RNAP的高转录活性增加了单位时间内的编辑概率,进而增强了突变效率,有望提高条形码的多样性^[83]。在体内连续进化的研究中,Muta-T7能够在转录区域内引入大量突变,而不受其他序列的限制,从而实现持续的转录偶联的基因突变^[64, 70-71]。

然而,想要真正地将Muta-T7系统作为分子记录器应用于细胞谱系追踪,仍然有一些挑战性的问题需要解决。首先,Muta-T7的长时间基因组编辑效率仍未得到充分刻画。尽管有研究表明,T7 RNAP与碱基编辑器结合可以推动植物、酵母和哺乳动物细胞的连续进化,但T7 RNAP在哺乳动物细胞中的转录活性和突变效率尚未达到有效细胞谱系追踪所需的水平,改进空间仍然显著,也需要系统性研究来验证Muta-T7在哺乳动物细胞中实现稳定、高效和长期基因编辑的能力。此外,在活体动物的细胞谱系追踪中,Muta-T7可能在哺乳动物细胞中受到转基因沉默的影响。这一限制可能影响Muta-T7在长时间内(例如数周或数月)的应用,降低突变效率,从而影响谱系树构建的质量。目前尚无关于Muta-T7或T7 RNAP转录系

统在哺乳动物细胞基因组中的活性图谱研究,而高通量的平行报告系统实验已经被证明可以定量解析内源启动子^[102-104]和增强子^[105]在基因组环境中的活性。因此,这项技术也可以被应用于Muta-T7或T7 RNAP转录系统,解析其启动子的广泛适用性和长期稳定性,以及T7系统与基因组环境之间的潜在相互作用。

3.4 三种新型分子记录器的对比分析

先导编辑是一项正被快速开发的热门技术,除了谱系追踪领域,基因编辑领域也在对先导编辑的效率、对细胞的影响等方面进行探索和优化。并且,在上文的Jay Shendure等的设计中,时序写入是先导编辑在进行细胞谱系追踪应用时的一个独有优势。但是由于其基于Cas9,可能受其固有的脱靶效应影响,在使用时需要通过设计pegRNA进行优化。

DNA结合蛋白融合碱基编辑器的技术路线已经在多个物种中得到了验证,其特异性强、编辑效率高的特点使其能够较好地进行谱系追踪的工作。但是其细胞类型特异性的问题需要得到解决后才能易于进行多通道的记录,例如在不同时间点启动不同的DNA结合蛋白进行谱系记录。

Muta-T7在体内连续进化领域也已经得到了充分的发展,其编辑效率得到了一定的优化,其可调性、正交性也已经在哺乳动物细胞中得到了验证。但是有关其在哺乳动物细胞中的长期应用,以及其与基因组的相互作用等方面的研究仍然需要补充。

4 总结与展望

新一代细胞谱系追踪技术依托高通量测序与基于DNA的分子记录器的结合,已经能够大规模回溯并重建细胞谱系发生树,为理解发育和肿瘤发生等重要生物学过程提供了前所未有的时间分辨率和数据规模。自2016年以来,基于Cas9的分子记录器在发育和肿瘤领域取得了一系列重要发现。这项技术在近十年间取得了长足的进展,但仍存在一些固有的挑战,导致无法根据测序信息

回溯谱系。这些挑战包括过低的编辑效率导致的谱系分辨率低下、DNA双链断裂带来记录信息丢失、条形码同质性引起的谱系合并。为了应对这些挑战,科研人员对这项技术进行了优化,旨在通过提高编辑精确性、降低删除率以及优化条形码设计等手段,进一步提升技术的可靠性和实用性。

同时,谱系发生树的重建与多组学联合细胞谱系分析是一个复杂且多层次的过程。利用生物信息学方法从分子记录器生成的数据中重建谱系发生树,并整合多组学数据深入解析细胞发育和肿瘤进化等生物学问题。有理论研究探讨了靶点数量、突变效率和细胞数量等因素对谱系重建精度的影响,以及一些算法相关的研究探索如何更好地重建谱系发生树。基于DNA的分子记录器与多组学数据的整合,为理解细胞状态变化与谱系关系提供了强大的工具。

