

# 基于 CRISPR-Cas9 的功能基因筛选研究进展

李欢欢, 黄承浩

厦门大学生命科学学院 分子疫苗学和分子诊断学国家重点实验室 国家传染病诊断试剂与疫苗工程技术研究中心, 福建 厦门 361102

李欢欢, 黄承浩. 基于 CRISPR-Cas9 的功能基因筛选研究进展. 生物工程学报, 2018, 34(4): 461-472.

Li HH, Huang CH. Functional genetic screening using CRISPR-Cas9 system. Chin J Biotech, 2018, 34(4): 461-472.

**摘 要:** 功能性基因的筛选是探索生物进程、研究疾病发生发展和诠释基因功能的重要方法, 在生物、医药、新治疗靶点筛选及肿瘤耐药等方面有广泛的应用。CRISPR-Cas9 (Clustered regularly interspaced short palindromic repeat sequences/CRISPR-associated protein 9) 技术作为近期热门的基因编辑工具, 能够高通量地对基因组进行精准修饰, 为实现功能性基因的筛选提供了简便高效的技术支持。文中对 CRISPR-cas9 技术应用于功能性基因筛选的方法及研究进展进行了综述。

**关键词:** CRISPR-Cas9, 基因敲除文库, 功能性基因筛选

## Functional genetic screening using CRISPR-Cas9 system

Huanhuan Li, and Chenghao Huang

National Institute of Diagnostics and Vaccine Development in Infectious Diseases, State Key Laboratory of Molecular Vaccinology and Molecular Diagnostics, School of Life Sciences, Xiamen University, Xiamen 361102, Fujian, China

**Abstract:** Functional genetic screening as an important method for exploring biological processes, diseases development research and functional annotation of genetic elements, has been widely used in pharmaceutical research, new therapeutic targets identifying and screening, and tumor resistance. CRISPR-Cas9 (Clustered regularly interspaced short palindromic repeat sequences/CRISPR-associated protein 9) is the newest tool in the geneticist's toolbox, allowing researchers to edit genome with unprecedented ease, accuracy and high-throughput. CRISPR-Cas9 system provides a high-throughput, practical and efficient tool for the discovery of functionally important genes responsible for certain phenotypes. In this review, we summarize the characterization of CRISPR/Cas9 system and applications of this new genetic toolkit in functional genetic screening.

**Keywords:** CRISPR/Cas9, genome-scale knockout library, functional genetic screening

**Received:** September 6, 2017; **Accepted:** October 30, 2017

**Supported by:** National Natural Science Foundation of China (No. 81571990).

**Corresponding author:** Chenghao Huang. E-mail: huangchenghao@xmu.edu.cn

国家自然科学基金 (No. 81571990) 资助。

网络出版时间: 2017-11-07

网络出版地址: <http://kns.cnki.net/kcms/detail/11.1998.Q.20171107.1703.003.html>

随着科学技术的发展,人类基因组计划初步完成,人们对基因的功能也有了一定的认识,但是仍然有更多的问题亟待探索,例如基因与疾病发生发展的关系,基因与肿瘤耐药性的关系等。如何利用人类基因组以及已有的技术手段研究疾病发生发展过程,寻找与疾病和健康相关的功能性基因,从而建立新的疾病诊断预防和治疗方法,发现新的治疗靶点及开发新的药物,是科学家们关注的热点。高通量功能性基因筛选以基因编辑技术为基础,通过在全基因组范围内干扰基因功能来筛选目的表型,是剖析基因功能、探索生物进程和疾病的重要工具。近年来,随着基因编辑技术的发展和成熟,功能性基因筛选也随之变得更加简便高效,在生命科学研究中发挥了重要作用。本文主要介绍应用 CRISPR-Cas9 文库进行功能性基因筛选的方法及目前的研究进展。

筛选功能性基因的常用策略主要分为功能获得型 (Gain-of-function) 筛选和功能缺失型 (Loss-of-function) 筛选。功能获得型筛选策略通常利用 cDNA 文库使基因过表达来筛选功能性基因<sup>[1]</sup>。由于细胞转录组非常复杂,难以构建覆盖整个细胞基因组的 cDNA 文库,导致 cDNA 文库的库容量有限;此外,cDNA 的表达并不受细胞内在调控,因此经常会出现基因异常高表达,影响对结果的判断,这些都限制了 cDNA 文库在功能性基因筛选中的应用。功能缺失型 (Loss-of-function) 筛选策略是通过抑制基因表达或基因敲除来筛选功能性基因。RNAi 是真核生物细胞内广泛存在的生物进程,通过 siRNA 介导其同源 RNA 降解实现抑制靶基因表达的作用<sup>[2]</sup>,因而曾被广泛应用于高通量的功能缺失型基因筛选<sup>[3]</sup>。与 cDNA 文库相比,RNAi 文库筛选具有高通量、简便、经济、高效的优点,但是其仍然存在难以忽视的问题。RNAi 技术是降低基因表达水平来影响表型,并不

能完全抑制基因表达,因此其造成的表型变化不够稳定明显,影响对基因功能的判定<sup>[4]</sup>。此外,细胞内源性的 RNAi 途径导致 RNAi 筛选存在着广泛的脱靶效应<sup>[5]</sup>,成为影响 RNAi 文库筛选的另一个重要问题。近年来 CRISPR-Cas9 技术大放异彩,可以用于酵母菌<sup>[6]</sup>、哺乳动物细胞<sup>[7]</sup>、斑马鱼<sup>[8]</sup>、水稻<sup>[9]</sup>、小鼠<sup>[10]</sup>等各类细胞及生物体的基因编辑,科学家们也开始利用 CRISPR-Cas9 技术进行功能性基因的筛选。CRISPR-Cas9 技术既可通过诱导基因突变进行功能缺失型 (Loss-of-function) 筛选,又可通过激活转录进行功能获得型 (Gain-of-function) 筛选,不仅可以作用于基因的编码区,还可以靶向基因非编码区<sup>[11]</sup>,与 RNAi 技术相比具有相对较低的脱靶效率和更广泛的作用位点,因而逐渐成为功能性基因筛选的热点方法。

