Supplemental Material for
CRISPR-ERA: a comprehensive design tool for CRISPR-mediated gene editing, repression and activation

Honglei Liu, Zheng Wei, Antonia Dominguez, Yanda Li, Xiaowo Wang*, Lei S. Qi*

Contents

Supplemental Methods

1 Construction of the sgRNA database

To construct the whole-genome sgRNA database, we first downloaded the genome sequence files in FASTA format and genome annotation files in RefFlat or GFF format, from UCSC genome browser or NCBI (Table S1). We searched for all possible sgRNA sequences with a PAM (NGG) sequence and 20-basepair (bp) DNA-binding sequences with patterns of $N_{20}NGG$. During the search, locations and strand information of each sgRNA was recorded. We used Bowtie to find all possible off-target sequences (both PAM=NGG, PAM=NAG were considered) that contained up to 3-bp mismatches for each sgRNA. For genome editing with Cas9 nickase, we searched for a pair of sgRNAs that contained up to 3-bp mismatches for each sgRNA, with an offset distance within -50 – 100 bp.

We extracted the location information of each gene from genome annotation files. For gene editing, only coding region (CDS) was considered, and UTR region was excluded. For repression, 3 kb region centered at TSS or the sense strand of the 5' end of the gene in bacteria was recorded. For activation, up to 1.5 kb upstream of TSS was recorded. (Fig. S1).

For each sgRNA, we computed the efficacy score (E) and specificity penalty score (S) by analyzing the sgRNA sequence feature (e.g., GC % content), location relative to target gene or isoform, and off-target sites.

2 The scoring method

We calculated E-score and S-score differently for different applications:

(1) Editing using nuclease: For non-bacteria organisms, E-score is based on GC content, poly-T sequence and exon location; for bacteria, E-score depends on GC content, poly-T sequence and distance to TSS. For both, S-score is evaluated based on the off-target number. Eight types of off-target "patterns" were computed: 0 mismatch, 1 mismatch, 2 mismatches, 3 mismatches with PAM=NGG or PAM=NAG. The penalty score for NAG off-target is smaller than NGG off-target.

(2) Editing using nickase: E-score equals the smaller E-score of the sgRNA pair. E-score will have an extra increase when the offset distance is within (0, 20 bp). S-score is based on the

off-target number *OFF*(*a, b*) of the sgRNA pair, where *a* (*a* = 0,1,2,3), *b* (*b* = 0,1,2,3) are the mismatch number of the off-target for each sgRNA in the pair, respectively.

(3) Activation and repression: E-score is based on GC content, poly-T sequence, and distance to TSS. S-score is evaluated based on the mismatch patterns. Eight types of off-target "patterns" were computed: 0 mismatch, 1 mismatch, 2 mismatches, 3 mismatches with PAM=NGG or PAM=NAG. The penalty score for NAG off-target is smaller than NGG off-target.

The detailed scoring matrix could be accessed at: http://crispr-era.stanford.edu/help.jsp. As CRISPR/Cas9 study is still fast evolving, we will keep on updating the scoring system once new determinants of CRISPR/Cas9 efficiency and specificity have been uncovered.

3 CRISPR-ERA workflow

We implement a web server (http://CRISPR-ERA.stanford.edu) to host the CRISPR-ERA application. There are three steps for the users to select or input before CRISPR-ERA runs (Figure S2): (1) the type of gene manipulation: gene editing using nuclease, gene editing using nickase, gene repression, or gene activation); (2) the host organism: *Escherichia coli*, *Bacillus subtilis, Saccharomyces cerevisiae, Drosophila melanogaster, Caenorhabditis elegans, Danio rerio, Rattus norvegicus, Mus musculus,* and *Homo sapiens;* (3) the input format: gene name, gene sequence, or TSS location for gene regulation; or gene name, gene sequence, or location region for gene editing. CRISPR-ERA provides two options for inputting user-defined target sequence, using a textbox (genomic region) or uploading files (DNA sequence in the FASTA format).

After user input, CRISPR-ERA uses an index-searching method to find sgRNAs in the predefined database. The sgRNAs are ranked according to the total of E score and S score. We provide the details of each sgRNA score using an interactive table. CRISPR-ERA also provides custom tracks in the UCSC genome browser to visualize the sgRNA design on the genome.

