

DOI:10.12113/202201013

BEguider: 一个单碱基编辑器 sgRNA 设计与编辑效率预测工具

高靖静, 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:单碱基编辑器是实用且高效的基因编辑工具,其编辑效率与单向导 RNA (single guide RNA, sgRNA) 序列的设计密切相关。目前单碱基编辑器 sgRNA 序列的设计缺少特定的法则,主要依靠经验和大量尝试完成。本研究基于卷积神经网络,开发了一个单碱基编辑器 sgRNA 序列设计工具 BEguider。BEguider 利用 TensorFlow 2 深度学习框架建立编辑效率预测模型,能够在人基因组范围内针对 NGG PAM 序列依赖的单碱基编辑器 ABE7.10-NGG 和 BE4-NGG 批量设计 sgRNA 序列,预测编辑效率。此外,通过整合 Cas-OFFinder, BEguider 能够提供对 sgRNA 脱靶情况的评估。利用 BEguider 设计 sgRNA 序列,有助于研究人员提高实验效率,节约实验成本。

关键词:CRISPR/Cas9; 单碱基编辑器; sgRNA 设计; 编辑效率; 卷积神经网络

中图分类号:Q78, TP391 文献标志码:A 文章编号:1672-5565(2023)02-106-08

BEguider: A base editor for sgRNA design and editing efficiency prediction

GAO Jingjing, WANG Xiaoyue*

(Department of Biochemistry and Molecular Biology, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing 100005, China)

Abstract: Base editors are practical and efficient gene editing tools, whose editing efficiencies often depend on the design of single guide RNA (sgRNA) sequences. At present, the design of sgRNA libraries lacks of specific rules and mainly relies on experience and attempts. On the basis of the convolutional neural network, BEguider was developed for sgRNAs design of base editors. BEguider used the deep learning framework TensorFlow 2 to build editing efficiency prediction models, which could design sgRNA sequences and predict editing probabilities for NGG PAM-dependent base editor variants ABE7.10-NGG and BE4-NGG within the scope of the human genome. Besides, BEguider could evaluate potential off-target sites of sgRNAs by integrating Cas-OFFinder. Using BEguider to design sgRNA sequences will facilitate future application of base editors and save experimental cost.

Keywords: CRISPR/Cas9; Base editor; sgRNA design; Editing efficiency; CNN

目前,在已知致病基因中约一半的变异为单核苷酸变异 (Single-nucleotide Variants, SNVs)^[1],这意味着开发高效校正 SNVs 的方法,具有广泛的科学研究价值和遗传病基因治疗潜力。CRISPR/Cas9 系统是目前广泛应用的基因编辑工具之一,但其主要通过切割 DNA 双链,借助细胞中 DNA 双链断裂修复机制来编辑基因,难以精确校正单个核苷酸位点的突变。单碱基编辑系统 (Base Editor, BE) 是在 CRISPR/Cas9 系统的基础上开发得到的新型碱基编

辑工具,主要有胞嘧啶碱基编辑器 (C to T) 和腺嘌呤碱基编辑器 (A to G) 两类。单碱基编辑系统能够实现单核苷酸水平的精准突变,其不但具有高效性和特异性,并且在细胞、小鼠模型和植物的基因组中产生插入缺失的比例一般低于 1%^[2-5]。因此,单碱基编辑系统在研究等位基因位点对基因功能影响,SNV 相关遗传病的致病机制和治疗方案等方面具有独特优势。

影响单碱基编辑系统编辑效率的因素包括靶基

收稿日期:2022-01-21;修回日期:2022-05-04;网络首发日期:2022-09-08.

网络首发地址:https://kns.cnki.net/kcms/detail/23.1513.Q.20220907.1145.002.html

基金项目:国家自然科学基金项目 (No.32070603).

* 通信作者:王晓月,女,教授,博士生导师,研究方向:癌症基因组学. E-mail: pumcwangxy@163.com.

因的选择、sgRNA 序列设计、脱靶效应和单碱基编辑系统向细胞内的递送方式等,sgRNA 序列设计是其中的关键因素之一^[6]。有研究报道,sgRNA 序列活性与 Cas9 蛋白序列偏好性^[7-9],以及脱氨酶的序列偏好性^[5, 10-12]有着密切联系。此外,若基因组上存在与设计的 sgRNA 序列相似度较高的片段,容易发生脱靶效应,在基因组靶位点之外产生编辑。因此,优化 sgRNA 设计有助于提高编辑效率,降低脱靶情况的发生。

sgRNA 设计工具的实现方式主要有三大类:一是基于序列比对的方法,即 sgRNA 和基因组的比对结果;二是基于概率假设的方法,即利用 sgRNA 上的一些特征(如 GC 含量)进行设计;三是基于学习的方法,利用大型数据集训练机器学习和深度学习模型来设计 sgRNA。目前针对单碱基编辑系统的 sgRNA 设计工具主要基于序列比对和 sgRNA 序列特征,且无法预测编辑效率。

