



BE-dot: 为单碱基编辑设计 sgRNA 及预测脱靶图谱的工具*

王泽鲁 梁俊波** 王晓月**

(中国医学科学院基础医学研究所, 北京协和医学院基础学院, 生物化学与分子生物学系, 北京 100005)

摘要 **目的** 单碱基编辑器作为修复基因组点突变的有力工具, 在生物技术开发和临床应用中具有极大潜力。对于目标改造的单核苷酸变异 (SNV), 首先要选择可用于编辑的单碱基编辑器 (base editors, BEs) 和单向导 RNA (single guide RNA, sgRNA)。目前, 尽管有较多工具实现了设计 sgRNA, 但缺少将设计 sgRNA 与评估单碱基编辑器特异性完整结合起来的工具。**方法** 纳入主流的 27 种胞嘧啶碱基编辑器 (CBEs) 和 12 种腺嘌呤碱基编辑器 (ABEs) 对 SNV 设计编辑方案, 并通过调用第三方工具 BE-Hive 对编辑方案进行效率预测。综合使用多个脱靶预测工具对编辑方案的脱靶图谱进行评估。最后结合 BEs 类型和脱靶位点, 分析得到所有可能的脱靶编辑产物, 再调用 ANNOVAR 这一变异注释工具进行脱靶产物的功能分析。**结果** 本文提出了一个综合性工具 BE-dot, 它可以实现对目标编辑的 SNV 从设计 sgRNA 到预测脱靶图谱的完整过程, 并对脱靶产物进行功能注释。利用密码子的简并性, BE-dot 除了提供 DNA 水平上的精确校正方案, 还可以在蛋白质水平进行同义校正。在对单碱基编辑系统做脱靶图谱的预测时, BE-dot 综合了 Cas-OFFinder、CALITAS、CFD、uCRISPR、BEdecoff 等多个工具, 可以更全面地评估单碱基编辑系统的特异性, 为用户选择 BEs 和 sgRNA 提供参考。此外, BE-dot 可以自动分析得到脱靶位点处所有可能的编辑产物, 并转换为 avinput 格式供 ANNOVAR 进行功能注释, 避免了以往手工注释的繁琐。**结论** BE-dot 能为单碱基编辑技术应用于纠正或引入 SNV 设计编辑方案, 并且能够从编辑效率、脱靶图谱、脱靶带来的功能影响等方面对编辑方案进行全面的评估。

关键词 单碱基编辑器, 单向导 RNA 设计, 脱靶图谱预测

中图分类号 Q31, Q78

DOI: 10.16476/j.pibb.2022.0046

基因组编辑工具, 尤其是基于 CRISPR 系统的单碱基编辑器自推出以来就显示了广阔的临床应用前景^[1]。单碱基编辑器 (base editors, BEs) 能高效地对指定位点的单个核苷酸变异 (SNV) 进行精准修复, 为治疗遗传性疾病提供了良好方案^[2-3]。主流的 BEs 由催化活性受损的 Cas9 酶连接脱氨酶构成, 在单向导 RNA (single guide RNA, sgRNA) 的指导下, 定位到预先设计的基因组位置发挥作用^[4]。根据实现的单碱基替换类型, BEs 可分为腺嘌呤碱基编辑器 (ABEs) 和胞嘧啶碱基编辑器 (CBEs)。ABEs 能实现 A 到 G 的直接转换, CBEs 能将 C 转换为 T^[3, 5]。综合 ABEs 和 CBEs, 它们能实现对所有转换类型 SNVs 的修复。在人类致病性遗传变异中, 点突变占比 58%, 点突变类型为“转换”的 SNVs 占 61%。BE 校正 SNVs 除了在 DNA

水平精准修复, 还可以在蛋白质水平进行修复。利用密码子的简并性, 蛋白质水平的校正增加了可选用的 BEs 范围, 也增加了在已有的 BEs 种类下可实现校正的 SNVs 范围^[6]。

近年来, 多种 ABEs 和 CBEs 被开发出来。2020 年, 研究者又开发出可介导 CG 碱基颠换的单碱基编辑工具 CGBEs, 在编辑效率和产物纯度得到极大提高后, CGBEs 在未来研究中的应用极具潜力^[7]。以上多种 BEs 在识别的原间隔序列邻近基序 (protospacer adjacent motif, PAM) 序列、具有编辑活性的序列范围等属性上有所差异^[6, 8-9]。对

* 国家自然科学基金 (32070603) 资助项目。

** 通讯联系人。

梁俊波 Tel: 010-69156413, E-mail: liangjunbo@ibms.pumc.edu.cn

王晓月 Tel: 010-69156413, E-mail: pumcwxxy@163.com

收稿日期: 2022-02-10, 接受日期: 2022-05-06

目标点突变进行编辑时, 需要结合点突变的类型及其背景序列特点, 选择合适的 BEs 和 sgRNA^[10]。因此, 借助计算工具来设计 sgRNA 十分必要。不同于 Cas9 在编辑位点产生双链 DNA 断裂 (DSBs), BEs 不会引入 DSBs, 从而被认为是更安全的基因组编辑工具^[2]。在应用于临床之前, 十分有必要对 BEs 的特异性进行详尽的评估。