理论上,一个理想的、适合于细胞谱系追踪的分子记录器,应该具有无细胞毒性和简洁性、无脱靶效应、可调的记录效率、长时间稳定性以及足够的记录空间等特点。我们关注到有一些其他类型的基于DNA的分子记录器,具有被应用于细胞谱系追踪的前景。先导编辑、DNA结合蛋白融合碱基编辑器以及Muta-T7技术,作为新型的DNA分子记录器,在细胞谱系追踪领域展现出了突破性的应用潜力,有望克服传统Cas9基因编辑技术在该领域的局限。它们基于碱基编辑原理,能够避免DNA双链断裂,以碱基替换而非碱基缺失的形式写入信息。当然,这些新型的分子记录器的一些问题仍然需要更深入的研究来解决和优化。先导编辑器能够以时序性写入信息,但是脱靶效应仍是需要解决的问题。DNA结合蛋白融合碱基编辑器以高编辑效率、高特异性为优势,但是其可拓展性仍需进一步的探索。Muta-T7已在连续进化体系中取得成功,但是在哺乳动物内的研究较少。

这三种新型的DNA分子记录器在细胞谱系追踪领域潜在的应用,能够提高基因编辑的准确性和效率,为深入研究细胞分化、发育和疾病发生机制提供新的工具和方法。我们期待这些技术的发展成熟能够带领细胞谱系追踪技术迈向新的发

展阶段,为生命科学研究和医学治疗带来更广阔的前景。

参 考 文 献

- [1] KESTER L, VAN OUDENAARDEN A. Single-cell transcriptomics meets lineage tracing[J]. *Cell Stem Cell*, 2018, 23(2): 166-179.
- [2] WAGNER D E, KLEIN A M. Lineage tracing meets single-cell omics: opportunities and challenges[J]. *Nature Reviews Genetics*, 2020, 21(7): 410-427.
- [3] WEINREB C, RODRIGUEZ-FRATICELLI A, CAMARGO F D, et al. Lineage tracing on transcriptional landscapes links state to fate during differentiation[J]. *Science*, 2020, 367(6479): eaaw3381.
- [4] VANHORN S, MORRIS S A. Next-generation lineage tracing and fate mapping to interrogate development[J]. *Developmental Cell*, 2021, 56(1): 7-21.
- [5] SANKARAN V G, WEISSMAN J S, ZON L I. Cellular barcoding to decipher clonal dynamics in disease[J]. *Science*, 2022, 378(6616): eabm5874.
- [6] WANG Y Q, ZHANG X, WANG Z. Cellular barcoding: from developmental tracing to anti-tumor drug discovery[J]. *Cancer Letters*, 2023, 567: 216281.
- [7] FENG J, PUCELLA J N, JANG G, et al. Clonal lineage tracing reveals shared origin of conventional and plasmacytoid dendritic cells[J]. *Immunity*, 2022, 55(3): 405-422.e11.
- [8] VAN EGEREN D, ESCABI J, NGUYEN M, et al. Reconstructing the lineage histories and differentiation trajectories of individual cancer cells in myeloproliferative neoplasms[J]. *Cell Stem Cell*, 2021, 28(3): 514-523.e9.
- [9] AALAM S M M, NGUYEN L V, RITTING M L, et al. Clonal tracking in cancer and metastasis[J]. *Cancer Metastasis Reviews*, 2024, 43(2): 639-656.
- [10] PORTA-PARDO E, VALENCIA A, GODZIK A. Understanding oncogenicity of cancer driver genes and mutations in the cancer genomics era[J]. *FEBS Letters*, 2020, 594(24): 4233-4246.
- [11] KAKIUCHI N, OGAWA S. Clonal expansion in non-cancer tissues[J]. *Nature Reviews Cancer*, 2021, 21(4): 239-256.
- [12] SINKALA M. Mutational landscape of cancer-driver genes across human cancers[J]. *Scientific Reports*, 2023, 13(1): 12742.
- [13] TARKOWSKI A K. Mouse chimaeras developed from fused eggs[J]. *Nature*, 1961, 190: 857-860.
- [14] MINTZ B. Genetic mosaicism in adult mice of quadruparental lineage[J]. *Science*, 1965, 148(3674): 1232-1233.