本研究使用 Song 等人^[13]发表的高通量单碱基编辑系统测序数据,利用卷积神经网络(Convolutional Neural Network, CNN)模型,结合 Cas-OFFinder 脱靶位点预测工具^[14],开发了一个 sgRNA 序列设计工具 BEguilder。该工具能针对人基因序列进行 sgRNA 设计、靶向编辑效率预测以及寻找 sgRNA 在基因组上可能的脱靶位点。

1 数据与方法

1.1 数据

本研究中,单碱基编辑器 BE4-NGG 和 ABE7.10-NGG 的 CNN 模型的训练数据来自于 Song 等人^[13]发表的高通量外源递送系统测序数据。数据被划分为训练集,验证集和测试集,比例为 8:1:1。

内源性位点数据来自于 NCBI BioProject 数据库的公开项目 PRJNA476544。

人类 hg38 参考基因组序列来自于 Ensembl 数据库。

1.2 方法

1.2.1 BEguilder 分析预测流程

首先,对输入的基因序列进行预处理,筛选候选 sgRNA 序列。分别在输入基因上的正链与负链上寻找是否存在 NGG PAM 序列。定位到 NGG PAM 序列后,向序列 5' 方向搜索,寻找编辑活性窗口内(即 sgRNA 第 4 到 8 位碱基)是否存在可编辑的底物核苷酸——A 或 C。若存在,则提取 PAM 序列前的 20 nt 核苷酸序列与 4 nt PAM 序列作为候选 sgRNA 序列。

其次,对于候选的 sgRNA 进行独热编码(One-hot Encoding)。为了使计算机能够识别并处理 sgRNA 序列,需要对每条序列进行独热编码,碱基 A 用向量[1, 0, 0, 0]表示,碱基 C 用向量[0, 1, 0, 0]表示,碱基 G 用向量[0, 0, 1, 0]表示,碱基 T 用向量[0, 0, 0, 1]表示。因此,每条长为 L nt 的 sgRNA 的序列可以被表示为一个 $L \times 4$ 的矩阵 $X_{input} \in R^{L \times 4}$ 。

最后,对候选 sgRNA 序列的编辑效率进行预测并输出。CNN 模型优势在于能够很好地学习训练样本的局部特征,达到准确的预测效果。神经网络会对输入的每个 sgRNA 独热编码矩阵进行预测分析,经过两层二维卷积层与两层全连接层处理,输出对应的预测概率(见表 1)。在卷积层与全连接层之间,以及两层全连接层之间各有一层随机失活层,随机失活比例为 0.3,防止 CNN 在较小的数据集上发生过拟合现象。

CNN 模型的评价指标有均方误差 MSE, Pearson 相关系数, Spearman 相关系数。

全流程如图 1c 所示。

表 1 BEguilder 中的卷积神经网络结构

Table 1 CNN structure of BEguilder

Layers	Filters	Kernel Size	Stride
Conv2D_1	128	(3, 4)	(1, 2)
Conv2D_2	256	(3, 4)	(1, 2)
Dense_1	256	/	/
Dense_2	1	/	/

1.2.2 BEguilder 的安装与使用

BEguilder 的代码已上传至 gitee 仓库,可通过 git clone 命令下载。使用 BEguilder 前,需要预先安装运行依赖的深度学习框架 TensorFlow 2 与第三方 python 程序包,建议通过开源软件包管理系统 Conda 进行安装(requirements_BEguilder.txt 已在软件中提供)。命令如下:

```
git clone https://gitee.com/gaojingjing123/beguilder
```

```
conda create --name <env> --file requirements_BEguilder.txt
```

BEguilder 支持三种输入形式(见图 2):