## 1 CRISPR-Cas9: 精准的基因编辑技术

CRISPR-Cas 是广泛存在于细菌和古菌中的获得性免疫系统,能够剪切外源基因,抵御病毒对细菌的入侵<sup>[12]</sup>,最早由 Ishino 等发现于 1987 年<sup>[13]</sup>,直到 2010 年 CRISPR 的机制和功能才被研究清楚<sup>[14-16]</sup>。目前被发现的 CRISPR 系统有 3 种类型,均由 3 种元件组成:一簇 CRISPR 相关基因 (Cas)、非编码 RNA 以及一系列重复序列<sup>[17]</sup>。Cas 基因可以和核酸酶、解旋酶、聚合酶结合,是 CRISPR-Cas 系统中执行剪切功能的元件<sup>[18]</sup>。一系列短重复序列被来源于外源基因的间隔序列 (Protospacer) 隔开,共同构成了 CRISPR RNA (crRNA) 阵列。通常间隔序列与前间区序列邻近基序 (PAM) 相连<sup>[19]</sup>。不同物种 PAM 序列不同,例如目前在哺乳动物细胞中应用广泛的 spCas9 识别 5'-NGG PAM 序列,而在细胞基因组中平均 8-12 bp 就有这样的序列出现<sup>[20-21]</sup>。基因组中广泛存在的 PAM 序列为 CRISPR-Cas9 基因编辑技术应用于全基因

组提供了可能。基因座上游是一段几百 bp 的非编码基因,被命名为前导序列<sup>[18]</sup>。CRISPR-Cas 系统中最常用于基因编辑的是 II 型 CRISPR-Cas 系统,由 Cas9 核酸酶、crRNA 和反式激活 crRNA (tracrRNA) 三部分组成<sup>[22]</sup>。执行基因剪切功能时,首先是 crRNA 转录为 pre-crRNA,同时与 crRNA 互补的 tracrRNA 也进行转录并激活 Cas9 及特异性的 RNA 核酸酶对 pre-crRNA 进行加工,成熟的 crRNA 与 tracrRNA 及 Cas9 核酸酶形成复合体,进而在向导 RNA 的指导下靶向特定位点进行切割<sup>[23]</sup>。研究者们将密码子优化的 Cas9 和必要的 RNA 元件在哺乳动物细胞中进行异源重组表达<sup>[24]</sup>,并将 crRNA 和 tracrRNA 融合表达形成一条嵌合单链向导 RNA (sgRNA)<sup>[25]</sup>,精简了 CRISPR-Cas9 系统的结构,极大地提高了 CRISPR-Cas9 系统应用于基因编辑的可行性,使其更加便捷高效。在 CRISPR-Cas9 系统中,sgRNA 和紧接着 sgRNA 的 PAM 是决定基因编辑准确性的关键因素。sgRNA 长 20 nt,通过与靶序列互补配对引导 Cas9 准确定位于靶基因,并在 PAM 上游约 3 bp 处进行 DNA 双链剪切<sup>[17]</sup>。断裂双链 (DSBs) 的 DNA 损伤修复途径主要有以下两种:缺少修复模板时,断裂双链通过非同源末端连接 (Nonhomologous end joining, NHEJ) 途径重新连接,这种过程容易产生插入或缺失突变使基因功能丧失,因而被应用于基因敲除;当存在同源修复模板时,断裂双链通过同源直接修复 (Homology-directed repair, HDR) 途径将修复模板重组进断裂部位,常被用于基因重组<sup>[20]</sup>。CRISPR-Cas9 技术除了可以进行基因敲除,还可以用于基因的上调或下调表达。CRISPR 干扰 (CRISPR interference, CRISPRi) 和 RNAi 功能相似,是在 CRISPR 技术基础上改造而来的,将 Cas9 突变使其丧失活性 (dead Cas9, dCas9) 无法对双链 DNA 进行切割,再与转录抑制因子联合作用,即可在 sgRNA 指导下抑制特定

基因的表达<sup>[26]</sup>。CRISPR 激活 (CRISPR activation, CRISPRa) 是利用 dCas9 和转录激活因子协作上调靶基因表达水平,可以被应用于功能获得型筛选<sup>[27]</sup>。

## 2 CRISPR-Cas9 基因文库

### 2.1 基因文库设计

目前研究者们已对如何将 CRISPR-Cas9 系统应用于功能性基因筛选进行了多种尝试并取得了成功 (表 1),为后来者提供了大量的经验和指导。CRISPR-Cas9 功能性基因筛选平台首先需要能够满足实验者需求的 CRISPR-Cas9 文库。有效的 sgRNA 是决定 CRISPR-Cas9 功能性基因筛选成功的关键因素,因此首先要设计针对全基因组的 sgRNA。设计 sgRNA 要考虑两点:一是 Cas9 作用必需的 PAM 序列,二是降低脱靶效率<sup>[17]</sup>。研究者们开发了多种用于设计 sgRNA 的平台。Chuai 等<sup>[28]</sup>总结了当前的 36 个设计 sgRNA 的网站,并根据平台设计规则将其分为 3 类:1) Alignment-based: 单纯依据 PAM 位置设计及评分;2) Hypothesis-driven: 依据 GC 含量、外显子位置等因素对 sgRNA 效率的作用设计及评分;3) Learning-based: 利用考虑影响 sgRNA 效率的不同特点的训练模式设计及评分。其中第二、三类平台在设计时考虑了不同序列和染色质特点,相比第一类平台更加全面高效,使用者可根据不同要求选择合适的平台。设计好的 sgRNA 合成后需要连接至合适的慢病毒载体上,Cas9 核酸酶既可以和 sgRNA 构建在同一个质粒上<sup>[29]</sup>,也可以构建在两个质粒上<sup>[30]</sup>。将构建好的 sgRNA 文库包装成慢病毒,并以低 MOI (通常为 0.3–0.5) 感染细胞构建稳定细胞株<sup>[29]</sup>,避免同一细胞多个基因同时被敲除对结果造成影响。