- [15] LIVET J, WEISSMAN T A, KANG H, et al. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system[J]. *Nature*, 2007, 450(7166): 56-62.
- [16] SNIPPERT H J, VAN DER FLIER L G, SATO T, et al. Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells[J]. *Cell*, 2010, 143(1): 134-144.
- [17] WEISSMAN T A, PAN Y A. Brainbow: new resources and emerging biological applications for multicolor genetic labeling and analysis[J]. *Genetics*, 2015, 199(2): 293-306.
- [18] NADALIN F, MARZI M J, PIRRA PISCAZZI M, et al. Multi-omic lineage tracing predicts the transcriptional, epigenetic and genetic determinants of cancer evolution[J]. *Nature Communications*, 2024, 15(1): 7609.
- [19] OREN Y, TSABAR M, CUOCO M S, et al. Cycling cancer persister cells arise from lineages with distinct programs[J]. *Nature*, 2021, 596(7873): 576-582.
- [20] RENZ P F, GHOSH DASTIDER U, BAGHAI SAIN S, et al. *In vivo* single-cell CRISPR uncovers distinct TNF programmes in tumour evolution[J]. *Nature*, 2024, 632(8024): 419-428.
- [21] YU C N, MANNAN A M, YVONE G M, et al. High-throughput identification of genotype-specific cancer vulnerabilities in mixtures of barcoded tumor cell lines[J]. *Nature Biotechnology*, 2016, 34(4): 419-423.
- [22] CORSELLO S M, NAGARI R T, SPANGLER R D, et al. Discovering the anti-cancer potential of non-oncology drugs by systematic viability profiling[J]. *Nature Cancer*, 2020, 1(2): 235-248.
- [23] XIA Y F, JI X D, JANG I S, et al. Genetic and pharmacological interrogation of cancer vulnerability using a multiplexed cell line screening platform[J]. *Communications Biology*, 2021, 4(1): 834.
- [24] MARTÍNEZ-JIMÉNEZ F, MUIÑOS F, SENTÍS I, et al. A compendium of mutational cancer driver genes[J]. *Nature Reviews Cancer*, 2020, 20(10): 555-572.
- [25] PARK S Y, MALI N M, KIM R, et al. Clonal dynamics in early human embryogenesis inferred from somatic mutation[J]. *Nature*, 2021, 597(7876): 393-397.
- [26] SPENCER CHAPMAN M, RANZONI A M, MYERS B, et al. Lineage tracing of human development through somatic mutations[J]. *Nature*, 2021, 595(7865): 85-90.
- [27] COORENS T H H, MOORE L, ROBINSON P S, et al. Extensive phylogenies of human development inferred from somatic mutations[J]. *Nature*, 2021, 597(7876): 387-392.
- [28] SHETH R U, WANG H H. DNA-based memory devices for recording cellular events[J]. *Nature Reviews Genetics*, 2018, 19(11): 718-732.
- [29] JANG H, YIM S S. Toward DNA-based recording of biological processes[J]. *International Journal of Molecular Sciences*, 2024, 25(17): 9233.
- [30] SUN J L, RAMOS A, CHAPMAN B, et al. Clonal dynamics of native haematopoiesis[J]. *Nature*, 2014, 514(7522): 322-327.
- [31] FIGUERES-OÑATE M, SÁNCHEZ-GONZÁLEZ R, LÓPEZ-MASCARAQUE L. Deciphering neural heterogeneity through cell lineage tracing[J]. *Cellular and Molecular Life Sciences*, 2021, 78(5): 1971-1982.
- [32] WOODWORTH M B, GIRSKIS K M, WALSH C A. Building a lineage from single cells: genetic techniques for cell lineage tracking[J]. *Nature Reviews Genetics*, 2017, 18(4): 230-244.
- [33] PEI W K, FEYERABEND T B, RÖSSLER J, et al. Polylox barcoding reveals haematopoietic stem cell fates realized *in vivo*[J]. *Nature*, 2017, 548(7668): 456-460.
- [34] PEI W K, WANG X, RÖSSLER J, et al. Using Cre-recombinase-driven Polylox barcoding for *in vivo* fate mapping in mice[J]. *Nature Protocols*, 2019, 14(6): 1820-1840.
- [35] CHOW K K, BUDDER M W, GRANADOS A A, et al. Imaging cell lineage with a synthetic digital recording system[J]. *Science*, 2021, 372(6538): eabb3099.
- [36] WAGNER D E, WEINREB C, COLLINS Z M, et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo[J]. *Science*, 2018, 360(6392): 981-987.
- [37] CONG L, RAN F A, COX D, et al. Multiplex genome engineering using CRISPR/Cas systems[J]. *Science*, 2013, 339(6121): 819-823.
- [38] RAN F A, HSU P D, WRIGHT J, et al. Genome engineering using the CRISPR-Cas9 system[J]. *Nature Protocols*, 2013, 8(11): 2281-2308.
- [39] SANDER J D, JOUNG J K. CRISPR-Cas systems for editing, regulating and targeting genomes[J]. *Nature Biotechnology*, 2014, 32(4): 347-355.
- [40] WANG J Y, DOUDNA J A. CRISPR technology: a decade of genome editing is only the beginning[J]. *Science*, 2023, 379(6629): eadd8643.
- [41] MCKENNA A, FINDLAY G M, GAGNON J A, et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing[J]. *Science*, 2016, 353(6298): aaf7907.
- [42] LI L, BOWLING S, MCGEARY S E, et al. A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells[J]. *Cell*, 2023, 186(23): 5183-5199.e22.
- [43] CHAN M M, SMITH Z D, GROSSWENDT S, et al. Molecular recording of mammalian embryogenesis[J]. *Nature*, 2019, 570(7759): 77-82.
- [44] FARZADFARD F, LU T K. Emerging applications for DNA writers and molecular recorders[J]. *Science*, 2018, 361(6405): 870-875.