目前, 虽然有较多工具可以实现 BEs 的 sgRNA 设计, 如 BE-designer^[5]、BE-FF^[6]、beditor^[11], 但缺少工具将下游的脱靶 (off-target, OT) 图谱分析及脱靶产物注释纳入^[12-13]。已有预测 CRISPR 系统脱靶的工具, 如 Cas-OFFinder^[14]、CFD^[15]、uCRISPR^[16]; 最新开发出的预测 BEs 脱靶编辑效率工具 BEdeepoff 包含了两个独立工具——ABEdeepoff 和 CBEdeepoff。它们分别基于 ABEmax 和 BE4max 在成对的 sgRNA-脱靶 DNA 高通量文库上的编辑效率构建得到^[17]。以上预测脱靶工具可以对脱靶位点以及脱靶位点处的编辑活性进行预测, 然而不同工具在预测结果上具有差异^[18-19], 综合使用多个脱靶预测工具可以为脱靶评估提供更全面的参考。

基于此, 本文提出了综合性工具 BE-dot, BE-dot 实现了设计 sgRNA, 综合多个工具预测脱靶图谱以及对脱靶产物进行功能注释的完整过程。借鉴已有工具 BE-FF 对 SNVs 的同义校正, BE-dot 在设计 sgRNA 时不仅提供了在 DNA 水平上对 SNVs 的精确校正方案, 还提供了在蛋白质水平校正突变的方案, 使得非转换类型的 SNVs 也有可能被 BEs 校正。在临床应用中除了必要考虑的单碱基编辑系统的脱靶效应, BEs 在靶向位点的编辑效率也是评估编辑方案优劣的重要方面, 尤其是针对某一 SNV 有多个可选用的 BE-sgRNA 方案时。因此, BE-dot 添加了 BE-Hive^[20] 作为预测候选 sgRNA 编辑效率的工具, 以对用户提供更全面的参考信息。另外, BE-dot 允许用户使用自定义的 BEs, 能更灵活、更具有实用性地设计 sgRNA。软件的下载地址为: <https://github.com/wendyw630/BE-dot>。

1 方 法

1.1 sgRNA 的设计

本研究面向的基因组为人类基因组 GRCh38, BE-dot 的总工作流程见图 1。针对用户需要编辑的 SNV, 用户调用 BE-dot 来设计 sgRNA 可通过输入方式一——rsID, 或输入方式二——SNV 及其上下

游各 50 nt 序列。若用户提供的为 rsID, BE-dot 调用 python Bio 程序包的 Entrez.efetch 模块, 以 URL 格式访问 NCBI 的 Entrez 数据库, 检索其子数据库“SNP”。将检索到的序列保存到本地目录下, 然后使用 Bio 程序包的 SeqIO 模块去解析它, 可以得到 rsID 在“SNP_ID”、“CLINICAL_SIGNIFICANCE”、“GENE_ID”、“CHRPOS”、“FXN_CLASS”、“DOCSUM”等方面的信息, 进一步解析“DOCSUM”中以 HGVS^[21] 表达式记录的 cDNA 水平的突变信息, 可以得到 SNV 的突变类型。若用户接受在蛋白质水平上的校正, 以方式二输入时还需提供 SNV 所在密码子阅读框的位置 (1 或 2 或 3)。

通过判断 SNV 类型, 选择适用的 BEs 类型为 ABEs 或 CBEs 或 CGBEs。对于突变为 T→C 或 A→G, 选用 CBEs; 突变为 G→A 或 C→T, 选用 ABEs; 突变为 G→C 或 C→G, 选用 CGBEs。确定 BEs 类型后, BE-dot 对于该类型内的所有 BEs 筛选。将 SNV 在某 BE 的编辑窗口的各个位置滑动, 若 SNV 在编辑窗口内的该位置时能满足 PAM 要求, 并同时满足编辑窗口内仅有 SNV 突变碱基这一个 BE 能编辑的碱基类型, 则该 BE 及相应位置的 sgRNA 可对该 SNV 进行精确修复; 若编辑窗口内 BE 可编辑的不只有 SNV 突变后的碱基, 则可能发生旁编辑 (bystander editing), 如果对 SNV 的编辑连同旁编辑可以修复 SNV 对蛋白质的影响, 则该 BE 及相应位置的 sgRNA 记为“同义修复”。

1.2 脱靶图谱的预测

BE-dot 提供预测脱靶图谱的工具具有 Cas-OFFinder、CALITAS、CFD-score、uCRISPR、BEdeepoff (表 1)。其中可用于预测脱靶位点的工具具有 Cas-OFFinder 和 CALITAS; 可用于预测 BEs 在脱靶位点处编辑活性的工具有 CFD-score、uCRISPR、BEdeepoff, 它们不做脱靶位点的预测, 都是基于 Cas-OFFinder 的搜索结果进行编辑活性预测。