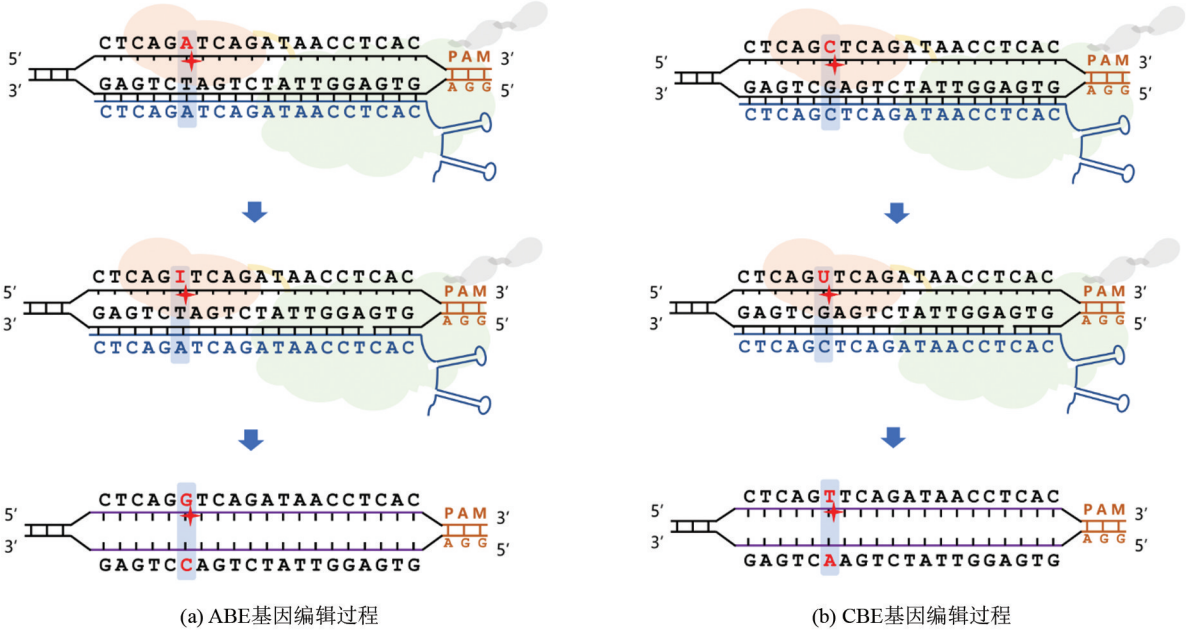
- (1) 输入基因名或序列 ID 与对应的基因序列。
- (2) 输入靶位点所在的染色体号与染色体坐标。
- (3) 输入靶位点的 rsID。

对于 ABE7.10-NGG 和 BE4-NGG, BEguilder 支持两种 sgRNA 设计模式(见图 2):

- (1)指定单碱基编辑器模式:输入编辑器名称,指定 ABE7.10-NGG 或 BE4-NGG 作为编辑工具。
- (2)ALL 模式:BEguider 给出最佳的单碱基编

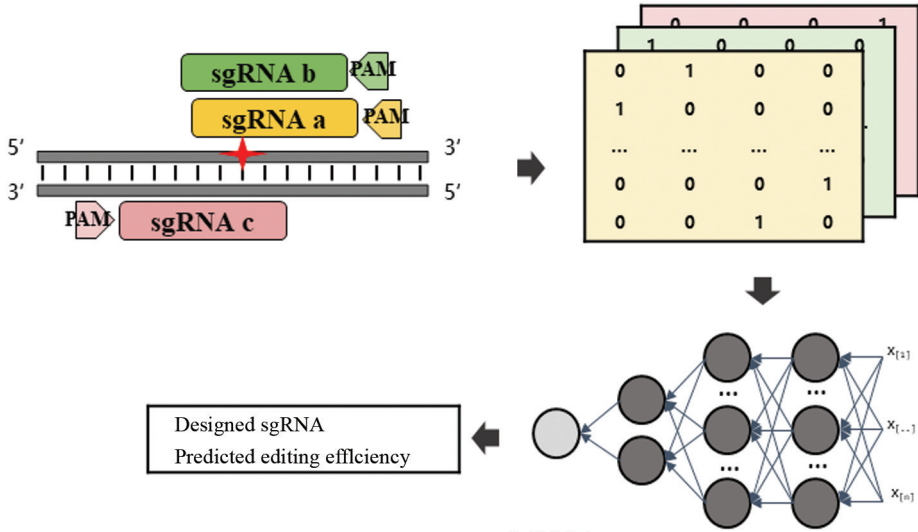
辑器与 sgRNA 组合。

BEguider 整合 Cas-OFFinder 脱靶位点搜索工具,提供可选的预测脱靶位点功能。



(a) ABE基因编辑过程

(b) CBE基因编辑过程



(c) BEguider工作流程图

图 1 单碱基编辑器编辑示意图与 BEguider 工作流程图

Fig. 1 Editing process of base editors and the workflow of BEguider

注:(a) ABE 将腺嘌呤脱氨得到次黄嘌呤,细胞内 DNA 复制时次黄嘌呤会与胞嘧啶配对,经 DNA 复制,A:T 配对被替换为 G:C 配对;(b) CBE 将胞嘧啶脱氨得到尿嘧啶,细胞内 DNA 复制时尿嘧啶会与腺嘌呤配对,经 DNA 复制,C:G 配对被替换为 T:A 配对(浅绿色为 Cas9n 核酸酶,淡橘色为 DNA 脱氨酶,浅灰色为尿嘧啶糖基化酶抑制剂,土黄色为 DNA 上的 PAM 序列,红色为点突变位点,蓝色为 sgRNA 序列,紫色为新合成的 DNA 双链);(c) sgRNA a,sgRNA b 和 sgRNA c 表示针对同一个基因位点可以设计多条 sgRNA。