- [45] SPANJAARD B, HU B, MITIC N, et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars[J]. *Nature Biotechnology*, 2018, 36(5): 469-473.
- [46] KALHOR R, KALHOR K, MEJIA L, et al. Developmental barcoding of whole mouse *via* homing CRISPR[J]. *Science*, 2018, 361(6405): eaat9804.
- [47] RAJ B, WAGNER D E, MCKENNA A, et al. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain[J]. *Nature Biotechnology*, 2018, 36(5): 442-450.
- [48] BOWLING S, SRITHARAN D, OSORIO F G, et al. An engineered CRISPR-Cas9 mouse line for simultaneous readout of lineage histories and gene expression profiles in single cells [J]. *Cell*, 2020, 181(6): 1410-1422.e27.
- [49] YANG D, JONES M G, NARANJO S, et al. Lineage tracing reveals the phylogenetics, plasticity, and paths of tumor evolution[J]. *Cell*, 2022, 185(11): 1905-1923.e25.
- [50] LOVELESS T B, GROTT S J H, SCHECHTER M W, et al. Lineage tracing and analog recording in mammalian cells by single-site DNA writing[J]. *Nature Chemical Biology*, 2021, 17(6): 739-747.
- [51] QUINN J J, JONES M G, OKIMOTO R A, et al. Single-cell lineages reveal the rates, routes, and drivers of metastasis in cancer xenografts[J]. *Science*, 2021, 371(6532): eabc1944.
- [52] WANG R, ZHANG R, KHODAVERDIAN A, et al. Theoretical guarantees for phylogeny inference from single-cell lineage tracing[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2023, 120(12): e2203352120.
- [53] SALVADOR-MARTÍNEZ I, GRILLO M, AVEROF M, et al. Is it possible to reconstruct an accurate cell lineage using CRISPR recorders?[J]. *eLife*, 2019, 8: e40292.
- [54] FORROW A, SCHIEBINGER G. LineageOT is a unified framework for lineage tracing and trajectory inference[J]. *Nature Communications*, 2021, 12(1): 4940.
- [55] ZAFAR H, TZEN A, NAVIN N, et al. SiFit: inferring tumor trees from single-cell sequencing data under finite-sites models [J]. *Genome Biology*, 2017, 18(1): 178.
- [56] ZAFAR H, LIN C, BAR-JOSEPH Z. Single-cell lineage tracing by integrating CRISPR-Cas9 mutations with transcriptomic data [J]. *Nature Communications*, 2020, 11(1): 3055.
- [57] KIM I S. DNA barcoding technology for lineage recording and tracing to resolve cell fate determination[J]. *Cells*, 2023, 13(1): 27.
- [58] BARON C S, VAN OUDENAARDEN A. Unravelling cellular relationships during development and regeneration using genetic lineage tracing[J]. *Nature Reviews Molecular Cell Biology*, 2019, 20(12): 753-765.
- [59] CHEN M Y, FU R J, CHEN Y Q, et al. High-resolution, noninvasive single-cell lineage tracing in mice and humans based on DNA methylation epimutations[J]. *Nature Methods*, 2025, 22: 488-498.
- [60] JONES M G, KHODAVERDIAN A, QUINN J J, et al. Inference of single-cell phylogenies from lineage tracing data using Cassiopeia[J]. *Genome Biology*, 2020, 21(1): 92.
- [61] CHOI J, CHEN W, MINKINA A, et al. A time-resolved, multi-symbol molecular recorder *via* sequential genome editing[J]. *Nature*, 2022, 608(7921): 98-107.
- [62] LU Z L, MO S L, XIE D, et al. Polyclonal-to-monoclonal transition in colorectal precancerous evolution[J]. *Nature*, 2024, 636(8041): 233-240.
- [63] MENGISTE A A, MCDONALD J L, NGUYEN TRAN M T, et al. MutaT7^{GDE}: a single chimera for the targeted, balanced, efficient, and processive installation of all possible transition mutations *in vivo*[J]. *ACS Synthetic Biology*, 2024, 13(9): 2693-2701.
- [64] CHEN H Q, LIU S, PADULA S, et al. Efficient, continuous mutagenesis in human cells using a pseudo-random DNA editor [J]. *Nature Biotechnology*, 2020, 38(2): 165-168.
- [65] LIAO H N, CHOI J, SHENDURE J. Molecular recording using DNA typewriter[J]. *Nature Protocols*, 2024, 19(10): 2833-2862.
- [66] CHEN W, CHOI J, LI X Y, et al. Symbolic recording of signalling and *cis*-regulatory element activity to DNA[J]. *Nature*, 2024, 632(8027): 1073-1081.
- [67] LIU K H, DENG S J, YE C, et al. Mapping single-cell-resolution cell phylogeny reveals cell population dynamics during organ development[J]. *Nature Methods*, 2021, 18(12): 1506-1514.
- [68] LIU Z, ZENG H, XIANG H M, et al. Achieving single-cell-resolution lineage tracing in zebrafish by continuous barcoding mutations during embryogenesis[J]. *Journal of Genetics and Genomics*, 2024, 51(9): 947-956.
- [69] MOLINA R S, RIX G, MENGISTE A A, et al. *In vivo* hypermutation and continuous evolution[J]. *Nature Reviews Methods Primers*, 2022, 2: 37.
- [70] BUTT H, RAMIREZ J L M, MAHFOUZ M. Synthetic evolution of herbicide resistance using a T7 RNAP-based random DNA base editor[J]. *Life Science Alliance*, 2022, 5(12): e202201538.
- [71] CRAVENS A, JAMIL O K, KONG D Z, et al. Polymerase-guided base editing enables *in vivo* mutagenesis and rapid protein engineering[J]. *Nature Communications*, 2021, 12(1): 1579.
- [72] ANZALONE A V, RANDOLPH P B, DAVIS J R, et al. Search-and-replace genome editing without double-strand breaks or