Supplemental Figures
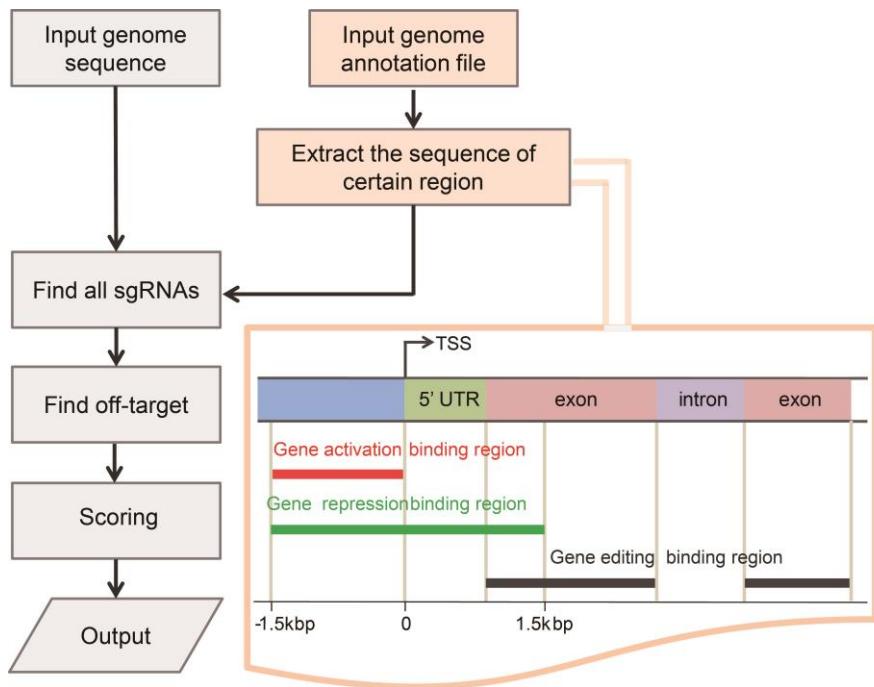
Fig. S1. The sgRNA database. To construct the genome-wide sgRNA library, we first search for all potential sgRNA sequences with default PAM (NGG) sequence and 20 bp DNA-binding length using the FASTA format genome sequence (Supplemental Table 1). We use Bowtie to find all possible off-target sequences (both PAM=NGG, PAM=NAG are considered). For the sgRNA pairs used in genome editing with nickase, CRISPR-ERA searches for genome-wide sgRNA pairs that have maximal 3 mismatches for each sgRNA and the offset distance within the range (-50 – 100 bp). The efficacy score (E) and specificity penalty score (S) are computed. For the genome-wide sgRNA database, the annotation file (Table S1) is the input to extract the sgRNA target sequences for different types of genome manipulation. All potential sgRNA sequences are designed within these regions.
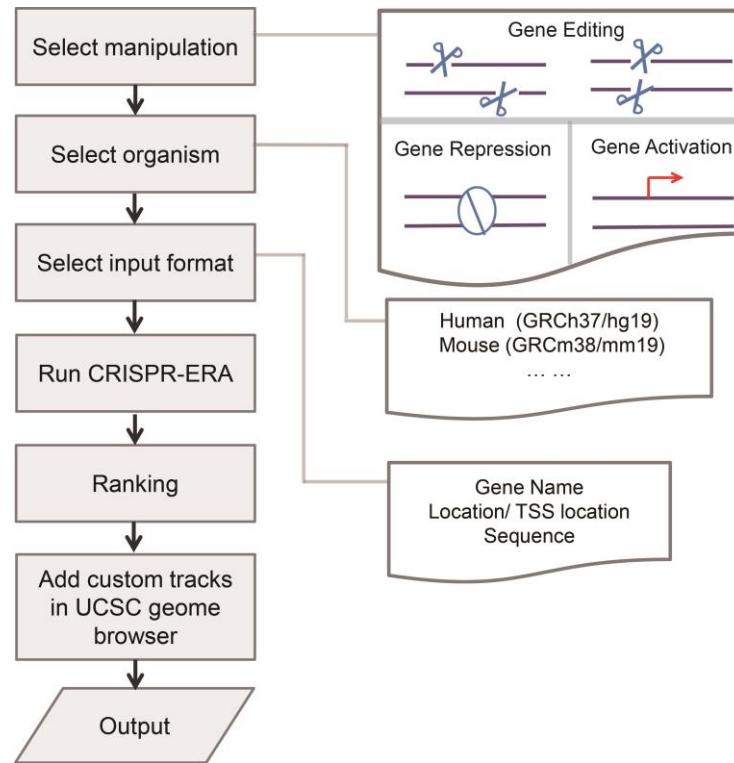
Fig. S2. CRISPR-ERA workflow. The users need to specific three options: (1) The type of gene manipulation (gene editing using nickase, gene editing using nuclease, gene repression, or activation); (2) host organism; and (3) sequence format (gene name, gene sequence or chromosome location). To input the sequence, we provide two options based on the webpage textbox or uploading files. CRISPR-ERA uses a index-searching method to define usable sgRNAs. Index is constructed according to the location or gene name. CRISPR-ERA also visualizes the sgRNAs using the custom tracks in the UCSC genome browser.

