百奥迈科生物技术有限公司

Biomics Biotechnologies Co., Ltd

Biomics CRISPR/sgRNA 设计程序

Biomics sgRNA design protocol　　Frank 20140315

CRISPR技术潜在脱靶效率较TALEN、ZFN等其它基因编辑工具较高，因此关键是设计特异性sgRNA序列。CRISPR技术直至去年才迎来技术应用的爆发期，因此其生物信息学相关软件开发应用一直滞后，这阻碍了整个技术应用领域的快速发展。目前sgRNA的在线设计网站有 1.Zhangfeng lab：http://crispr.mit.edu/；2. Jack Lin's lab CRISPR/Cas9 gRNA finder: http://spot.colorado.edu/~slin/cas9.html； Boutros lab: http://www.e-crisp.org/E-CRISP/; Zinc Finger Consortium: http://zifit.partners.org/ZiFiT/ChoiceMenu.aspx等网站。这些网站虽然一定程度上满足了部分研究者的需要，但仍然存在很多不足和缺陷。例如MIT等网站等待分析结果需要数日，E-CRISP网站的分析结果不提供具体的脱靶数据，ZIFI网站可供选择物种十分有限，jack Lin的网站则几乎要靠自己一个一个收集脱靶信息。上述国外网站不仅不能囊括所有的物种信息，而且由于国内网络不稳定经常不能登录成功，极大的影响了科研工作者的工作效率。这些问题给设计者带来了一定困难，鉴于此BIOMICS及时推出了本地化运行sgRNA软件设计服务。当日提交设计订单，第二工作日给出sgRNA设计结果及，不限物种。Biomics提供的分析报告有详细的off-target数据，客户可以根据要求定制，从而自主选择靶点，提高科研工作效率！

1.客户提供敲除靶基因信息或自己已经设计的靶点序列信息

Oryctolagus cuniculus apolipoprotein E (APOE), mRNA

NCBI Reference Sequence: NM_001082643.1

GenBank Graphics

>gi|130488074|ref|NM_001082643.1| Oryctolagus cuniculus apolipoprotein E (APOE), mRNA