### 2.2 常用筛选方式

CRISPR-Cas9 文库高通量筛选常用模式可以

表 1 CRISPR-Cas9 系统在功能基因筛选中的应用

Table 1 Functional genetic screening using CRISPR-Cas9 system

Species	sgRNAs	Target genes	sgRNAs/gene	Type of perturbation	Cell type(s)	Selection	Reference
Human	64 751	10 808	3–4	KO	A375	Resistance to Vemurafenib	[9]
Human	73 151	7 114	10	KO	KBM7	Resistance to 6-thioguanine	[34]
Human	869	291	2–3	KO	HeLa	Resistance to diphtheria	[35]
						Resistance to chimaeric anthrax	
Human	91 320	17 232	6	KO	HPAF- II	Proliferation	[36]
Human	23 652	73	3	KO	HELA A549	Proliferation	[37]
					293T		
Mouse	67 405	20 611	3	KO	NSCLC	Proliferation	[38]
Human	77 406	20 121	3–4	KO	293FT	Resistance to WNV	[39]
Human	187 536	18 543	10	KO	GXRCas9	Resistance to HIV	[40]
Human	70 290	23 430	3	Activation	A549	Resistance to IAV	[41]
Human	18 315	NF1: 6 682 NF2: 6 934		KO	A375	Resistance to Vemurafenib	[43]
		CUL3: 4 699					
Human	12 472	671	20	KO	Huh7.5	Proliferation	[44]
Human	HBE1	281	50	Activation	K562	Protein expression	[45]
	10 739	433	30				
	HER2						
	12 189						
Human	123 411	19 050	6	KO	CEM T	Antigen antibody interactions	[48]
					Raji/CD4 B		
Mouse	87 897	19 150	1–5	KO	iBMDMs	GSDMD	[49]

分为阳性筛选 (Positive selection screen) 和阴性筛选 (Negative selection screen)<sup>[30]</sup>。阳性筛选是对已成功整合 sgRNA 的细胞文库施加一定的筛选压力, 仅使少数目的表型的细胞能够存活, 达到富集关键基因的目的。阴性筛选与之相反, 存活的细胞并不是目的表型细胞, 因此需要比较不同时间点 sgRNA 的丰度找出差异 sgRNA 来确定关键基因。两种筛选方式通常都需要与高通量测序技术结合, 通过高通量测序获得样品中细胞的 sgRNA 序列信息, 再通过生物信息学分析排除假阳性结果继而确定可能的候选基因 (图 1)。CRISPR 系统后续分析同样是影响实验结果的重要问题, Chuai 等<sup>[28]</sup>同样整理了 7 种用于后续结

果分析的平台。例如 MAGeCK 是 Li 等<sup>[31]</sup>开发的用来分析 CRISPR 高通量筛选结果的算法, 之后 Li 等又将其拓展为 MAGeCK-VISPR 算法<sup>[32]</sup>, 为 CRISPR 高通量筛选提供了综合质控、分析及可视化的数据分析方法。

### 3 CRISPR-Cas9 文库应用于功能性基因筛选的研究进展

#### 3.1 药物作用基因筛选

CRISPR-Cas9 文库早期应用于药物作用基因筛选的尝试较多。Vemurafenib 是突变的蛋白激酶 BRAF 的抑制剂, 被批准用于晚期黑色素瘤的治疗, 对存在 V600E BRAF 突变的黑色素瘤有治疗

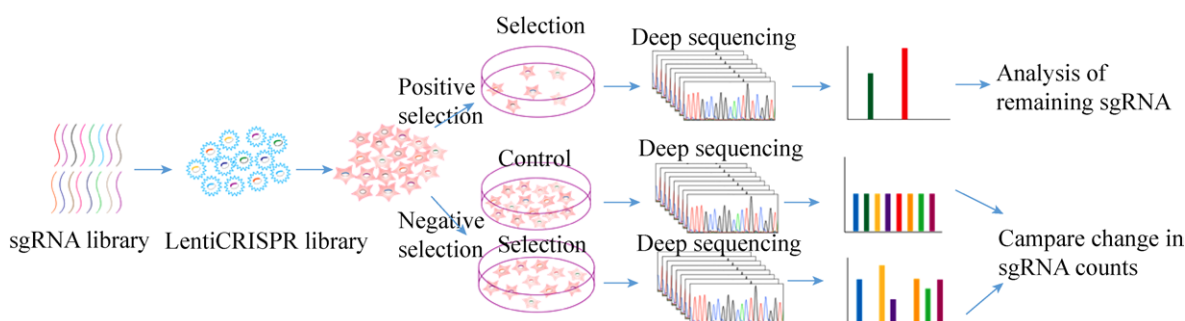


图1 CRISPR-Cas9 系统功能基因筛选流程

Fig. 1 Flowchart of functional genetic screening using CRISPR-Cas9 system.

效果。Shalem 等<sup>[29]</sup>设计了靶向全基因组 18 080 个基因,包含 64 751 条 sgRNA 的 CRISPR-Cas9 基因敲除文库,并命名为 GeCKO (Genome-scale CRISPR-Cas9 knockout) 文库。Shalem 等利用慢病毒载体将 GeCKO 文库转染进有 V600E BRAF 突变的黑色素瘤细胞 A375 中,加入 Vemurafenib 抑制 A375 细胞生长,从而富集少数具有 Vemurafenib 药物抗性的细胞。通过对这部分细胞的 sgRNA 序列进行分析,Shalem 等筛选出了 *nf1*、*med12*、*nf2*、*cul3*、*tada2b* 和 *tada1* 等敲除后能够使细胞对 Vemurafenib 产生抗性的基因,其中 *nf2*、*cul3*、*tada2b* 和 *tada1* 之前尚未被报道过与 Vemurafenib 耐药性有关。这些候选基因提出了新的 Vemurafenib 肿瘤耐药机制假说。研究者还将筛选结果与之前使用 RNAi<sup>[33]</sup>筛选 Vemurafenib 耐药基因的结果进行比较评估,两者结果高度一致,证明了 CRISPR-Cas9 用于功能性基因筛选的可行性。

同期,Wang 及其团队也采用了类似的筛选方式研究了 DNA 错配修复的相关基因<sup>[34]</sup>。研究者使用能够引起细胞损伤的核苷类似物 6-硫代鸟嘌呤作为筛选抗性,MMR 通路正常的细胞无法修复 6-硫代鸟嘌呤引起的损伤而导致细胞滞留在 G<sub>2</sub>-M 期无法进行正常的分裂,而 MMR 缺陷细胞则不能识别这种损伤从而继续进行分裂。Wang 等设计合成了靶向 7 114 个基因的 sgRNA 文库,并