A. Top 50 sgRNAs design of mouse gene *Sox2* for genome editing using nickase.
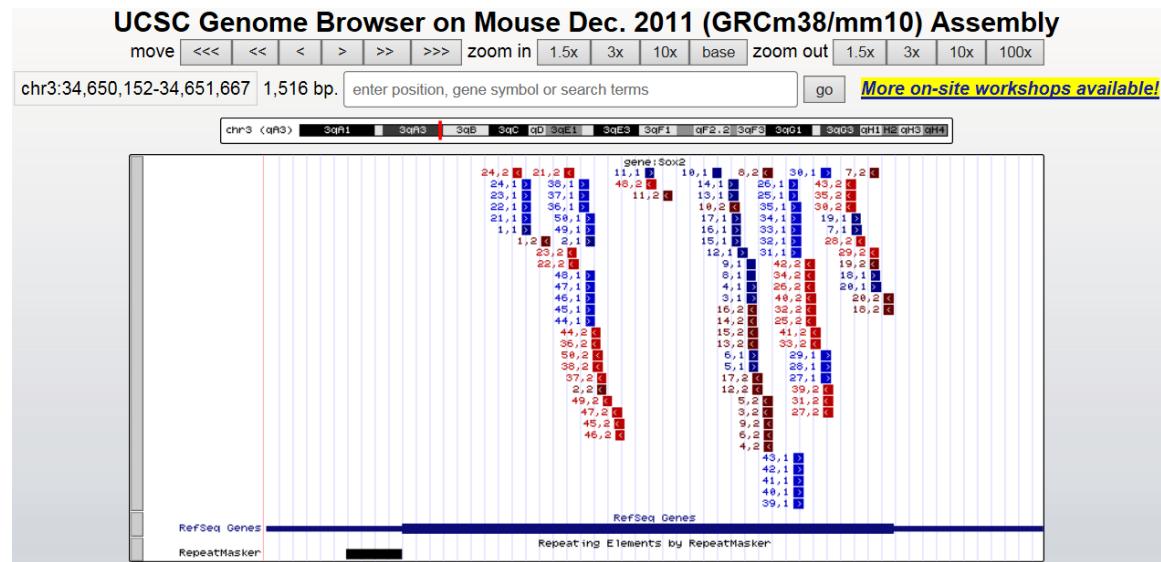
B. Top 50 sgRNAs design of mouse gene *Sox2* for genome editing using nuclease.
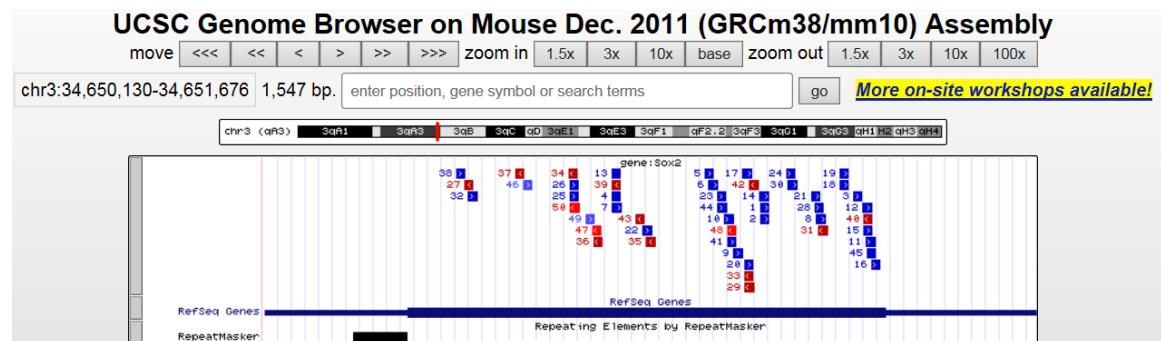
C. Top 50 sgRNAs design of mouse gene *Sox2* for gene repression.