- donor DNA[J]. *Nature*, 2019, 576(7785): 149-157.
- [73] CHEN P J, LIU D R. Prime editing for precise and highly versatile genome manipulation[J]. *Nature Reviews Genetics*, 2023, 24(3): 161-177.
- [74] NELSON J W, RANDOLPH P B, SHEN S P, et al. Engineered pegRNAs improve prime editing efficiency[J]. *Nature Biotechnology*, 2022, 40(3): 402-410.
- [75] KOCAK D D, JOSEPHS E A, BHANDARKAR V, et al. Increasing the specificity of CRISPR systems with engineered RNA secondary structures[J]. *Nature Biotechnology*, 2019, 37(6): 657-666.
- [76] COELHO M A, DE BRAEKELEER E, FIRTH M, et al. CRISPR GUARD protects off-target sites from Cas9 nuclease activity using short guide RNAs[J]. *Nature Communications*, 2020, 11(1): 4132.
- [77] LI A, MITSUNOBU H, YOSHIOKA S, et al. Cytosine base editing systems with minimized off-target effect and molecular size[J]. *Nature Communications*, 2022, 13(1): 4531.
- [78] CABRERA A, EDELSTEIN H I, GLYKOFRYDIS F, et al. The sound of silence: transgene silencing in mammalian cell engineering[J]. *Cell Systems*, 2022, 13(12): 950-973.
- [79] ALLSHIRE R C, MADHANI H D. Ten principles of heterochromatin formation and function[J]. *Nature Reviews Molecular Cell Biology*, 2018, 19(4): 229-244.
- [80] ALHAJI S Y, NGAI S C, ABDULLAH S. Silencing of transgene expression in mammalian cells by DNA methylation and histone modifications in gene therapy perspective[J]. *Biotechnology & Genetic Engineering Reviews*, 2019, 35(1): 1-25.
- [81] LYKO F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation[J]. *Nature Reviews Genetics*, 2018, 19(2): 81-92.
- [82] HUGHES A L, KELLEY J R, KLOSE R J. Understanding the interplay between CpG island-associated gene promoters and H3K4 methylation[J]. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms*, 2020, 1863(8): 194567.
- [83] LI X Y, CHEN W, MARTIN B K, et al. Chromatin context-dependent regulation and epigenetic manipulation of prime editing[J]. *Cell*, 2024, 187(10): 2411-2427.e25.
- [84] FONFARA I, CURTH U, PINGOUD A, et al. Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases[J]. *Nucleic Acids Research*, 2012, 40(2): 847-860.
- [85] HWANG B, LEE W, YUM S Y, et al. Lineage tracing using a Cas9-deaminase barcoding system targeting endogenous L1 elements[J]. *Nature Communications*, 2019, 10(1): 1234.
- [86] ZHANG X H, ZHU B Y, CHEN L, et al. Dual base editor catalyzes both cytosine and adenine base conversions in human cells[J]. *Nature Biotechnology*, 2020, 38(7): 856-860.