Cas-OFFinder 在更新的 3.0 及更高版本中增加了对 DNA 凸起、RNA 凸起的考虑, 即认为存在 sgRNA 与 DNA 配对中两者碱基数不相等的脱靶。在本研究中, BE-dot 设置的默认参数是: 存在 DNA 凸起或者 RNA 凸起的最大值为 1, sgRNA-DNA 错配数最大值为 3。CALITAS 设置的默认参数中 DNA 凸起和 RNA 凸起总数最大为 2, sgRNA-DNA 错配数为 3。

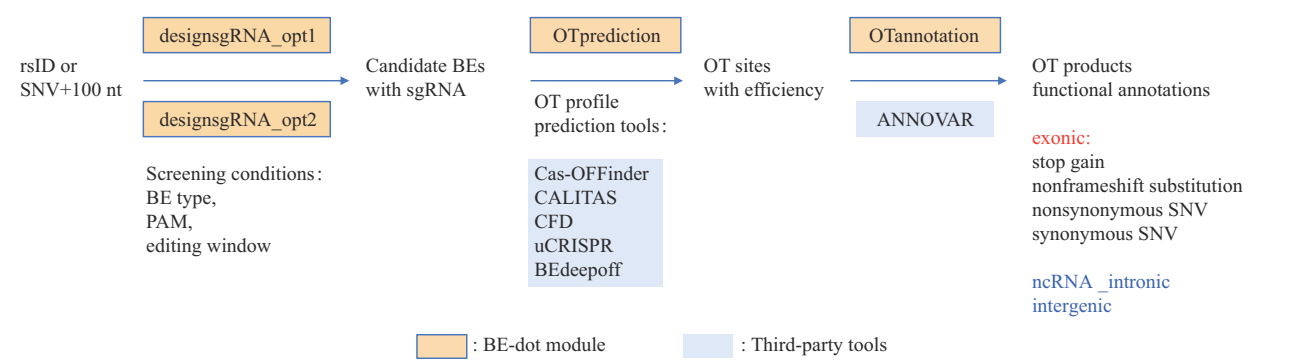


Fig. 1 The overall workflow of BE-dot

Table 1 List of 5 OT prediction tools contained in BE-dot

Tool	Mechanism	OT site prediction	Reference
Cas-OFFinder	Alignment-based	mismatch	[14]
CALITAS	Alignment-based	mismatch	[22]
CFD	Hypothesis-driven	mismatch+score	[15]
uCRISPR	Energy-based	mismatch+score	[16]
BEdeepoff	Learning-based	mismatch+score	[17]

CFD得分可以拆解为独立 sgRNA-DNA 匹配情况的得分和 PAM 序列得分，其中 PAM 序列得分规则是基于观测 CRISPR/Cas9 结合相同 sgRNA、不同 PAM 序列时的切割活性得到的，其 PAM 序列的偏好性在 PAM 序列不是 NGG 的单碱基编辑系统中很难适用。BE-dot 舍弃了 CFD 得分中对 PAM 序列打分的部分，保留了其对 sgRNA-DNA 匹配情况的打分。

BEdeepoff 是基于成对的 sgRNA-脱靶 DNA 高通量试验，分别检测 ABEmax 和 BE4max 单碱基编辑器在脱靶 DNA 上的编辑效率，构建得到预测模型 ABEddeepoff、CBEddeepoff。在设计 sgRNA-脱

靶 DNA 文库中，考虑了含有碱基插入、缺失的情况，因此 BEdeepoff 可对含有 DNA 或 RNA 凸起的 OT 位点做脱靶活性的预测。由于 CFD 和 uCRISPR 只针对无 DNA 或 RNA 凸起的 OT 位点做脱靶活性的预测，因此 Cas-OFFinder 找到的 OT 位点中与 sgRNA 等长度匹配的 OT 作为 CFD 和 uCRISPR 的输入。

用户运行 BE-dot 的脱靶预测模块时，可参考操作说明（图 2）输入相应的参数。其中，OTprediction 模块对部分参数提供了默认值，它们是 mismatch_number（默认值取 3）、DNAbulge（默认值取 1）、RNAbulge（默认值取 1）。

```
usage: BE-dot.py OTprediction [-h] -BE BE -grna GRNA [-genome <file>]
                             [-mis MISMATCH_NUMBER] [--DNAbulge DNABULGE]
                             [--RNAbulge RNABULGE] [-o OUTPUTPATH]

optional arguments:
  -h, --help            show this help message and exit
  -BE BE                BE to do OT-prediction
  -grna GRNA            Single gRNA sequence to analyse (20nt)
  -genome <file>, --target_genome <file>
                        Genome to search off-target sites
  -mis MISMATCH_NUMBER, --mismatch_number MISMATCH_NUMBER
                        Maximum mismatch number considered in aligning.
                        [default:3]
  --DNAbulge DNABULGE  Maximum DNA bulge number considered in aligning.
                        [default:1]
  --RNAbulge RNABULGE  Maximum RNA bulge number considered in aligning.
                        [default:1]
  -o OUTPUTPATH, --outputPath OUTPUTPATH
                        Path of output file(s)
```