CCCAGCCGAGGTGCAGGCGAGGTTCCGCAGGAGGCGACTGACCAACGCAGGCGGGAAGATGAAGGTGT

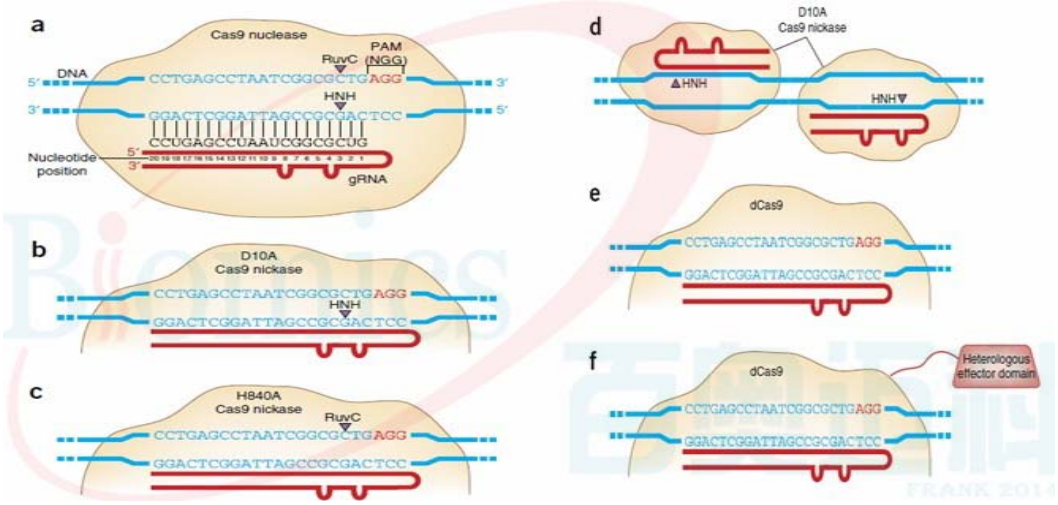
GGTGGGCTGTGTTGGCGGCGGCGATCCTGGCAGGATGCCGGGCCCAGACGGAGCAGGAGGTGGAGGTG

CCGGAGCAGGCCAGGTGGAAGGCCGGCCAGCCCTGGGAGCTGGCCCTGGGCCGCTTCTGGGATTACCTG

CGCTGGGTACAGTCGCTGTCTGATCAGGTGCAGGAGGAGCTGCTCAGCTCGCAGGTCACCCAGGAACTGA

CGATGCTGATGGAAGAAACCATGAAGGAGGTGAAGGCCTACAAGTCGGAGCTGGAGGAGCAGCTGAGCC

2.根据客户试验需要选择合适的 Cas9/sgRNA 设计模式：Single sgRNA/ Pair sgRNA

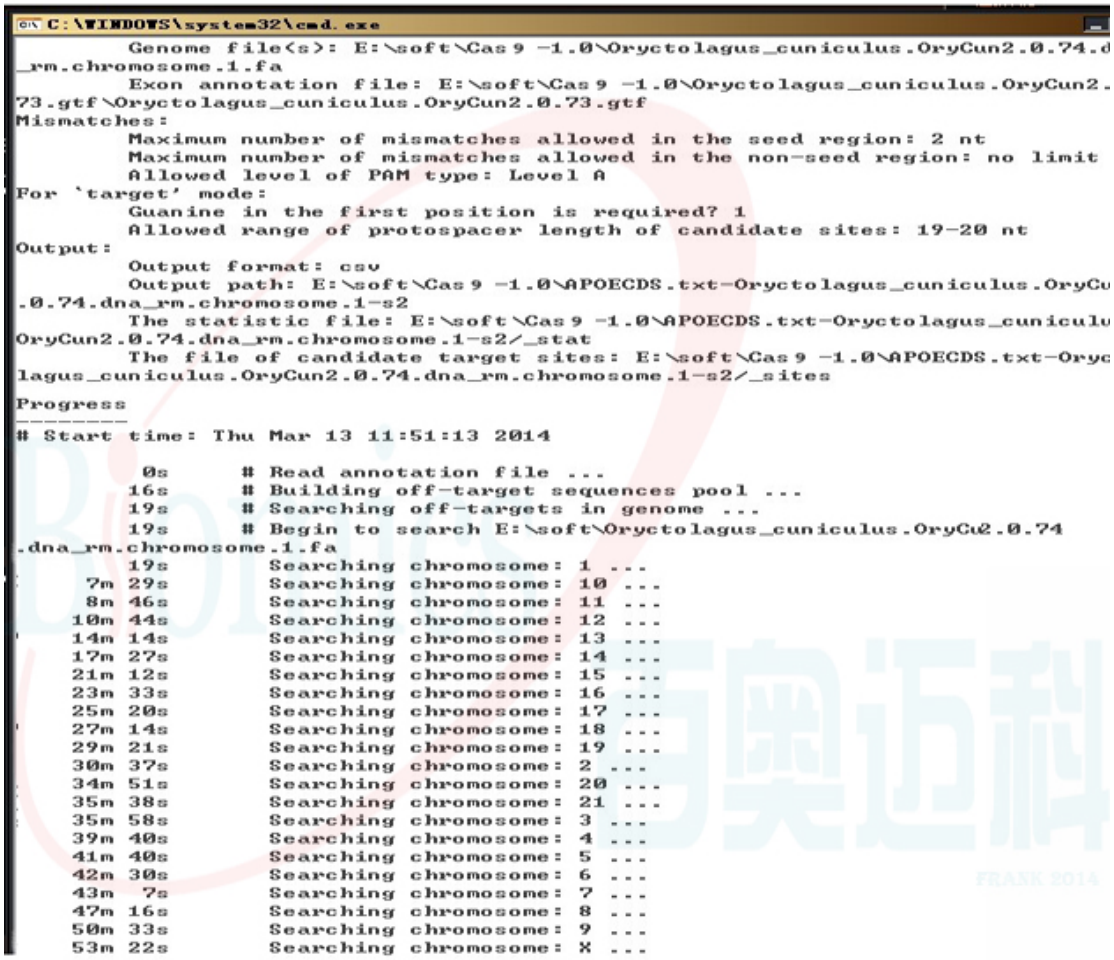


3.选择基因组物种信息：例如 Oryctolagus_cuniculus（兔）

Show All entries | Show/hide columns | Filter

★	Species	DNA (FASTA)	cDNA (FASTA)	CDS (FASTA)	ncRNA (FASTA)	Protein sequence (FASTA)	Annotated sequence (EMBL)	Annotated sequence (GenBank)	Gene sets	Whole databases	Variation (GVF)	Variation (VCF)	Variatio (VEP)
Y	**Human** *Homo sapiens*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	GVF	VCF	VEP
Y	**Mouse** *Mus musculus*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	GVF	VCF	VEP
Y	**Zebrafish** *Danio rerio*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	GVF	VCF	VEP
	Platypus *Ornithorhynchus anatinus*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	GVF	VCF	VEP
	Rabbit *Oryctolagus cuniculus*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	-	-	VEP
	Rat *Rattus norvegicus*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	GVF	VCF	VEP
	Saccharomyces cerevisiae *Saccharomyces cerevisiae*	FASTA	FASTA	FASTA	FASTA	FASTA	EMBL	GenBank	GTF	MySQL	GVF	VCF	VEP