且每个基因至少设计了 10 条 sgRNA 以确保文库的 knockout 效率。与 Shalem 的方法略有差别的是,Wang 等并未采用单质粒的慢病毒载体系统,而是先构建稳定表达 Cas9 的 KBM7 细胞,再转染了整个 sgRNA 文库。与单质粒系统相比,双质粒系统可以通过挑选高效的 Cas9 表达细胞株来提高文库的 knockout 效率。完成建库工作后研究者利用 6-硫代鸟嘌呤的细胞致死作用,杀死 MMR 通路正常细胞,富集 MMR 通路突变细胞。对存活细胞的 sgRNA 测序分析,研究者发现被富集细胞的 sgRNA 主要靶向 MMR 通路的 4 个关键基因 *msh2*、*msh6*、*mlh1* 和 *pms2*。

sgRNA 文库的构建是实现功能性基因筛选的关键。覆盖全基因的文库虽然更加全面,但是由于库容量庞大,在实际操作过程中需要大量的细胞,工作量及操作难度均较大。在研究炭疽毒素与白喉毒素毒性作用相关基因的工作中,Zhou 等<sup>[35]</sup>通过对已掌握知识进行整理,挑选出了可能与研究内容相关的 291 基因,并设计了靶向这 291 个基因的包含 869 条 sgRNA 的慢病毒聚焦型人源文库。该课题组使用炭疽毒素和白喉毒素分别进行阳性筛选,经过三轮毒素处理,待对照组细胞全部死亡后,对实验组存活细胞进行测序分析。除了已知的炭疽毒素受体 *Antxr1*,研究者还成功鉴定 *plxna1*、*pecr*、*fzd10* 和 *CD81* 等与炭疽毒素作用

机制相关的基因。已知白喉毒素相关基因 *hbegf* 及新的与白喉毒素相关基因 *rab2a* 也被富集。这种聚焦型的文库缩小了筛选范围,降低了操作难度和实验成本,尤其适用于研究者对筛选功能基因已有一定了解的情况。根据实际情况选择适合的文库容量能够更有效地进行文库筛选工作,提高实验成功率。

### 3.2 肿瘤功能基因筛选

肿瘤是机体在遗传及环境等多重因素的影响下,致癌基因和抑癌基因发生突变而导致的细胞生长不受调控的疾病,严重威胁着人类的生命健康,引起了科学家的高度重视。运用高通量筛选技术鉴别参与肿瘤发生发展及转移的关键基因是研发抗肿瘤新药的重要手段。CRISPR 文库也被应用于肿瘤治疗靶点和新型抗肿瘤药物筛选,取得了良好的成果。

胰腺癌生长快、转移率高且预后差,被称为“癌中之王”,其中胰腺导管腺癌(PDA)是胰腺癌的一个主要的类型。Steinhart 等<sup>[36]</sup>选取了一株 RNF43 突变的胰腺导管腺癌细胞(PDAC)HPAF-II 来筛选胰腺癌新的治疗靶标。Steinhart 及其团队中将靶向全基因组 17 232 个基因的 TKO gRNA 文库转导进 HPAF-II 细胞,通过对不同时间点细胞的 sgRNA 的丰度变化进行分析从而鉴别影响细胞增殖的基因。Steinhart 等鉴别出了 2 174 个影响 HPAF-II 细胞增殖的基因,其中参与 Wnt 信号通路的 FZD5,对 PDAC 的增殖具有重要的意义。随后研究者还发现特异性结合 FZD5 和 FZD8 的抗体能够抑制 RNF43 突变 PDAC 细胞生长,在肿瘤异种移植模型中也有类似作用,并且对 RNF43 突变的结直肠癌也有抑制效果。因此 Steinhart 等提出 FZD5 可以作为胰腺癌的潜在治疗靶点,基于 CRISPR 的高通量功能性基因筛选是筛选细胞表面特异性治疗靶标的有效手段。

一些抗肿瘤药物仅对发生特异性基因突变的

细胞起作用,其作用机制是当肿瘤细胞中某个基因发生突变时抑制另一基因表达,因而能够特异性地杀伤突变率高的肿瘤细胞而不影响正常细胞,这一作用方式被称为合成致死性。Shen 等<sup>[37]</sup>从这一思路出发,设计了双重 gRNA 文库筛选合成致死相互作用网络。与一般的 sgRNA 文库不同,双重 gRNA 文库中每个载体包含两条 gRNA,一条靶向肿瘤细胞中高突变的肿瘤抑制基因,另一条靶向一种能够被肿瘤治疗药物抑制的基因。Shen 等设计了靶向 73 对癌症基因,共计 141 912 种基因组合的双重 gRNA 文库,并在人宫颈癌细胞、肺癌细胞和人胚肾细胞中进行了筛选,通过检测不同时间点 gRNA 丰度变化,进一步分析筛选出 120 种合成致死相互作用,为癌症新药的开发提供了新的靶点。

CRISPR-Cas9 功能性筛选技术也是研究肿瘤生长与转移的重要工具之一。Chen 等<sup>[38]</sup>用包含 67 405 条 sgRNA 的基因文库诱变处理一种非转移的鼠非小细胞肺癌细胞株,并将其移植于免疫缺陷鼠身上,发现肿瘤很快就出现了转移。Chen 等将肺转移细胞与原发部位细胞进行测序比较 sgRNA 丰度变化,筛选出 *nf2*、*pten*、*cdkn2a*、*trim72*、*fga*、*miR-345* 及 *miR-152* 七个与肿瘤转移相关的基因。Chen 等的工作证明了 CRISPR-Cas9 全基因组筛选不仅可以在各种细胞中开展,并且在体内也同样适用。Chen 及其团队所提供的完整的体内筛选方案在之后也被更广泛地应用于各种疾病的体内研究工作。与体外筛选相比,动物体内筛选模型具有更复杂的微环境,比体外更接近真实的肿瘤微环境,从而有可能筛选到依赖于肿瘤微环境的肿瘤抑制基因。