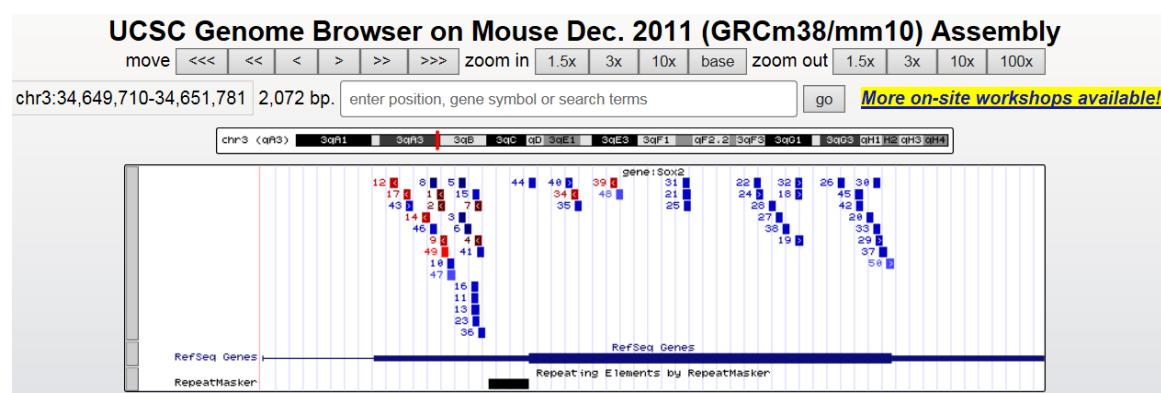
D. Top 50 sgRNAs design of mouse gene *Sox2* for gene activation.
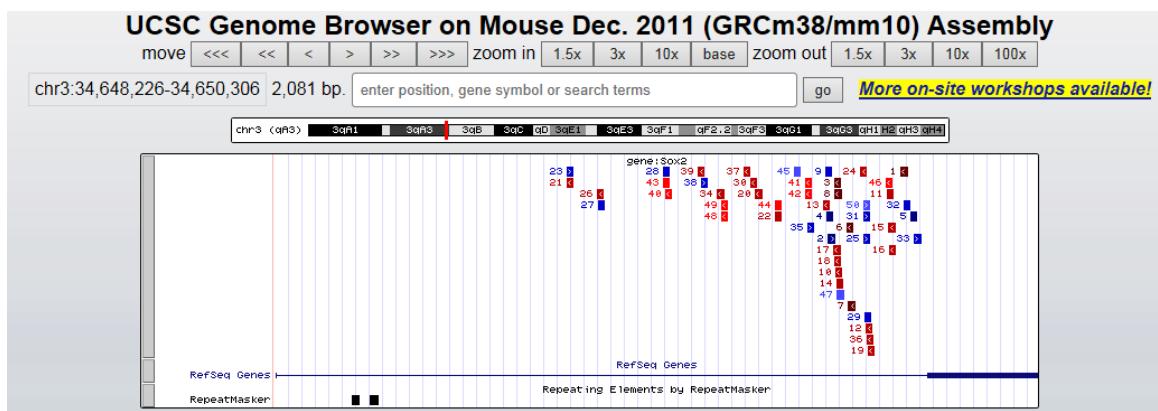
Fig. S3. Visualization of sgRNA designs in UCSC genome browser using murine *Sox2* as an example.