Fig. 2 Screen shot of command lines of BE-dot OT prediction

1.3 使用ANNOVAR对脱靶编辑产物进行功能注释

对于潜在的脱靶位点，可能存在sgRNA与这段DNA配对，使得BE在此编辑产生试验设计之外的点突变。BE-dot的OTannotation模块实现了将某个OT位点上所有的编辑结果进行穷举，并将各种点突变组合自动转换为ANNOVAR^[23]的输入文件格式。

该模块需要用户提供BE名称，BE-dot将结合BE的碱基转换类型和编辑窗口，遍历脱靶序列上可能的编辑位点，列举所有可能的脱靶编辑组合情况。

2 结 果

2.1 设计sgRNA的应用举例

rs80357410是人类17号染色体的第43 124 027位碱基发生了T到C的错义突变，ClinVar数据库中的注释显示其位于*BRCAl*基因，多项研究表明该SNV与乳腺癌和卵巢癌相关^[24-25]。对此，使用BE-dot的sgRNA设计功能选择可校正此点突变的BEs

和sgRNA。

运行如下命令：python BE-dot.py designsgRNA_opt1 --jobID job001 --upSeq GCGTTGAAGAAGTACAAAATGTCATTAATGCTATGCAGAAAATCTTAGAG--downSeq GTCCCATCTGGTAAGTCAGCACAAAGAGTGATTAATTTGGGATTCC TATG --mut C --wt T --codon_frame 1 --outputPath /path/，因为该SNV在dbSNP数据库中有相关记录，所以可运行命令：python BE-dot.py designsgRNA_opt2 --rsID rs80357410 --outputPath /path/

sgRNA的设计结果见表2，BE-dot提供的碱基编辑器中能对该SNV精确修复的有Target-AID-NG、xBE3、SpRY-PmCDA1、SpRY-BE4max。此外，能在蛋白质水平修复的碱基编辑器有BE-PLUS、evoCDA1-BE4max。尽管它们的编辑窗口内除了目标编辑的C，编辑位点的上游7 nt处还存在一个C，但该C位于的密码子为ATC，与非目标编辑产物ATT同样翻译为异亮氨酸。由于BE-Hive中目前包含的BEs类型有限，在本次校正中能进行编辑效率预测的只有CP-CBEmax-variants，其预测值(Z-score)为-0.19，略低于平均水平。

Table 2 BE-dot's sgRNA design scheme for correcting rs80357410¹⁾

BE	sgRNA	BE-Hive
	CTATGCAGAAAATCTTAGAGCGTCCCATCTGGTAAGTCAGC ^a	
Target-AID-NG	AGCGTCCCATCTGGTAAGTCag	NA
xBE3	TTAGAGCGTCCCATCTGGTAag	NA
SpRY-PmCDA1	GAGCGTCCCATCTGGTAAGTcag	NA
SpRY-BE4max	TTAGAGCGTCCCATCTGGTAagt	NA
SpRY-BE4max	CTTAGAGCGTCCCATCTGGTaag	NA
SpRY-BE4max	TCTTAGAGCGTCCCATCTGGtaa	NA
BE-PLUS ²⁾	AAATCTTAGAGCGTCCCATCtgg	NA
CP-CBEmax-variants ²⁾	AAATCTTAGAGCGTCCCATCtgg	-0.19

¹⁾The marked sequence contains a SNV, upstream 20 nt, downstream 20 nt. The letter in red is a mutational base. Among sgRNA bases, letters in blue represent the editing window of the corresponding BE and lowercase letters are PAM sequence recognized by BE on target site. ²⁾Synonymous correcting BEs, BEs without "b" mark are precise correcting BEs.

BE-dot纳入了CGBEs，可以实现对突变类型为C→G和G→C的SNV的纠正，使得以往不能直接利用ABEs、CBEs进行编辑的SNV有了可能的编辑方案。统计ClinVar数据库中被记录为“致病性”和“可能具有致病性”的SNV记录共43 925条，其中C:G突变为T:A的SNV占比最大，有20 918个；A:T突变为G:C的SNV占比15%，有6 410个；C:G→G:C占11%，有3 704个（图3a）。利用BE-dot对以上3种突变类型的SNV设计编辑方案，

综合精确修复和同义修复的设计结果，可以对15 724个具有和可能具有致病性的C:G→T:A突变进行纠正，A:T→G:C可编辑的有5 272个，C:G→G:C可编辑的有1 306个（图3b）。