1.2.3 BEguider 在内源性位点的预测效果评估

将内源性 sgRNA 序列输入 BEguider,得到预测编辑效率,用 Pearson 相关系数和 Spearman 相关系数评估预测编辑效率与实验编辑效率之间的相关性。其中,ABE7.10-NGG 的 sgRNA 序列有 94 条, BE4-NGG 的 sgRNA 序列有 102 条。

2 结果

2.1 BEguider 神经网络模型训练结果

BEguider 的训练数据来自于外源靶序列,即慢病毒将靶序列整合进宿主细胞基因组,再由单碱基

编辑器对靶位点编辑,最后 PCR 扩增出靶位点片段进行高通量测序获得编辑结果。BEguider 中包括两个 CNN 模型,分别为 ABE7.10-NGG 和 BE4-NGG 提供预测功能。图 3 为 BEguider 在训练集与验证集上 MSE 随训练轮次的变化。在模型训练过程中验证集的数据特征可能会泄露到模型中,因此模型评价需要在模型从未接触的测试集数据中进行。在测试集

中 ABE7.10-NGG 的 MES 为 0.007 8, Spearman 相关系数为 0.901 3, Pearson 相关系数为 0.904 5 (见图 4a); BE4-NGG 的 MES 为 0.001 9, Spearman 相关系数为 0.853 1, Pearson 相关系数为 0.867 3 (见图 4b)。这说明 BEguider 能够较好地预测未知 sgRNA 的编辑效率。

```
python BEguider.py -h
usage: BEguider.py [-h] (-g GENES | -c CHROMOSOME | -s RSID) -b BASEEDITOR [-f OFFTARGET]
                    [-m MISMATCH] [-o OUTPUT]

Program: BEguider
Version: 0.1
Author : Jingjing Gao
Email  : gaojingjing@ibms.pumc.edu.cn

Optional Base Editors: ['ABE7.10-NGG', 'BE4-NGG']

The pairings of Base Editors and SNPs: CBE -- C>T; G>A   ABE -- A>G; T>C

Examples:
  1) use genes as input file:
      Genes,Seqs
      BRCA1,TGGCTGAAGAATTGCTAAGC
      AATCAGGAAAGCTGGTGG
      ARIDIA,CTCCACCGAAGGGAGGACCC
      ACTGCCCCAGCCGGGGTCTCG
  2) use chromosomes and coordinates as input file:
      Chrom,Coordinate,Type
      chr1,145634,r
      chrX,87632,a
  3) use rsIDs as input file:
      SNP,Type
      rs5297,r
      rs12603332,a

Optional Arguments:
-h, --help          show this help message and exit
-g GENES, --genes GENES
                    A txt file including genes and sequences separated
                    by comma(.).
-c CHROMOSOME, --chromosome CHROMOSOME
                    A txt file including chromosomes, coordinates and
                    genetic type separated by comma(.). Genetic type: 'r' means editing
                    wild genes, 'a' means editing mutant genes.
-s RSID, --rsID RSID A txt file including rsID and genetic type
                    separated by comma(.). Genetic type: 'r' means editing wild genes, 'a'
                    means editing mutant genes.
-b BASEEDITOR, --BaseEditor BASEEDITOR
                    Base Editors: ALL / A Specific BE name. Design
                    sgRNA for a specific BE in optional BEs or for both BEs.
-p PAM, --pam PAM   PAM sequence: NGG. (Default = NGG)
-f OFFTARGET, --oftarget OFFTARGET
                    True / False. True: predict off-target sites in hg38
                    genome. (Default = False)
-m MISMATCH, --mismatch MISMATCH
                    Allowed maximum mismatch site between sgRNA
                    and genome while searching for off-target sites. (Default = 3)
-o OUTPUT, --output OUTPUT
                    Output directory. (Default = current directory)
```