Supplemental Tables

Table S1. Database and software used in CRISPR-ERA.

A. Database

	Genome version	Genome database	Annotation file database
Human	GRCh37/hg19	UCSC genome browser	RefFlat format
Mouse	GRCm38/mm10	UCSC genome browser	RefFlat format
Rat	RGSC 5.0/rn5	UCSC genome browser	RefFlat format
Zebrafish	Zv9/danRer7	UCSC genome browser	RefFlat format
D.melanogaster	BDGP R5/dm3	UCSC genome browser	RefFlat format
C.elegans	WS220/ce10	UCSC genome browser	RefFlat format
S.cerevisiae	SacCer_Apr2011/sacCer3	NCBI	GFF format
E. coli	NC_000913	NCBI	GFF format
B. subtilis	NC_000964	NCBI	GFF format

B. Software

	Ref.
UCSC genome browser	(Kent, *et al.*, 2002)
Blat	(Kent, 2002)
Bowtie	(Langmead *et al.*, 2009)

Table S2. Detailed information of the top 50 sgRNA designs for editing (using nickase (A) or nuclease (B)) or regulating (repression (C) or activation (D)) murine *Sox2*.

A. Top 50 sgRNAs design of mouse gene *Sox2* for genome editing using nickase.

ID	sgRNA sequence1	Location	Strand	sgRNA sequence2	Location	Strand	Offset distance	Target Gene	Transcript ID	Target Exon	E score	S score	E+S score
1	AAAGTTTCCACTCCGCGCCC	chr3:34650648	+	CGTTCATCGACGAGGCCAAG	chr3:34650686	–	17	Sox2	NM_011443	1	20	0	20
2	ATCCTTCTTCATGAGCGTCT	chr3:34650773	+	CACGCTTCCCGGAGGCTTGC	chr3:34650795	–	1	Sox2	NM_011443	1	20	0	20
3	CGAGTAGGACATGCTGTAGG	chr3:34651088	+	CGGTATGGCGCTGGGCTCCA	chr3:34651119	–	10	Sox2	NM_011443	1	20	0	20
4	CGAGTAGGACATGCTGTAGG	chr3:34651088	+	GGTATGGCGCTGGGCTCCAT	chr3:34651120	–	11	Sox2	NM_011443	1	20	0	20
5	CTGCGAGTAGGACATGCTGT	chr3:34651091	+	CGGTATGGCGCTGGGCTCCA	chr3:34651119	–	7	Sox2	NM_011443	1	20	0	20
6	CTGCGAGTAGGACATGCTGT	chr3:34651091	+	GGTATGGCGCTGGGCTCCAT	chr3:34651120	–	8	Sox2	NM_011443	1	20	0	20
7	CTGGGCCATGTGCAGTCTAC	chr3:34651292	+	GTGCCCGGCACGGCCATTAA	chr3:34651327	–	14	Sox2	NM_011443	1	20	0	20
8	GAGTAGGACATGCTGTAGGT	chr3:34651087	+	CGGTATGGCGCTGGGCTCCA	chr3:34651119	–	11	Sox2	NM_011443	1	20	0	20
9	GAGTAGGACATGCTGTAGGT	chr3:34651087	+	GGTATGGCGCTGGGCTCCAT	chr3:34651120	–	12	Sox2	NM_011443	1	20	0	20
10	GCAGGGCGCTGACGTCGTAG	chr3:34651020	+	AGCTCGCAGACCTACATGAA	chr3:34651054	–	13	Sox2	NM_011443	1	20	0	20
11	GCGCGTAGCTGTCCATGCGC	chr3:34650888	+	CGGCAGCTACAGCATGATGC	chr3:34650924	–	15	Sox2	NM_011443	1	20	0	20
12	GGTGGGCGAGCCGTTCATGT	chr3:34651070	+	TACTCGCAGCAGGGCACCCC	chr3:34651099	–	8	Sox2	NM_011443	1	20	0	20
13	GTAGGTCTGCGAGCTGGTCA	chr3:34651052	+	AGCATGTCCTACTCGCAGCA	chr3:34651090	–	17	Sox2	NM_011443	1	20	0	20
14	GTAGGTCTGCGAGCTGGTCA	chr3:34651052	+	CAGCATGTCCTACTCGCAGC	chr3:34651089	–	16	Sox2	NM_011443	1	20	0	20
15	GTTCATGTAGGTCTGCGAGC	chr3:34651058	+	AGCATGTCCTACTCGCAGCA	chr3:34651090	–	11	Sox2	NM_011443	1	20	0	20
16	GTTCATGTAGGTCTGCGAGC	chr3:34651058	+	CAGCATGTCCTACTCGCAGC	chr3:34651089	–	10	Sox2	NM_011443	1	20	0	20
17	GTTCATGTAGGTCTGCGAGC	chr3:34651058	+	TACTCGCAGCAGGGCACCCC	chr3:34651099	–	20	Sox2	NM_011443	1	20	0	20
18	TAATGGCCGTGCCGGGCACC	chr3:34651329	+	TGCCCCTGTCGCACATGTGA	chr3:34651355	–	5	Sox2	NM_011443	1	20	0	20