2.2 预测BE的脱靶图谱可以指导BEs和sgRNA的选择

脱靶图谱是评估BEs的重要方面，具有高度特异性的BEs才有可能发展到临床应用。对于rs80357410设计的8个可选用的BEs，结合各自对

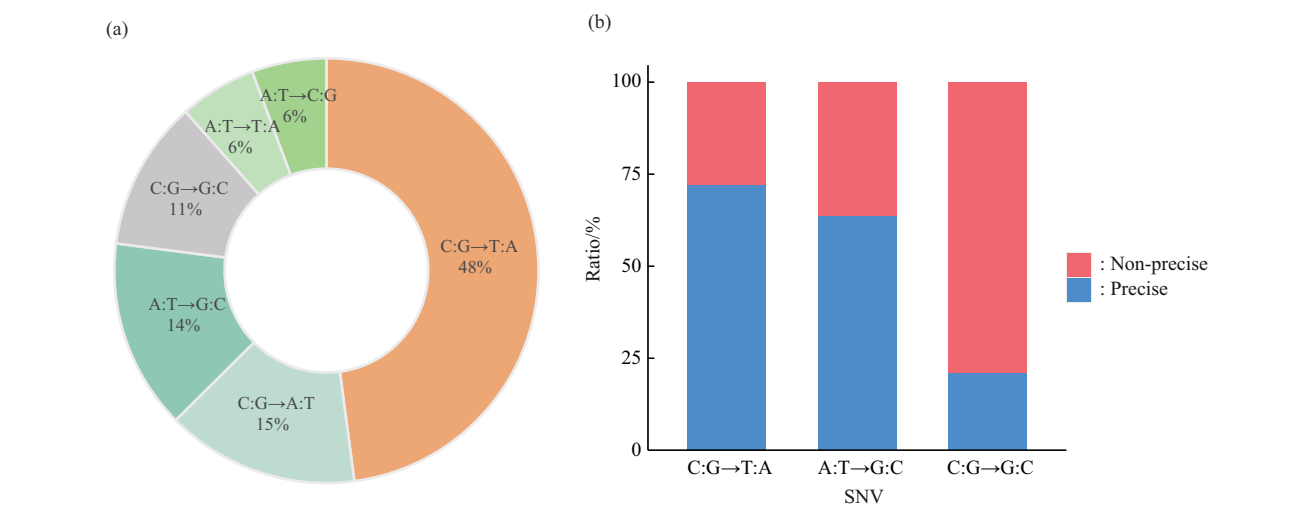


Fig. 3 Pathogenic SNV type and BEs editable SNV ratio

应的sgRNA进行脱靶数量的预测,得出的OT数量的预测结果见图4a。预测结果中,Cas-OFFinder预测的脱靶数量显著多于CALITAS预测数量(配对 t 检验, $P=0.004\ 534$)。相较于PAM序列为NG和NRN的BE类型,PAM序列为NGG的BE-PLUS和evoCDA1-BE4max在全基因组范围内预测的脱靶位点数量最少,特异性最高。在编辑效率等方面可以满足要求的情况下,应优先选择BE-PLUS或evoCDA1-BE4max。

BE-PLUS在脱靶位点处的编辑活性预测由CFD、uCRISPR、BEdeepoff给出(表3)。表3保

留了BEdeepoff预测脱靶编辑效率大于0.2的记录。由于CFD和uCRISPR只能对sgRNA和DNA长度一致的匹配做出活性预测,因此表中列出的脱靶序列来自Cas-OFFinder和CALITAS预测结果中的无DNA或RNA凸起的部分。两个软件在这4个脱靶位点处的预测得分排名是基本一致的,其中脱靶活性最高的位点位于17号染色体43 124 016~43 124 038位,与sgRNA序列比对时仅5'端第12位碱基存在错配。综合以上分析,需要特别注意BE-PLUS对应的gRNA在17号染色体43 124 016~43 124 038位的脱靶。

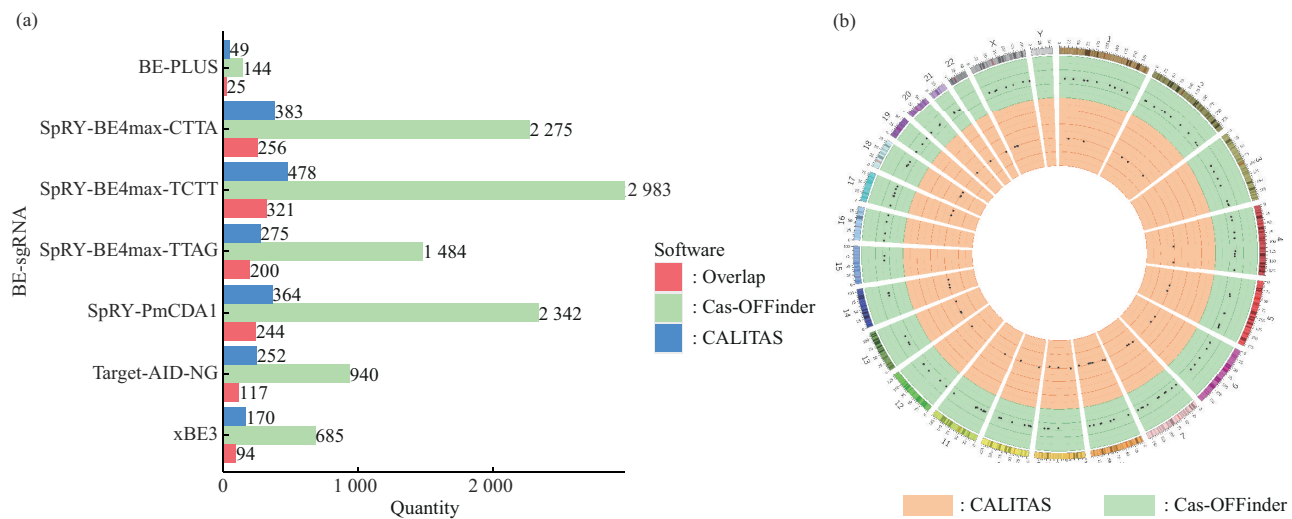


Fig. 4 OT quantity and BE-PLUS system's OT site distribution

(a) Number of OT sites predicted by Cas-OFFinder and CALITAS for each editing scheme. (b) The Circos plot shows the OT site distribution of BE-PLUS editing scheme on the reference genome (hg38).