- [87] YE L J, ZHAO D D, LI J, et al. Glycosylase-based base editors for efficient T-to-G and C-to-G editing in mammalian cells[J]. *Nature Biotechnology*, 2024, 42(10): 1538-1547.
- [88] CHEN Z H, SCHNEIDER T D. Information theory based T7-like promoter models: classification of bacteriophages and differential evolution of promoters and their polymerases[J]. *Nucleic Acids Research*, 2005, 33(19): 6172-6187.
- [89] DIETZ A, WEISSER H J, KÖSSEL H, et al. The gene for *Klebsiella* bacteriophage K11 RNA polymerase: sequence and comparison with the homologous genes of phages T7, T3, and SP6[J]. *Molecular & General Genetics*, 1990, 221(2): 283-286.
- [90] WANG W Y, LI Y, WANG Y Q, et al. Bacteriophage T7 transcription system: an enabling tool in synthetic biology[J]. *Biotechnology Advances*, 2018, 36(8): 2129-2137.
- [91] RIO D C. Expression and purification of active recombinant T7 RNA polymerase from *E. coli*[J]. *Cold Spring Harbor Protocols*, 2013, 2013(11): pdb.prot078527.
- [92] LEE S S, KANG C. Two base pairs at-9 and-8 distinguish between the bacteriophage T7 and SP6 promoters[J]. *The Journal of Biological Chemistry*, 1993, 268(26): 19299-19304.
- [93] IMBURGIO D, RONG M, MA K, et al. Studies of promoter recognition and start site selection by T7 RNA polymerase using a comprehensive collection of promoter variants[J]. *Biochemistry*, 2000, 39(34): 10419-10430.
- [94] ZONG Y Q, ZHANG H M, LYU C, et al. Insulated transcriptional elements enable precise design of genetic circuits[J]. *Nature Communications*, 2017, 8(1): 52.
- [95] MOORE C L, PAPA L J 3RD, SHOULDERS M D. A processive protein chimera introduces mutations across defined DNA regions *in vivo*[J]. *Journal of the American Chemical Society*, 2018, 140(37): 11560-11564.
- [96] PARK H J, KIM S H. Gene-specific mutagenesis enables rapid continuous evolution of enzymes *in vivo*[J]. *Nucleic Acids Research*, 2021, 49(6): e32.
- [97] SEO D J, KOH B H, EOM G E, et al. A dual gene-specific mutator system installs all transition mutations at similar frequencies *in vivo*[J]. *Nucleic Acids Research*, 2023, 51(10): e59.
- [98] MENGISTE A A, WILSON R H, WEISSMAN R F, et al. Expanded MutaT7 toolkit efficiently and simultaneously accesses all possible transition mutations in bacteria[J]. *Nucleic Acids Research*, 2023, 51(6): e31.
- [99] DIONISI S, PIERA K, BAUMSCHLAGER A, et al. Implementation of a novel optogenetic tool in mammalian cells based on a split T7 RNA polymerase[J]. *ACS Synthetic Biology*, 2022, 11(8): 2650-2661.
- [100] GHADERI M, SABAHI F, SADEGHI-ZADEH M, et al.