### 3.3 病毒感染基因筛选

许多疾病都与病毒感染相关,筛选出病毒受体的病毒在宿主细胞中复制关键基因有助于开发新的传染病预防及治疗策略。西尼罗病毒(West

Nile virus, WNV) 是一种脑炎病毒, 能够引起急性神经感染并导致大量神经元细胞死亡。为了研究 WNV 导致细胞死亡的机制, Ma 及其团队<sup>[39]</sup>构建了靶向人类基因组 20 121 个基因的 sgRNA 文库, 并将每条 sgRNA 的靶点都设计在转录起始位点附近以确保基因能够被完全敲除。Ma 等将构建好的文库转导进 293FT 细胞中, 并感染 WNV 病毒株进行了第一轮筛选。Ma 等对第一轮筛选结果进行分析, 发现有部分丰度较高的 sgRNA 其靶向基因的其他 sgRNA 并未被富集, 这些结果可能是由假阳性造成的。因此 Ma 及其团队又针对初次实验结果富集到的 sgRNA 设计了亚基因敲除文库进行重复实验, 最终筛选出 7 个与 WNV 诱导细胞死亡相关的关键基因: *emc2*、*emc3*、*sel1l*、*derl2*、*ubeg2*、*ube2j1* 和 *hrd1*。研究者证实了敲除这些基因能够有效保护细胞不被 WNV 杀死, 这 7 个基因可以作为新的治疗靶点以降低 WNV 导致的机体损伤。

HIV 引发的艾滋病 (AIDS) 严重威胁着人类生命健康, 阐明 HIV 突破宿主细胞防御系统的分子机制, 开发 HIV 治疗的新靶点, 进而提出新的 HIV 防治策略, 具有非常重大的理论意义和实际应用价值。Park 等<sup>[40]</sup>为了阐明 HIV 在 T 细胞上的受体, 设计了适用于 CRISPR 筛选的 CD4<sup>+</sup> T 细胞 (GXRCas9 细胞) 筛选模型, 并用包含 187 536 条 sgRNA (靶向 18 543 个基因) 的慢病毒文库感染该细胞, 获得了超过 18 万种缺失不同受体的 T 细胞库。之后研究人员用 HIV 病毒株 JR-CSF 感染这些缺失不同受体的 T 细胞, 随后通过流式分选 GFP 阴性和阳性的 T 细胞 (被病毒感染的细胞能够检测到 GFP 的表达), 并对 GFP 阴性细胞群与未感染 HIV 病毒细胞测序, 分析两群细胞 sgRNA 的丰度差异, 最终发现了 5 个 sgRNA 丰度变化最大的基因。其中 CD4 和 CCR5 是 HIV 感染 T 细胞的受体, 酪蛋白磺基转移酶 2 (Tyrosylprotein

sulfotransferase 2, TPST2) 和溶质家族 35 成员 B2 (Solute carrier family 35 member B2, SLC35B2) 对 CCR5 进行修饰以便于与 HIV 结合。另一种基因编码白细胞粘附因子 (Activated leukocyte cell adhesion molecule, ALCAM), 与 HIV 在细胞间的传播相关。筛选出的 5 个基因被敲除后均不影响 T 细胞的存活, 但能够使 T 细胞抵抗 HIV 的感染, 因此可以作为 HIV 治疗的潜在靶标, 为预防和治疗艾滋病提供新的理念和途径。

目前功能缺失型 (Loss-of-function) 筛选已被广泛应用于寻找病毒感染、复制和传播所需基因, 但过表达基因影响病毒感染的研究还比较欠缺。基于此 Heaton 等<sup>[41]</sup>应用 CRISPR 协同激活调节 (CRISPR synergistic activation mediator, CRISPR SAM) 技术筛选过表达后抑制甲型流感病毒 (IVA) 感染的基因。研究者首先将 dCas9-VP64 融合蛋白和 MS2-p65-HSF1 转录因子转导进 A549-CR 细胞中, 并将靶向全基因组 23 430 个基因转录起始位点上游 200 bp 处的 sgRNA 文库导入转录激活效率最高的细胞株, 促进基因的过表达。之后研究者用 IVA 感染细胞, 通过流式分选出未被病毒感染的细胞进行测序分析, 发现靶向 *b4galnt2* 基因的 sgRNA 丰度最高。经过多重验证, Heaton 等证明了 *b4galnt2* 过表达能够限制多种亚型的甲型流感病毒感染, 未来可能被应用于防治人流感及禽流感。

为了研究影响溶瘤病毒感染及杀伤肿瘤细胞的关键基因, 我们同样采用基于 GeCKO v2 全基因组敲除文库的功能缺失型 (Loss-of-function) 筛选方法。我们尝试了两种筛选方案: 第一种是阵列筛选 (Arrayed screen), 即将已成功转导 GeCKO v2 文库的病毒抵抗型细胞分入 96 孔板, 通过免疫斑点法检测病毒复制情况, 挑选病毒复制提高的孔进行下一步验证; 第二种方法是合并筛选 (Pooled screen), 即将成功转导 GeCKO v2



文库的细胞感染病毒,利用高通量测序分析病毒感染后细胞 sgRNA 丰度变化情况。与合并筛选相比,阵列筛选每孔细胞数较少,因此在培养过程中易出现 sgRNA 丢失的情况,同时实验结果还与用于检测病毒复制抗体的亲和性有关,容易出现假阳性,降低了实验的准确性,且该方法工作量较大,操作难度高。合并筛选是较为便捷准确的筛选方式,需要注意的是筛选所用细胞的选择,可依据正向筛选和负向筛选分别选择病毒敏感型细胞和病毒抵抗型细胞。