Table 3 BE-PLUS cleavage efficiency on OT sites corresponding to different prediction tools

OT seq	Chrom	Strand	Start_position	CFD	uCRISPR	BEdeepoff
AAATCTTAGAGtGTCCCATCTGG	17	-	43 124 016	0.54	3.15	0.71
AAATCTTAcAGtGTCCaATCTGG	17	-	43 168 169	0.14	0.77	<0.2
AAATtTTAGAGgGTCCCAaAGG	14	+	78 409 864	0.14	0.61	<0.2
AAATCTTAGAGCagaCCATCTGG	21	+	20 364 525	0.00	0.38	<0.2
AgAT-TTAGtGtGTCCCATCAGG	6	+	106 973 011	NA	NA	0.29
cAgTCTTAGA-CGTCCCGTCAGG	X	-	153 703 521	NA	NA	0.29
AAATCT-AGAGttTCCCATCAGG	2	+	89 961 281	NA	NA	0.28
AAAT-cTAGAGttTCCCATCAGG	2	-	89 221 144	NA	NA	0.28

2.3 对脱靶编辑产物进行功能注释的应用举例

对选定的 BE 和 sgRNA 进行脱靶图谱的分析后，还可以利用 BE-dot 对高风险的脱靶位点进行编辑产物的分析。结合 rs80357410 示例中找到的表 3 所示的 4 个具有最高脱靶编辑活性的位点，将 Cas-OFFinder 运行结果中这 4 个位点的相应记录提取出来，作为输入文件 4ots_BE_PLUS.txt，运行命令 python BE-dot.py OTannotation -BE BE-PLUS -i 4ots_BE_PLUS.txt -o /path/，即可得到所有的脱靶

SNV 组合情况，共得到 7 条突变记录（表 4）。调用 ANNOVAR 文件做功能注释后得到发生在外显子区的脱靶 SNV 有 1 个，属于 *BRCAl* 基因位于 17 号染色体的第 43 124 035 位，是同义突变；发生在内含子和基因间区的 SNV 共有 6 个。

综合以上分析，BE-PLUS 对应的 gRNA 尽管在基因组上有个别高脱靶活性位点，但其脱靶编辑并不会对基因组或转录产物、翻译产物产生有害的影响。

Table 4 The list of all possible editing products on BE-PLUS 4 OT hot sites

Chrom	Start	End	Wt	Mut	OT site, strand [editing position]
17	43 124 035	43 124 035	G	A	CCAGATGGGACACTCTAAGATTT, -[5]
17	43 168 188	43 168 188	G	A	CCAGATTGGACACTGTAAAGATTT, -[5]
17	43 168 184	43 168 184	G	A	CCAGATTGGACACTGTAAAGATTT, -[9]
17	43 168 184	43 168 188	GTAAG	ATAAA	CCAGATTGGACACTGTAAAGATTT, - [5, 9]
21	20 364 530	20 364 530	C	T	AAATCTTAGAGCAGACCATCTGG, +[5]
21	20 364 537	20 364 537	C	T	AAATCTTAGAGCAGACCATCTGG, +[12]
21	20 364 530	20 364 537	CTTAGAGC	TTTAGAGT	AAATCTTAGAGCAGACCATCTGG, +[5, 12]

3 讨 论

随着单碱基编辑技术的广泛应用，对某一点突变位点设计特异性强的 sgRNA 已经成为普遍需求。对此，已有多个 sgRNA 设计工具问世，如 BE-designer、BE-FF、beditor。已有的可用于预测 BE 脱靶效应的工具也经历了由经典基于比对的 Cas-OFFinder 增加到基于假设、基于机器学习、基于能量等多种类型的工具 [14, 18, 26-27]。尽管相关的工具数量较多，目前仍没有工具对 sgRNA 设计、脱靶效应评估及下游的功能注释等完整过程进行整合，没有对 BEs 脱靶提供功能相同的多种工具的综合评价。

基于此，本文开发了 BE-dot，实现了仅需用户提供 SNV，即可完成 sgRNA 设计、脱靶效应预

测和脱靶产物功能注释的完整流程。为用户提供在 DNA 水平和蛋白质水平修复某一 SNV 时所有可选用的 BEs 以及可设计的 sgRNA。BE-dot 利用多个脱靶预测工具预测脱靶图谱，更容易确定脱靶热点，避免了使用单一工具的偏差。此外，BE-dot 可以自动列举脱靶位点处所有可能的编辑产物，并转换为 ANNOVAR 变异注释所要求的格式，使得用户能方便快速地获得脱靶编辑产生的功能影响。