- Construction of an eGFP expression plasmid under control of T7 promoter and IRES sequence for assay of T7 RNA polymerase activity in mammalian cell lines[J]. Iranian Journal of Cancer Prevention, 2014, 7(3): 137-141.
- [101] QIN C R, XIANG Y H, LIU J, et al. Precise programming of multigene expression stoichiometry in mammalian cells by a modular and programmable transcriptional system[J]. Nature Communications, 2023, 14(1): 1500.
- [102] AKHTAR W, DE JONG J, PINDYURIN A V, et al. Chromatin position effects assayed by thousands of reporters integrated in parallel[J]. Cell, 2013, 154(4): 914-927.
- [103] VANHILLE L, GRIFFON A, MAQBOOL M A, et al. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq[J]. Nature Communications, 2015, 6: 6905.
- [104] VAN ARENSBERGEN J, FITZPATRICK V D, DE HAAS M, et al. Genome-wide mapping of autonomous promoter activity in human cells[J]. Nature Biotechnology, 2017, 35(2): 145-153.
- [105] ANDERSSON R, SANDELIN A. Determinants of enhancer

and promoter activities of regulatory elements[J]. Nature Reviews Genetics, 2020, 21(2): 71-87.



通讯作者: 钱珑(1985—),女,副研究员。研究方向为DNA分子信息存储、合成生物学元件的大数据挖掘与设计、合成基因线路的可预测设计、基因组科学和基因组进化。

E-mail: long.qian@pku.edu.cn



第一作者: 姜百翼(2000—),男,博士研究生。研究方向为哺乳动物细胞合成生物学。

E-mail: 2201111972@stu.pku.edu.cn