### 3.4 非编码基因功能研究

基因组中大多数 DNA 并不编码蛋白质,但却与基因表达调控密切相关,研究表明超过 90% 遗传变异相关疾病发生在基因非编码区域<sup>[42]</sup>。阐明这些非编码基因如何参与基因表达调控有助于人们更好地研究基因异常导致的疾病。在之前的工作中,Shalem 等<sup>[29]</sup>利用 CRISPR-Cas9 文库发现了 *nf1*、*nf2*、*clu3* 等功能缺失后使黑色素瘤细胞对 Vemurafenib 产生抗性的基因。在此基础上,Sanjana 等<sup>[43]</sup>设计了 3 个分别靶向 *cul3*、*nf1* 和 *nf2* 基因 5'端和 3'端紧邻的 100 kb 范围内基因区域的 sgRNA 文库,来研究非编码区对相关基因的调控机制。Sanjana 等将 sgRNA 文库转导进 BRAF 突变的 A375 细胞中,将 Vemurafenib 处理后的存活细胞测序分析。研究者对富集 sgRNA 最多的 *cul3* 基因进行了重点分析,发现在 *cul3* 基因周围有 24 个基因位点突变能够导致 *cul3* 表达降低。Sanjana 及其团队的工作证明了 CRISPR-Cas9 文库不仅可以用于筛选功能性基因,还可以用来研究非编码基因功能。

对于非编码长链 RNA (lncRNA),插入或缺失突变导致的变化可能并不足以产生明显的表型变化。针对这一情况,Zhu 等<sup>[44]</sup>设计了配对向导 RNA (Paired-guide RNA, pgRNA),对人肝癌细胞系 Huh7.5OC 中的近 700 个癌症或其他疾病相关长

链非编码 RNA 的基因进行了功能筛选。pgRNA 是一对靶向不同位点的 sgRNA,能够在同一基因的不同位点造成双链断裂,因此有可能产生大片段的缺失,造成可观察的表型变化。为了验证 pgRNA 的有效性,研究人员设计了靶向 *cspg4* 基因座的 pgRNA,使用双 U6 启动子分别启动两个 gRNA 的表达,结果表明所有的 6 对 pgRNA 均能够对基因正确切割。基于此,Zhu 等针对 Huh7.5OC 中的近 700 个癌症或其他疾病相关长链非编码 RNA 设计了 gRNA,剔除其中特异性和效率较低的 gRNA 及靶向启动子和外显子的 gRNA,获得了可用于筛选的基因文库。之后研究者将文库转导进稳定表达 Cas9 的 Huh7.5OC,持续培养 30 d 后对 CRISPR 筛选前和筛选后的细胞测序分析来鉴定能够影响细胞增殖和生存的 lncRNA。经过多重的验证,Zhu 等鉴定出了 51 个与肿瘤细胞增殖相关的非编码 RNA,证实了利用 pgRNA 研究 lncRNA 的方法具有较高的精确性和特异性。同时,研究人员在 Hela 细胞中的重复实验结果也证明了这一点。但是这一筛选策略仍存在一些问题。首先删除 lncRNA 可能影响邻近功能元件如增强子和 microRNA 的功能,因此需要在设计 pgRNA 时尽可能地排除靶向增强子和 microRNA 区域的 gRNA。由于 pgRNA 间具有相同的 U6 启动子和 3'支架序列,不同的 pgRNA 间有可能(约 7.5%)会发生重组导致错误的配对,使用不同的 U6 启动子和 3'支架序列可以降低重组率。此外,这种筛选策略并不能揭示 lncRNA 的调节机制,还需要后续的工作进行探究。

Klann 及其团队<sup>[45]</sup>介绍了一种以 CRISPR-Cas9 技术为基础的表观遗传调节元件筛选方法(CERES)。识别调节元件所有可能的转录因子结合位点需要针对该调节元件区域设计高密度的 sgRNA。然而在许多基因区域 PAM 序列可能并不在转录因子结合位点,导致 sgRNA 不能充分覆盖



所有的转录因子结合位点。因此 Klann 等选用了转录抑制的 dCas9KRAB 酶和转录激活的 dCas9p300 酶。dCas9KRAB 是将 KRAB (Krüppel-associated box) 区域与 dCas9 融合表达, 招募使组蛋白 H3K9 发生甲基化的蛋白进而形成异染色质导致靶序列上的基因抑制<sup>[46]</sup>。E1A 相关蛋白 p300 组蛋白乙酰转移酶核心区与 dCas9 融合后能够结合到靶 DNA 增强子或启动子上, 促进组蛋白 H3K27 发生乙酰化, 从而促使基因激活<sup>[47]</sup>。这两种方法即使在转录因子结合位点没有 PAM 序列的情况下仍然能够对调节元件活性进行调节。研究人员根据人白血病细胞系 K562 DNase-seq 数据设计了针对  $\beta$ -globin 基因座周围 4.5 Mb 区域的 10 739 条 sgRNA。这些 sgRNA 靶向该区域 281 个 DNase I 超敏感位点 (DNase I hypersensitive sites, DHSs)。研究者将构建好的 sgRNA 导入改造过的细胞 ( $\beta$ -globin 的转录与荧光蛋白相关联) 中, 并分别将荧光表达最高和最低的 10% 细胞分选出来分析 sgRNA 的相对丰度变化。Klann 及其团队的工作为高通量注释调节元件的作用提高了效率。

### 3.5 其他工作

除了应用于肿瘤、病毒、非编码 RNA 的相关功能基因研究外, CRISPR-Cas9 高通量筛选在抗体靶标筛选、细胞信号通路研究等生物医学的其他领域中也有着广泛的应用。

单克隆抗体特异性靶标抗原及其表位的识别是抗体研究中的重要工作。传统的抗原特异性鉴别方式是 Western blotting (蛋白印迹) 和 Immunoprecipitation (免疫沉淀), 但是当抗体不能被 Western blotting 和 Immunoprecipitation 检测到时, 就需要使用基于基因的技术来鉴定抗体靶标。Zotova 等<sup>[48]</sup>以 MT2 细胞 (人 T 细胞嗜淋巴病毒 I 型 HTLV-1 慢性感染 T 细胞) 作为免疫原免疫小鼠获得了一株 HTLV-1 生物膜特异性单克隆抗体 BF4。BF4 能够和未感染的淋巴细胞、中性粒