根据现有的研究结果 [2, 28]，关于诱导产生 BEs 脱靶的因素尚无明确的结论。BE-dot 考虑的脱靶类型主要是由 sgRNA 序列相似性引起的脱靶。另外，当前版本的 BE-dot 的功能仅面向基因组 GRCh38，将来会包含多种基因组，扩大其应用的范围。

4 结 论

本研究建立了一个对 SNVs 设计单碱基编辑方案并对编辑方案的脱靶图谱进行全面评估的综合分析工具——BE-dot (<https://github.com/wendyw630/BE-dot>)。BE-dot 在设计单碱基编辑方案中包含了 27 种 CBEs 和 12 种 ABEs。BE-dot 不仅提供了在 DNA 水平上的精确修复方案, 还提供了在蛋白质水平上的同义修复方案, 旨在利用已有的 BEs 对尽可能多的 SNVs 提供编辑方案。同时, 它调用第三方工具 BE-Hive 对编辑方案的靶标编辑效率进行预测。在评估编辑方案的脱靶效应时, BE-dot 纳入了多个脱靶预测工具, 便于确定脱靶热点。此外, 为了方便用户查看脱靶编辑对基因功能的影响, BE-dot 能程序化地列举所有可能的脱靶编辑产物并分析了脱靶产物所位于的基因组区域以及对基因功能的影响。使用 BE-dot 的 sgRNA 设计功能时, 用户需要提供 SNV 的 rsID 或包含 SNV 在内共计 101 nt 的 DNA 序列及 SNV 所处的氨基酸密码子阅读框。提供 BE-sgRNA 编辑方案以及目标编辑的基因组文件即可进行脱靶位点的预测。

与其他软件相比, BE-dot 首次实现了设计 SNV 修复方案、脱靶评估以及脱靶功能影响的完整过程。对编辑方案从靶标编辑效率以及脱靶效应等方面进行了全面的评估。旨在为生物医学试验引入或修复 SNVs 设计编辑方案并对方案提供全面的评估和参考。

参 考 文 献

- [1] Zuo E, Sun Y, Wei W, *et al.* Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science*, 2019, **364**(6437): 289-292
- [2] Lei Z, Meng H, Lv Z, *et al.* Detect-seq reveals out-of-protospacer editing and target-strand editing by cytosine base editors. *Nat Methods*, 2021, **18**(6): 643-651
- [3] Wang Y, Gao R, Wu J, *et al.* Comparison of cytosine base editors and development of the BEable-GPS database for targeting pathogenic SNVs. *Genome Biol*, 2019, **20**(1): 218
- [4] Huang T P, Newby G A, Liu D R. Precision genome editing using cytosine and adenine base editors in mammalian cells. *Nat Protoc*, 2021, **16**(2): 1089-1128
- [5] Hwang G-H, Park J, Lim K, *et al.* Web-based design and analysis tools for CRISPR base editing. *BMC Bioinformatics*, 2018, **19**(1): 542
- [6] Rabinowitz R, Abadi S, Almog S, *et al.* Prediction of synonymous corrections by the BE-FF computational tool expands the targeting scope of base editing. *Nucleic Acids Res*, 2020, **48**(W1): W340-W347
- [7] Koblan L W, Arbab M, Shen M W, *et al.* Efficient C•G-to-G•C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat Biotechnol*, 2021, **39**(11): 1414-1425
- [8] Walton R T, Christie K A, Whittaker M N, *et al.* Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science*, 2020, **368**(6488): 290-296
- [9] Ren Q, Sretenovic S, Liu S, *et al.* PAM-less plant genome editing using a CRISPR-SpRY toolbox. *Nat Plants*, 2021, **7**(1): 25-33
- [10] Siegner S M, Karasu M E, Schröder M S, *et al.* PnB Designer: a web application to design prime and base editor guide RNAs for animals and plants. *BMC Bioinformatics*, 2021, **22**(1): 101-101
- [11] Dandage R, Després P C, Yachie N, *et al.* beditor: A computational workflow for designing libraries of guide RNAs for CRISPR-mediated base editing. *Genetics*, 2019, **212**(2): 377-385
- [12] Haeussler M, Schöning K, Eckert H, *et al.* Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR. *Genome Biol*, 2016, **17**(1): 148
- [13] Chen C L, Rodiger J, Chung V, *et al.* SNP-CRISPR: a web tool for SNP-specific genome editing. *G3 (Bethesda)*, 2020, **10**(2): 489
- [14] Bae S, Park J, Kim J S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, 2014, **30**(10): 1473-1475
- [15] Doench J G, Fusi N, Sullender M, *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol*, 2016, **34**(2): 184-191
- [16] Zhang D, Hurst T, Duan D, *et al.* Unified energetics analysis unravels SpCas9 cleavage activity for optimal gRNA design. *Proc Natl Acad Sci USA*, 2019, **116**(18): 8693
- [17] Zhang C, Wang D, Qi T, *et al.* BEdeepoff: an *in silico* tool for off-target prediction of ABE and CBE base editors. *bioRxiv*, 2021. <http://www.biorxiv.org/content/10.1101/2021.03.14.435296v1>
- [18] Chuai G, Ma H, Yan J, *et al.* DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biol*, 2018, **19**(1): 80
- [19] Yan J, Xue D, Chuai G, *et al.* Benchmarking and integrating genome-wide CRISPR off-target detection and prediction. *Nucleic Acids Res*, 2020, **48**(20): 11370-11379
- [20] Arbab M, Shen M W, Mok B, *et al.* Determinants of base editing outcomes from target library analysis and machine learning. *Cell*, 2020, **182**(2): 463-480.e430
- [21] Den Dunnen J T, Dalgleish R, Maglott D R, *et al.* HGVS recommendations for the description of sequence variants: 2016 update. *Hum Mutat*, 2016, **37**(6): 564-569
- [22] Fennell T, Zhang D, Isik M, *et al.* CALITAS: A CRISPR-Cas-aware aligner for *in silico* off-target search. *CRISPR J*, 2021, **4**(2): 264-274
- [23] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, 2010, **38**(16): e164
- [24] Findlay G M, Daza R M, Martin B, *et al.* Accurate classification of BRCA1 variants with saturation genome editing. *Nature*, 2018, **562**(7726): 217-222
- [25] Rebbeck T R, Friebe T M, Friedman E, *et al.* Mutational spectrum in a worldwide study of 29,700 families with BRCA1 or BRCA2 mutations. *Hum Mutat*, 2018, **39**(5): 593-620
- [26] Cancellieri S, Canver M C, Bombieri N, *et al.* CRISPRitz: rapid, high-throughput and variant-aware *in silico* off-target site identification for CRISPR genome editing. *Bioinformatics*, 2020, **36**(7): 2001-2008
- [27] Alkan F, Wenzel A, Anthon C, *et al.* CRISPR-Cas9 off-targeting assessment with nucleic acid duplex energy parameters. *Genome Biol*, 2018, **19**(1): 177
- [28] McGrath E, Shin H, Zhang L, *et al.* Targeting specificity of APOBEC-based cytosine base editor in human iPSCs determined by whole genome sequencing. *Nat Commun*, 2019, **10**(1): 5353