4.Biomics CRISPR program 设计 sgRNA，全基因组分析脱靶位点、种子序列错配分析



5.分析报告输出

A、潜在靶点输出

Guide #21	GAAGGCCTACAAGTCGGAGCTGG	191F
Guide #40	CATGCTGGGCCAGAGCACCGAGG	
Guide #18	GCGCAGCTTGCGCAGGTGCGAGG	366R
Guide #17	GAACGCGCGCGCGAGCTCCTCGG	342R
Guide #16	GCGCGAGCTCCTCGGTGCTCTGG	334R
Guide #23	GCAGCTGAGCCCCATGGCGCAGG	218F
Guide #6	CTCCTCCAGCTCCGACTTGTAGG	
Guide #39	GGCGAGGCGCAGGCCATGCTGGG	333F
Guide #9	CCACCTGCAGCTCCTTGGACAGG	
Guide #10	GCCTGCCACCTGCAGCTCCTTGG	237R
Guide #11	GTTGCAGACGTCCTCCATGTCGG	270R
Guide #12	CGCCGCGGTACTGCGCCAGGCGG	292R
Guide #13	CCTCGCCGCGGTACTGCGCCAGG	
Guide #14	GCATGGCCTGCGCCTCGCCGCGG	307R

B、具体靶点分析，例如>73f GCTTCTGGGATTACCTGCGCTGG 脱靶分析如下表

1	基因组位置	非种子区/种子区错配碱基	靶点位置及方向	脱靶位点错配分析--非种子区/种子区		错配碱基数	脱靶外显子可能区域
2	# Location	Site	Target	Mm.Type	PAM	Mm.All	Exon_info
3	4:84291426-84291449:+	GgTaCTGG_GATTACCcGCGC-TGGG	73f	A12	A:NGG	3	
4	12:93987293-93987316:-	cCacCTGG_GATTACCTGgGC-AGGG	73f	A13	A:NGG	4	
5	9:100688371-100688394:-	GtaaCTGG_GATTtCCTGCGC-CGGG	73f	A13	A:NGG	4	MC4R (ENSOCUG00000025457)
6	1:91601619-91601642:-	cCTTggcG_cATTACCTGCGC-AGGA	73f	A14	A:NGG	5	
7	1:101612927-101612950:+	tCTTCcca_GATTcCCTGCGC-CGGG	73f	A14	A:NGG	5	
8	12:146225339-146225362:-	tCcaCTGt_GATTACCTGCGg-AGGC	73f	A14	A:NGG	5	
9	15:18439216-18439239:-	tgTgCTaG_GtTTACCTGCGC-TGGT	73f	A14	A:NGG	5	
10	19:19601763-19601786:-	aCacaTGG_GAgTACCTGCGC-TGGA	73f	A14	A:NGG	5	
11	9:64770940-64770963:+	GCgcCgaG_GcTTACCTGCGC-CGGG	73f	A14	A:NGG	5	NPC1 (ENSOCUG00000009850)
12	X:82176756-82176779:-	aCTTtgGa_GATTACCTaCGC-TGGT	73f	A14	A:NGG	5	
13	X:92855823-92855846:-	tCacCTGc_GATTACCTGgGC-AGGG	73f	A14	A:NGG	5	
14	GL018710:2965967-2965990:+	ctgcCTGG_GATTACCTGtGC-AGGG	73f	A14	A:NGG	5	
15	GL018755:1500551-1500574:-	agaTgTGG_GATcACCTGCGC-AGGA	73f	A14	A:NGG	5	
16	11:59269468-59269491:-	GCTgtctt_GATTACCTGCcC-AGGC	73f	A15	A:NGG	6	
17	12:84204671-84204694:-	aCTgCaca_GATgACCTGCGC-AGGG	73f	A15	A:NGG	6	
18	13:142175197-142175220:-	atTTaccG_GATcACCTGCGC-GGGA	73f	A15	A:NGG	6	
19	14:89828397-89828420:+	agTgaaGG_GATTtCCTGCGC-TGGG	73f	A15	A:NGG	6	
20	15:72106358-72106381:+	cagcCaGG_GATTACCTGCaC-AGGA	73f	A15	A:NGG	6	FRAS1 (ENSOCUG00000017867)
21	17:83394534-83394557:-	tggaCgGG_GATcACCTGCGC-AGGT	73f	A15	A:NGG	6	
22	18:44317992-44318015:+	GCTaagat_GATTACCTGCtC-AGGT	73f	A15	A:NGG	6	ALDH18A1 (ENSOCUG00000005956)
23	19:21128287-21128310:+	GgaggTcG_GgTTACCTGCGC-TGGC	73f	A15	A:NGG	6	
24	19:50668606-50668629:-	cggcaTGG_GATTACCTGCcC-TGGG	73f	A15	A:NGG	6	
25	2:62131667-62131690:-	GaagCaGa_GATTACCaGCGC-GGGT	73f	A15	A:NGG	6	

6. 筛选最优 sgRNA 靶点，设计并合成 sgRNA DNA oligo 用于构建 sgRNA 表达载体或体外转录合成相应 sgRNA

筛选候选靶点一般原则是 sgRNA 尽量与脱靶位点匹配低，与潜在脱靶位点不匹配数越多表明脱靶可能越低，尤其是种子区与脱靶位点不匹配则脱靶率更低。另外尽可能减少脱靶效应发生在其他蛋白编码区。

百奥迈科生物技术有限公司
地址：江苏省南通市经济与技术开发区常兴路 76 号 邮编：226016
网址：www.biomics.cn；www.premirna.com
电话：0513-85175207
传真：0513-85175202
电邮：sales@biomics.cn（销售），jane@biomics.cn（技术服务）