细胞及 HTLV-1 感染细胞表面的病毒生物膜结合。Zotova 等的研究思路是在 BF4 阳性细胞上转导 CRISPR knockout 文库, 分选出阴性细胞, 即可能是 BF4 抗原被敲除的细胞。基于这样的思路研究者在 CEM T 和 Raji/CD4 B 细胞上转导 GeCKO 文库, 并将不与 BF4 结合的细胞分选出来。经过两轮的重重复分选, 阴性细胞比例达到了 99% 以上。研究人员对这部分细胞进行测序分析, 发现约 80% 的 sgRNA 靶向 CD82, 经过验证, 证实了 BF4 确实是 CD82 特异性抗体。在筛选抗体靶标的研究中还有一些需要注意的地方, 首先这种筛选方式不适合筛选抗体抗原是细胞生存所需的基因, 因为这部分基因被敲除后会导致细胞无法存活而丢失, 无法在后续的分选工作中被鉴定出。Cas9 文库的代表性和多样性也是需要关注的问题, 不同的细胞和慢病毒载体都会对此造成影响。

Shi 及其团队<sup>[49]</sup>在 *Nature* 上发表了关于细胞焦亡机制的相关研究。细胞焦亡是机体在感知病原微生物浸染后启动的免疫防御反应, 炎症激活的 caspase-1 和识别细菌脂多糖的 caspase-4、caspase-5 和 caspase-11 都能够引起细胞焦亡, 但是其机制仍然未知。研究人员首先建立了脂多糖 (LPS) 电转方法, 使 LPS 能够诱导 90% 以上的细胞焦亡。接着在能够对脂多糖刺激正常反应的 Tlr4<sup>-/-</sup>骨髓源永生生化巨噬细胞 (iBMDMs) 转导 CRISPR knockout 基因文库, 将脂多糖刺激细胞焦亡后存活的细胞测序分析。分析结果显示, 靶向 gasdermin D (GSDMD) 基因的 5 条 sgRNA 中, 有 4 条在富集到的前 30 中, 其中有 2 条在前 10。后续的结果进一步证明了 GSDMD 的 N 端能够诱导细胞焦亡。在此研究中, 研究人员利用 CRISPR 基因文库进行了全基因组范围的遗传筛选, 成功地筛选到了敲除后能够抑制细胞焦亡的基因 GSDMD, 并阐明了 GSDMD 作为炎症 caspase 底物蛋白, 在被切割后能够引发细胞焦亡的分子机制。

## 4 结语

CRISPR-Cas9 技术作为近年来热门的基因编辑技术, 凭借其精准、简便、高效的特点受到众多研究者的关注, 短短几年间已成为基因编辑领域中最热门的技术, 被广泛应用于生物、医学等领域。基于 CRISPR-Cas9 的高通量功能性筛选技术也逐渐取代了 RNAi 和 cDNA 文库, 为功能基因的研究提供了高通量高效率的筛选工具。尽管与传统的筛选方法相比, CRISPR-Cas9 更加高效可靠且特异性更高, 但其仍然存在一些问题, 其中最令人关注的就是脱靶效率。CRISPR-Cas9 系统依据 20 nt 的 sgRNA 定位靶基因, 因此有可能由于错配而出现脱靶问题。除了设计特异性更高的 sgRNA 外, 研究者们提供了其他的降低脱靶效率的策略。例如可以使用只切开 DNA 双链一侧的 Cas9n 切口酶<sup>[50]</sup>, 需要一对 sgRNA 才能完成 DNA 双链剪切, 因此可以降低脱靶效率。Vidigal 等<sup>[51]</sup>则是将 20 nt 的 sgRNA 缩短为 17 nt, 剪切效率不变而脱靶效率降低。另一种策略是将 dCas9 与 Fok I 核酸酶融合表达来提高筛选特异性<sup>[52]</sup>。这些方法虽然能够降低脱靶效率, 但是还需要进一步改进才能应用于高通量的功能性基因筛选。

基于 CRISPR-Cas9 功能性基因筛选技术已经在细胞生存必需基因鉴定、药物或毒素抗性基因筛选、潜在治疗靶标筛选、肿瘤转移等相关基因筛选方面取得了重要进展, 为开发有效的靶点特异性治疗药物奠定了基础, 未来也必然会推动整个生命科学的快速发展。

## REFERENCES

- [1] Yang XP, Boehm JS, Yang XP, et al. A public genome-scale lentiviral expression library of human ORFs. *Nat Meth*, 2011, 8(8): 659–661.
- [2] Rana TM. Illuminating the silence: understanding the structure and function of small RNAs. *Nat Rev Mol Cell Biol*, 2007, 8(1): 23–36.
- [3] Mohr SE, Smith JA, Shamu CE, et al. RNAi screening comes of age: improved techniques and complementary approaches. *Nat Rev Mol Cell Biol*, 2014, 15(9): 591–600.
- [4] Sachse C, Krausz E, Krönke A, et al. High-throughput RNA interference strategies for target discovery and validation by using synthetic short interfering RNAs: functional genomics investigations of biological pathways. *Methods Enzymol*, 2005, 392: 242–277.
- [5] Qiu SB, Adema CM, Lane T. A computational study of off-target effects of RNA interference. *Nucleic Acids Res*, 2005, 33(6): 1834–1847.
- [6] Gratz SJ, Wildonger J, Harrison MM, et al. CRISPR/Cas9-mediated genome engineering and the promise of designer flies on demand. *Fly (Austin)*, 2013, 7(4): 249–255.
- [7] Jinek M, East A, Cheng A, et al. RNA-programmed genome editing in human cells. *eLife*, 2013, 2: e00471.
- [8] Hruscha A, Krawitz P, Rechenberg A, et al. Efficient CRISPR/Cas9 genome editing with low off-target effects in zebrafish. *Development*, 2013, 140(24): 4982–4987.
- [9] Shan QW, Wang YP, Li J, et al. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat Biotechnol*, 2013, 31(8): 686–688.
- [10] Fujii W, Onuma A, Sugiura K, et al. Efficient generation of genome-modified mice via offset-nicking by CRISPR/Cas system. *Biochem Biophys Res Commun*, 2014, 445(4): 791–794.
- [11] Shalem O, Sanjana NE, Zhang F. High-throughput functional genomics using CRISPR-Cas9. *Nat Rev Genet*, 2015, 16(5): 299–311.
- [12] Wei CX, Liu JY, Yu ZS, et al. TALEN or Cas9-rapid, efficient and specific choices for genome modifications. *J Genet Genomics*, 2013, 40(6): 281–289.
- [13] 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.
- [14] Horvath P, Romero DA, Coûté-Monvoisin AC, et al. Diversity, activity, and evolution of CRISPR loci in *Streptococcus thermophilus*. *J Bacteriol*, 2008, 190(4): 1401–1412.
- [15] Horvath P, Coûté-Monvoisin AC, Romero DA, et al. Comparative analysis of CRISPR loci in lactic acid