图 2 BEguider 的命令行截图说明

Fig. 2 Screenshot for commands of BEguider

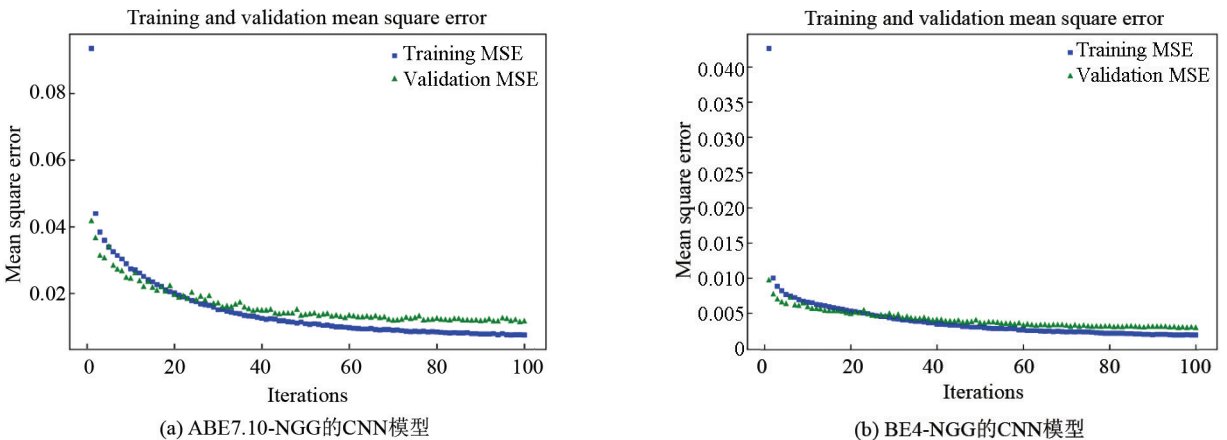
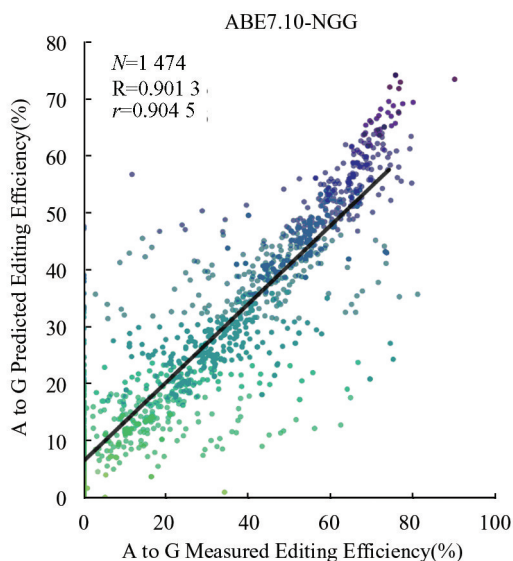


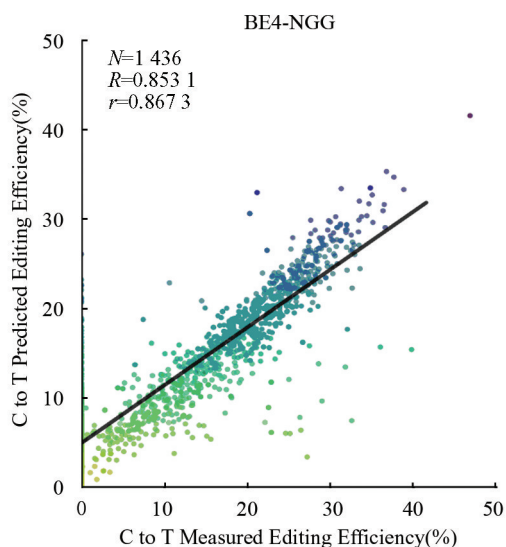
图 3 BEguider 训练集与验证集上的 MSE

Fig. 3 MSE of BEguider on training set and validation set

注:(a)、(b) 横坐标表示训练轮次,纵坐标表示每个轮次结束计算的 MSE



(a) ABE7.10-NGG的CNN模型



(b) BE4-NGG的CNN模型

图4 BEguilder在测试集上预测结果与测序结果的相关性

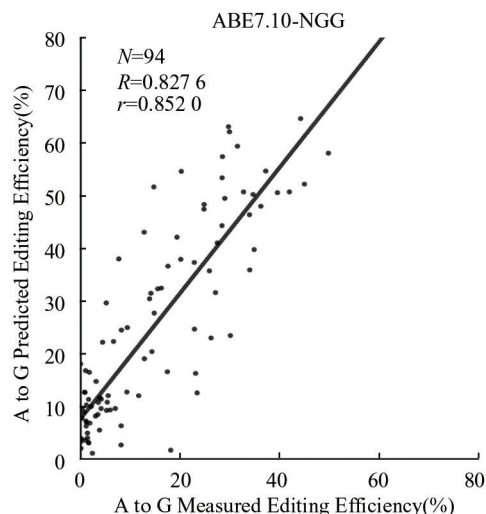
图4 Correlation between predicted efficiencies of BEguilder and measured efficiencies on the test set

注:(a)、(b)横坐标表示测序得到的编辑效率,纵坐标表示BEguilder预测的编辑效率。N为纳入评估的sgRNA的数量,R为Spearman相关系数,r为Pearson相关系数。