BE-dot: a Tool for sgRNA Design and Off-target Profile Prediction of Base Editing*

WANG Ze-Lu, LIANG Jun-Bo**, WANG Xiao-Yue**

(Department of Biochemistry and Molecular Biology, Institute of Basic Medical Sciences CAMS, Institute of Basic Medicine PUMC, Beijing 100005, China)

Abstract Objective As a powerful tool in correcting genomic point mutations, base editors (BEs) show a promising prospect for biotechnology development and therapeutic applications. While editing the target single-nucleotide variant (SNV), it's primary to select competent BEs and design single guide RNA (sgRNA). Currently, although there are multiple sgRNA design tools, no tools are available for integrating the design of sgRNA with the assessment of the specificity of BEs. **Methods** 27 cytosine base editors (CBEs) and 12 adenine base editors (ABEs) were used to design base editing schemes. BE-Hive, a third-party tool, was provoked to predict the editing efficiency. The off-target profiles of editing schemes were evaluated by using a combination of multiple off-target prediction tools. Finally, all possible off-target editing products were calculated by considering both base editor types and off-target sites, and then ANNOVAR, a variant annotation tool, was called for functional analysis of off-target products. **Results** We propose a comprehensive tool, BE-dot, which enables the complete process from a given SNV to designing sgRNAs, predicting off-target profiles, and annotating off-target products' functions. Besides providing precise correction schemes at DNA level, in order to expand the range of editable SNVs, BE-dot can perform synonymous corrections at protein level by degeneracy. When predicting off-target profiles of single base editing systems, BE-dot integrates multiple tools such as Cas-OFFinder, CALITAS, CFD, uCRISPR and BEdeepoff, which allows BE-dot to evaluate the specificity of single base editing systems more comprehensively and provide users with consultations about BEs and sgRNA selection. In addition, BE-dot can automatically analyze all possible editing products at off-target sites, and convert them into avinput format for functional annotation by ANNOVAR, avoiding the tedious manual annotation. **Conclusion** BE-dot designs editing schemes for applying base editing to correct or introduce SNVs, and comprehensively evaluates the editing scheme in terms of editing efficiency, off-target profile, and off-target functional impact.

Key words base editor, sgRNA design, off-target profile prediction

DOI: 10.16476/j.pibb.2022.0046

* This work was supported by a grant from The National Natural Science Foundation of China (32070603).

** Corresponding author.

LIANG Jun-Bo. Tel: 86-10-69156413, E-mail: liangjunbo@ibms.pumc.edu.cn

WANG Xiao-Yue. Tel: 86-10-69156413, E-mail: pumcwangxy@163.com

Received: February 10, 2022 Accepted: May 6, 2022