19	TGGGCCATGTGCAGTCTACT	chr3:34651291	+	GTGCCCGGCACGGCCATTAA	chr3:34651327	−	15	Sox2	NM_011443	1	20	0	20
20	TTAATGGCCGTGCCGGGCAC	chr3:34651330	+	TGCCCCTGTCGCACATGTGA	chr3:34651355	−	4	Sox2	NM_011443	1	20	0	20
21	AAAGTTTCCACTCCGCGCCC	chr3:34650648	+	AGCACCCGGATTATAAATAC	chr3:34650731	−	62	Sox2	NM_011443	1	15	0	15
22	AAAGTTTCCACTCCGCGCCC	chr3:34650648	+	ATTATAAATACCGGCCGCGG	chr3:34650740	−	71	Sox2	NM_011443	1	15	0	15
23	AAAGTTTCCACTCCGCGCCC	chr3:34650648	+	CGGATTATAAATACCGGCCG	chr3:34650737	−	68	Sox2	NM_011443	1	15	0	15
24	AAAGTTTCCACTCCGCGCCC	chr3:34650648	+	GATCAGCAAGCGCCTGGGCG	chr3:34650630	−	−39	Sox2	NM_011443	1	15	0	15
25	AACCACGGGGGGGCTGGAGC	chr3:34651169	+	CAGGGCGCCCTGCCAGGCCG	chr3:34651203	−	13	Sox2	NM_011443	1	15	0	15
26	AACCACGGGGGGGCTGGAGC	chr3:34651169	+	TCCAGGGCGCCCTGCCAGGC	chr3:34651201	−	11	Sox2	NM_011443	1	15	0	15
27	ACATGCTGATCATGTCCCGG	chr3:34651233	+	ATGATCAGCATGTACCTCCC	chr3:34651237	−	−17	Sox2	NM_011443	1	15	0	15
28	ACATGCTGATCATGTCCCGG	chr3:34651233	+	ATGGCCCAGCACTACCAGAG	chr3:34651300	−	46	Sox2	NM_011443	1	15	0	15
29	ACATGCTGATCATGTCCCGG	chr3:34651233	+	GTGCCCGGCACGGCCATTAA	chr3:34651327	−	73	Sox2	NM_011443	1	15	0	15
30	ACATGCTGATCATGTCCCGG	chr3:34651233	+	TGCGCCCAGTAGACTGCACA	chr3:34651281	−	27	Sox2	NM_011443	1	15	0	15
31	AGAGGTAACCACGGGGGGGC	chr3:34651175	+	ATGATCAGCATGTACCTCCC	chr3:34651237	−	41	Sox2	NM_011443	1	15	0	15
32	AGAGGTAACCACGGGGGGGC	chr3:34651175	+	CAGGGCGCCCTGCCAGGCCG	chr3:34651203	−	7	Sox2	NM_011443	1	15	0	15
33	AGAGGTAACCACGGGGGGGC	chr3:34651175	+	CTGCCAGGCCGGGGACCTCC	chr3:34651212	−	16	Sox2	NM_011443	1	15	0	15
34	AGAGGTAACCACGGGGGGGC	chr3:34651175	+	TCCAGGGCGCCCTGCCAGGC	chr3:34651201	−	5	Sox2	NM_011443	1	15	0	15
35	AGAGGTAACCACGGGGGGGC	chr3:34651175	+	TGCGCCCAGTAGACTGCACA	chr3:34651281	−	85	Sox2	NM_011443	1	15	0	15
36	AGCGTCTTGGTTTTCCGCCG	chr3:34650760	+	AAGGATAAGTACACGCTTCC	chr3:34650784	−	3	Sox2	NM_011443	1	15	0	15
37	AGCGTCTTGGTTTTCCGCCG	chr3:34650760	+	CACGCTTCCCGGAGGCTTGC	chr3:34650795	−	14	Sox2	NM_011443	1	15	0	15
38	AGCGTCTTGGTTTTCCGCCG	chr3:34650760	+	GATAAGTACACGCTTCCCGG	chr3:34650787	−	6	Sox2	NM_011443	1	15	0	15
39	AGGAAGAGGTAACCACGGGG	chr3:34651179	+	ATGATCAGCATGTACCTCCC	chr3:34651237	−	37	Sox2	NM_011443	1	15	0	15
40	AGGAAGAGGTAACCACGGGG	chr3:34651179	+	CAGGGCGCCCTGCCAGGCCG	chr3:34651203	−	3	Sox2	NM_011443	1	15	0	15