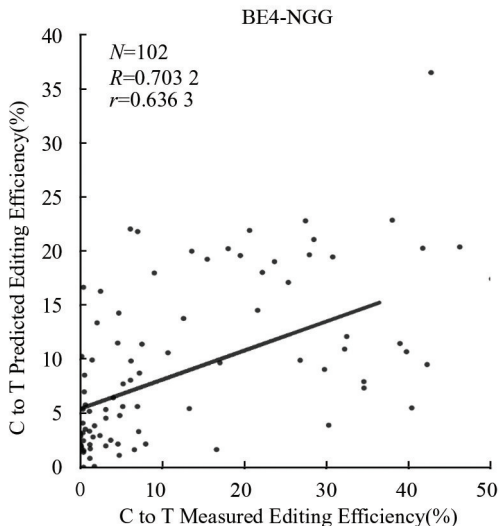
2.2 BEguilder预测内源性位点编辑效率的评估结果

为了扩大BEguilder的应用范围,需要评估BEguilder对内源基因组位点编辑效率预测的效果。对于ABE7.10-NGG,BEguilder的预测编辑效率与实验编辑效率的Spearman相关系数0.8276,Pearson相关系数为0.852,MSE为0.021(见图5a)。对于BE4-NGG,BEguilder的预测编辑效率与实验编辑效率的Spearman相关系数0.7032,Pearson相关系数为0.6363,MSE为0.0264(见图5b)。说明BEguilder能够准确提取sgRNA序列特征,建立sgRNA序列组

成与编辑效率之间的联系,对于内源性位点的编辑效率也有较好的预测效果。



(a) BEguilder对ABE7.10-NGG的预测效果



(b) BEguilder对BE4-NGG的预测效果

图5 BEguilder对内源性位点预测编辑效率的效果评估

图5 Evaluation of the predictive performance of BEguilder at endogenous sites

注:N为纳入评估的内源性位点数量,R为Spearman相关系数,r为Pearson相关系数。

2.3 BEguilder设计sgRNA的流程与结果

以输入形式为rsID为例,选取10个SNV位点,将rsID输入BEguilder,使用推荐单碱基编辑器与sgRNA组合的模式,默认不开启脱靶位点预测,在linux系统中运行命令如下:

```
python BEguilder.py -s SNP_example.txt -b ALL -o ./temp.
```

BEguilder一共设计出12条sgRNA,输出信息包括单碱基编辑器、SNV的相关信息、sgRNA所在的正负链情况、设计的sgRNA序列以及对应的预测编辑效率。结果如表2所示。

表 2 以 10 个 SNV 位点为例的 sgRNA 设计结果
Table 2 Output results of sgRNAs for ten SNVs using BEguider

Base-Editor	SNP-Site	Strand	Designed-sgRNA	PAM	Pred-Efficiency
ABE7.10-NGG	rs1567488305_chr16:56334784	+	GCAGAACATCCTCCGAACCA	GGGT	0.559 0
ABE7.10-NGG	rs888485088_chr16:59592	-	TGGGACCAACAAAGGTGGGT	AGGT	0.534 7
ABE7.10-NGG	rs1567488305_chr16:56334784	+	AGCAGAACATCCTCCGAACC	AGGG	0.355 2
ABE7.10-NGG	rs936032752_chr6:87084	-	CAGCAATAACACCTGCAGAC	TGGT	0.312 1
ABE7.10-NGG	rs368024573_chr12:21064	+	CTCAAGAAGGTAATGAGGA	TGGT	0.254 0
ABE7.10-NGG	rs183323267_chr12:140200	+	TCAAGAGACTCTATTCTAG	GGGA	0.232 8
BE4-NGG	rs1812418017_chr8:116854264	+	TTATTCCACAAAGGCTGAGC	AGGT	0.204 7
BE4-NGG	rs1001551731_chr16:59616	+	CCCCCCAACCCACATCCT	TGGT	0.162 4
BE4-NGG	rs571345175_chr12:49632	-	GCGCCACTGCACTCCAGCCT	GGGT	0.107 5
BE4-NGG	rs140218451_chr20:89849	+	CTTTCTCTTTTATTTTGGAG	AGGT	0.093 7
BE4-NGG	rs571345175_chr12:49632	-	TGCGCCACTGCACTCCAGCC	TGGG	0.088 4
BE4-NGG	rs1006009587_chr20:75692	+	ACTACGCTGACATAATCAAA	TGGT	0.075 9

注: '+' :表示模板链; '-' :表示互补链;将输入的单链作为模板链.