- bacteria genomes. *Int J Food Microbiol*, 2009, 131(1): 62–70.
- [16] Garneau JE, Dupuis MÈ, Villion M, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, 2010, 468(7320): 67–71.
- [17] Ran FA, Hsu PD, Wright J, et al. Genome engineering using the CRISPR-Cas9 system. *Nat Protoc*, 2013, 8(11): 2281–2308.
- [18] Karginov FV, Hannon GJ. The CRISPR system: small RNA-guided defense in bacteria and archaea. *Mol Cell*, 2010, 37(1): 7–19.
- [19] Grissa I, Vergnaud G, Pourcel C. The CRISPRdb database and tools to display CRISPRs and to generate dictionaries of spacers and repeats. *BMC Bioinform*, 2007, 8: 172.
- [20] Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science*, 2013, 339(6121): 819–823.
- [21] Hsu PD, Scott DA, Weinstein JA, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol*, 2013, 31(9): 827–832.
- [22] Gasiunas G, Barrangou R, Horvath P, et al. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA*, 2012, 109(39): E2579–E2586.
- [23] Jinek M, Jiang FG, Taylor DW, et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science*, 2014, 343(6176): 1247997.
- [24] Mali P, Yang L, Esvelt KM, et al. RNA-guided human genome engineering via Cas9. *Science*, 2013, 339(6121): 823–826.
- [25] Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 2012, 337(6096): 816–821.
- [26] Gilbert LA, Larson MH, Morsut L, et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*, 2013, 154(2): 442–451.
- [27] Perez-Pinera P, Kocak DD, Vockley CM, et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods*, 2013, 10(10): 973–976.
- [28] Chuai GH, Wang QL, Liu Q. *In silico* meets *in vivo*: towards computational CRISPR-based sgRNA design. *Trends Biotechnol*, 2017, 35(1): 12–21.
- [29] Shalem O, Sanjana NE, Hartenian E, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science*, 2014, 343(6166): 84–87.
- [30] Xu H, Xiao TF, Chen CH, et al. Sequence determinants of improved CRISPR sgRNA design. *Genome Res*, 2015, 25(8): 1147–1157.
- [31] Li W, Xu H, Xiao TF, et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol*, 2014, 15: 554.
- [32] Li W, Köster J, Xu H, et al. Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol*, 2015, 16: 281.
- [33] Whittaker SR, Theurillat JP, van Allen E, et al. A genome-scale RNA interference screen implicates NF1 loss in resistance to RAF inhibition. *Cancer Discov*, 2013, 3(3): 350–362.
- [34] Wang T, Wei JJ, Sabatini DM, et al. Genetic screens in human cells using the CRISPR-Cas9 system. *Science*, 2014, 343(6166): 80–84.
- [35] Zhou YX, Zhu SY, Cai CZ, et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature*, 2014, 509(7501): 487–491.
- [36] Steinhart Z, Pavlovic Z, Chandrashekhar M, et al. Genome-wide CRISPR screens reveal a Wnt-FZD5 signaling circuit as a druggable vulnerability of *RNF43*-mutant pancreatic tumors. *Nat Med*, 2017, 23(1): 60–68.
- [37] Shen JP, Zhao DX, Sasik R, et al. Combinatorial CRISPR-Cas9 screens for *de novo* mapping of genetic interactions. *Nat Methods*, 2017, 14(6): 573–576.
- [38] Chen SD, Sanjana NE, Zheng KJ, et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell*, 2015, 160(6): 1246–1260.
- [39] Ma HM, Dang Y, Wu YG, et al. A CRISPR-based screen identifies genes essential for west-nile-virus-induced cell death. *Cell Rep*, 2015, 12(4): 673–683.
- [40] Park RJ, Wang T, Koundakjian D, et al. A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nat Genet*, 2017, 49(2): 193–203.
- [41] Heaton BE, Kennedy EM, Dumm RE, et al. A CRISPR activation screen identifies a pan-avian influenza virus inhibitory host factor. *Cell Rep*, 2017, 20(7): 1503–1512.

- 
- [42] Thurman RE, Rynes E, Humbert R, et al. The accessible chromatin landscape of the human genome. *Nature*, 2012, 489(7414): 75–82.
- [43] Sanjana NE, Wright J, Zheng KJ, et al. High-resolution interrogation of functional elements in the noncoding genome. *Science*, 2016, 353(6307): 1545–1549.
- [44] Zhu SY, Li W, Liu JZ, et al. Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. *Nat Biotechnol*, 2016, 34(12): 1279–1286.
- [45] Klann TS, Black JB, Chellappan M, et al. CRISPR-Cas9 epigenome editing enables high-throughput screening for functional regulatory elements in the human genome. *Nat Biotechnol*, 2017, 35(6): 561–568.
- [46] Thakore PI, D'Ippolito AM, Song LY, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods*, 2015, 12(12): 1143–1149.
- [47] Hilton IB, D'Ippolito AM, Vockley CM, et al. Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol*, 2015, 33(5): 510–517.
- [48] Zotova A, Zotov I, Filatov A, et al. Determining antigen specificity of a monoclonal antibody using genome-scale CRISPR-Cas9 knockout library. *J Immunol Methods*, 2016, 439: 8–14.
- [49] Shi JJ, Zhao Y, Wang K, et al. Cleavage of GSDMD by inflammatory caspases determines pyroptotic cell death. *Nature*, 2015, 526(7575): 660–665.
- [50] Ran FA, Hsu PD, Lin CY, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 2013, 154(6): 1380–1389.
- [51] Vidigal JA, Ventura A. Rapid and efficient one-step generation of paired gRNA CRISPR-Cas9 libraries. *Nat Commun*, 2015, 6: 8083.
- [52] Tsai SQ, Wyvekens N, Khayter C, et al. Dimeric CRISPR RNA-guided *Fok I* nucleases for highly specific genome editing. *Nat Biotechnol*, 2014, 32(6): 569–576.

(本文责编 郝丽芳)