41	AGGAAGAGGT AACCACGGGG	chr3:34 651179	+	CTGCCAGGCC GGGGACCTCC	chr3:34 651212	–	12	Sox2	NM_011443	1	15	0	15
42	AGGAAGAGGT AACCACGGGG	chr3:34 651179	+	TCCAGGGCGC CCTGCCAGGC	chr3:34 651201	–	1	Sox2	NM_011443	1	15	0	15
43	AGGAAGAGGT AACCACGGGG	chr3:34 651179	+	TGCGCCCAGT AGACTGCACA	chr3:34 651281	–	81	Sox2	NM_011443	1	15	0	15
44	ATCCTTCTTC ATGAGCGTCT	chr3:34 650773	+	AAGGATAAGT ACACGCTTCC	chr3:34 650784	–	-10	Sox2	NM_011443	1	15	0	15
45	ATCCTTCTTC ATGAGCGTCT	chr3:34 650773	+	ACAGCATGGC GAGCGGGGTT	chr3:34 650830	–	36	Sox2	NM_011443	1	15	0	15
46	ATCCTTCTTC ATGAGCGTCT	chr3:34 650773	+	CAGCATGGCG AGCGGGGTTG	chr3:34 650831	–	37	Sox2	NM_011443	1	15	0	15
47	ATCCTTCTTC ATGAGCGTCT	chr3:34 650773	+	CGGGAACAGC ATGGCGAGCG	chr3:34 650825	–	31	Sox2	NM_011443	1	15	0	15
48	ATCCTTCTTC ATGAGCGTCT	chr3:34 650773	+	GACAGCTACG CGCACATGAA	chr3:34 650892	–	98	Sox2	NM_011443	1	15	0	15
49	ATCCTTCTTC ATGAGCGTCT	chr3:34 650773	+	GAGGCTTGCT GGCCCCCGGC	chr3:34 650806	–	12	Sox2	NM_011443	1	15	0	15
50	ATCCTTCTTC ATGAGCGTCT	chr3:34 650773	+	GATAAGTACA CGCTTCCCGG	chr3:34 650787	–	-7	Sox2	NM_011443	1	15	0	15