2.4 与其它单碱基编辑器 sgRNA 设计的比较

将 BEguider 与 5 种目前可用的碱基编辑设计工具进行比较 (见表 3)。Benchling、BE-Designer 和 beditor 主要功能是在指定的单碱基编辑器与点突变上给出编辑结果,而 BEable-GPS、BE-FF 能找到合适

的碱基编辑器来还原给定点突变的变化。BEguider 的优势在于提供了靶序列的多种输入方式,能够根据需要选择编辑野生型基因位点或突变型基因位点,同时批量为多个基因位点设计 sgRNA,预测每条 sgRNA 的编辑效率。

表 3 多个 sgRNA 设计工具的比较
Table 3 Comparison of different sgRNA design tools

Tool	Supported BEs	Input format	Suitable sgRNA sequences	Recommend BEs	Predict editing efficiency	Multiple SNVs analysis	Off-target site assessment
BEguider	ABE and CBE	target sequences/ genomic coordinates/ rsID	Y	Y	Y	Y	Y
Benchling (www.benchling.com)	CBE only	DNA sequence file	Y	N	N	N	Y (CBE-NGG only)
beditor ^[15]	CBE and ABE	genomic coordinates and amino acid information	Y	N	N	Y	Y
BE-Designer ^[16]	CBE and ABE	target sequences	Y	N	N	Y	Y
BE-FF ^[17]	CBE and ABE	genomic coordinates/ rsID	Y	Y	N	Y	Y (CBE-NGG only)
BEable-GPS ^[18]	CBE only	target sequences / rsID	Y	Y (CBE only)	N	N	N

注: Y: yes; N: no.

3 讨论

传统的基因编辑工具,如锌指蛋白核酸酶(Zinc Finger Nuclease, ZFN),转录激活样效应因子核酸酶(Transcription Activator-Like Effector Nuclease, TALEN)等,需要通过改造蛋白结构域才能在新的靶位点进行编辑,过程复杂,而单碱基编辑系统仅需要合成新 sgRNA 序列即可靶向新的基因位点,因此单碱基编辑系统的高效性、精确性、稳定性和易操作性,使其成为目前最为流行的基因编辑工具之一。

单碱基编辑系统提高了在细胞内功能性 SNVs 位点筛选与验证的效率,为 SNV 相关疾病的机制研究和治疗方案的探索提供了新工具和新方法。例如,Weng 等人^[19]利用单碱基编辑系统从染色体 17q21 上 136 个 SNVs 位点中筛选到了功能 SNP 位点 rs12603332。在肠类器官中,ABE 能够校正与囊性纤维化高度相关的 *CFTR* 基因的无义突变,编辑效率最高为 9.3%^[20]。在小鼠模型中,CBE 能够校正致聋基因 *TMC1* 的点突变,编辑效率最高为 51%^[21]。

具有高活性的 sgRNA 序列是使用单碱基编辑器进行研究的基础。BEguider 利用 CNN 神经网络,学习了大量单碱基编辑器序列偏好特征与 sgRNA 编辑活性的序列特征,能够准确预测所设计的 sgRNA 编辑活性。此外,对于由其它支持 ABE7.10-NGG 和 BE4-NGG sgRNA 序列设计工具得到的 sgRNA 序列,只要将其整理为 BEguider 要求的输入格式,也可输入 BEguider 进行编辑效率预测与脱靶情况评估。

目前,BEguider 还存在一定局限性。由于数据集的限制,BEguider 仅支持 ABE7.10-NGG 和 BE4-NGG 的 sgRNA 设计,且只能设计 PAM 序列为 NGG 的 sgRNA。此外,以染色体坐标或 rsID 作为输入时,BEguider 需在线从 NCBI 获取数据,因此在设计大量 SNVs 时运行速度会受到网速、NCBI 响应速度等因素的限制。未来,BEguider 将继续整合更多类型的单碱基编辑器数据,为用户提供更多单碱基编辑器的选择,并且继续优化提高 sgRNA 设计效率。

4 结论

本研究利用 CNN 模型,开发了针对 ABE7.10-NGG 和 BE4-NGG 的 sgRNA 序列设计工具 BEguider,支持 Windows、Linux 和 MacOS 操作系统。BEguider 能够同时为多个人类基因位点设计 sgRNA,并提供 3 种输入格式,2 种设计模式与可选

的脱靶评估功能。用户可以指定 ABE7.10-NGG 或 BE4-NGG,也可由 BEguider 提供最佳的单碱基编辑器与 sgRNA 序列组合,并给出预测的编辑效率。BEguider 能够大大简化 sgRNA 的设计流程。随着单碱基编辑器的应用越来越广泛,简便有效的 sgRNA 设计方案能够有利于提高实验效率,帮助研究人员更好地利用单碱基编辑器作为研究工具。

参考文献(References)

- [1] SUN H, YU G. New insights into the pathogenicity of non-synonymous variants through multi-level analysis[J]. *Scientific Reports*, 2019, 9(1): 1667–1678. DOI: 10.1038/s41598-018-38189-9.
- [2] RYU S M, KOO T, KIM K, et al. Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy [J]. *Nature Biotechnology* 2018, 36(6): 536–539. DOI: 10.1038/nbt.4148.
- [3] YE H W, CHIANG H, REES H A, et al. In vivo base editing of post-mitotic sensory cells[J]. *Nature Communications*, 2018, 9(1): 2184–2193. DOI: 10.1038/s41467-018-04580-3.
- [4] KANG B C, YUN J Y, KIM S T, et al. Precision genome engineering through adenine base editing in plants [J]. *Nature Plants*, 2018, 4(7): 427–431. DOI: 10.1038/s41477-018-0178-x.
- [5] ARBAB M, SHEN M W, MOK B, et al. Determinants of base editing outcomes from target library analysis and machine learning[J]. *Cell*, 2020, 182(2): 463–480. DOI: 10.1016/j.cell.2020.05.037.
- [6] SCHATOFF E M, ZAFRA M P, DOW L E. Base editing the mammalian genome [J]. *Methods*, 2019, 164–165: 100–108. DOI: 10.1016/j.ymeth.2019.02.022.
- [7] KLEINSTIVER B P, PATTANAYAK V, PREW M S, et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects [J]. *Nature*, 2016, 529(7587): 490–495. DOI: 10.1038/nature16526.
- [8] SLAYMAKER I M, GAO L, ZETSCHE B, et al. Rationally engineered Cas9 nucleases with improved specificity [J]. *Science*, 2016, 351(6268): 84–88. DOI: 10.1126/science.aad5227.
- [9] CASINI A, OLIVIERI M, PETRIS G, et al. A highly specific SpCas9 variant is identified by in vivo screening in yeast[J]. *Nature Biotechnology* 2018, 36(3): 265–271. DOI: 10.1038/nbt.4066.
- [10] LIU L D, HUANG M, DAI P, et al. Intrinsic nucleotide preference of diversifying base editors guides antibody ex vivo affinity maturation[J]. *Cell Reports*, 2018, 25(4): 884–892. DOI: 10.1016/j.celrep.2018.09.090.
- [11] GEHRKE J M, CERVANTES O, CLEMENT M K, et al.

- An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities [J]. *Nature Biotechnology*, 2018, 36(10): 977-982. DOI: 10.1038/nbt.4199.
- [12] WANG X, LI J, WANG Y, et al. Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion [J]. *Nature Biotechnology*, 2018, 36(10): 946-949. DOI: 10.1038/nbt.4198.
- [13] SONG M, KIM H K, LEE S, et al. Sequence-specific prediction of the efficiencies of adenine and cytosine base editors[J]. *Nature Biotechnology*, 2020, 38(9): 1037-1043. DOI: 10.1038/s41587-020-0573-5.
- [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 [J]. *Bioinformatics*, 2014, 30(10): 1473-1475. DOI: 10.1093/bioinformatics/btu048.
- [15] DANDAGE R, DESPRES P C, YACHIE N, et al. beditor: A computational workflow for designing libraries of guide RNAs for CRISPR-mediated base editing [J]. *Genetics*, 2019, 212(2): 377-385. DOI: 10.1534/genetics.119.302089.
- [16] HWANG G H, PARK J, LIM K, et al. Web-based design and analysis tools for CRISPR base editing[J]. *BMC Bioinformatics*, 2018, 19(1): 542-548. DOI: 10.1186/s12859-018-2585-4.
- [17] 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 [J]. *Nucleic Acids Research*, 2020, 48(W1): W340-W347. DOI: 10.1093/nar/gkaa215.
- [18] WANG Y, GAO R, WU J, et al. Comparison of cytosine base editors and development of the BEable-GPS database for targeting pathogenic SNVs[J]. *Genome Biology*, 2019, 20(1): 218-224. DOI: 10.1186/s13059-019-1839-4.
- [19] WENG N, MILLER M, PHAM A K, et al. Single-base editing of rs12603332 on Chromosome 17q21 with a Cytosine Base Editor regulates ORMDL3 and ATF6alpha expression[J]. *Allergy*, 2021, 77(4): 1139-1149. DOI: 10.1111/all.15092.
- [20] GEURTS M H, DE POEL E, AMATNGALIM G D, et al. CRISPR-based adenine editors correct nonsense mutations in a cystic fibrosis organoid biobank[J]. *Cell Stem Cell*, 2020, 26(4): 503-510. DOI: 10.1016/j.stem.2020.01.019.
- [21] YE H W, SHUBINA-OLEINIK O, LEVY J M, et al. In vivo base editing restores sensory transduction and transiently improves auditory function in a mouse model of recessive deafness[J]. *Science Translational Medicine*, 2020, 12(546): 9101-9112. DOI: doi.org/10.1126/scitranslmed.aay9101.