B. Top 50 sgRNAs design of mouse gene *Sox2* for genome editing using nuclease.

ID	sgRNA sequence	Target Gene	Transcript ID	Target Exon	Chromosome	Location	Strand	E score	S score	E+S score
1	ATGGAGCCCAGCGCCATACC	Sox2	NM_011443	1	3	34651120	+	15	0	15
2	CATGGAGCCCAGCGCCATAC	Sox2	NM_011443	1	3	34651121	+	15	0	15
3	CGGGCCGCTCTGGTAGTGCT	Sox2	NM_011443	1	3	34651307	+	15	0	15
4	CTCGCCATGCTGTTCCCGCC	Sox2	NM_011443	1	3	34650823	+	15	0	15
5	CTGACGTCGTAGCGGTGCAT	Sox2	NM_011443	1	3	34651009	+	15	0	15
6	GCAGGGCGCTGACGTCGTAG	Sox2	NM_011443	1	3	34651017	+	15	0	15
7	GCTCGCCATGCTGTTCCCGC	Sox2	NM_011443	1	3	34650824	+	15	0	15
8	GGTACATGCTGATCATGTCC	Sox2	NM_011443	1	3	34651233	+	15	0	15
9	GGTGGGCGAGCCGTTCATGT	Sox2	NM_011443	1	3	34651067	+	15	0	15
10	GTAGGTCTGCGAGCTGGTCA	Sox2	NM_011443	1	3	34651049	+	15	0	15
11	GTGCCGTTAATGGCCGTGCC	Sox2	NM_011443	1	3	34651333	+	15	0	15
12	TAATGGCCGTGCCGGGCACC	Sox2	NM_011443	1	3	34651326	+	15	0	15
13	TCGCCATGCTGTTCCCGCCG	Sox2	NM_011443	1	3	34650822	+	15	0	15
14	TGGAGCCCAGCGCCATACCG	Sox2	NM_011443	1	3	34651119	+	15	0	15
15	TTAATGGCCGTGCCGGGCAC	Sox2	NM_011443	1	3	34651327	+	15	0	15
16	CAGGGCAGTGTGCCGTTAA	Sox2	NM_011443	1	3	34651343	+	15	-1	14
17	CGAGTAGGACATGCTGTAGG	Sox2	NM_011443	1	3	34651085	+	15	-1	14
18	GTGCAGTCTACTGGGCGCAG	Sox2	NM_011443	1	3	34651280	+	15	-1	14
19	TGCAGTCTACTGGGCGCAGC	Sox2	NM_011443	1	3	34651279	+	15	-1	14
20	CTGCGAGTAGGACATGCTGT	Sox2	NM_011443	1	3	34651088	+	15	-2	13
21	GATCATGTCCCGGAGGTCCC	Sox2	NM_011443	1	3	34651223	+	15	-2	13
22	GCGCGTAGCTGTCCATGCGC	Sox2	NM_011443	1	3	34650885	+	15	-2	13
23	GGTCATGGAGTTGTACTGCA	Sox2	NM_011443	1	3	34651034	+	15	-2	13
24	AGAGGTAACCACGGGGGGGC	Sox2	NM_011443	1	3	34651172	+	15	-3	12
25	CGCGGCCGGTATTTATAATC	Sox2	NM_011443	1	3	34650739	+	15	-3	12
26	GCGGCCGGTATTTATAATCC	Sox2	NM_011443	1	3	34650738	+	15	-3	12
27	ACAGCCCGGACCGCGTCAAG	Sox2	NM_011443	1	3	34650527	-	15	-4	11
28	ACATGCTGATCATGTCCCGG	Sox2	NM_011443	1	3	34651230	+	15	-4	11
29	AGCATGTCCTACTCGCAGCA	Sox2	NM_011443	1	3	34651090	-	15	-4	11
30	AGGAAGAGGTAACCACGGGG	Sox2	NM_011443	1	3	34651176	+	15	-4	11
31	ATGATCAGCATGTACCTCCC	Sox2	NM_011443	1	3	34651237	-	15	-4	11
32	ATGGGCCTCTTGACGCGGTC	Sox2	NM_011443	1	3	34650535	+	15	-4	11
33	CAGCATGTCCTACTCGCAGC	Sox2	NM_011443	1	3	34651089	-	15	-4	11
34	CGGATTATAAATACCGGCCG	Sox2	NM_011443	1	3	34650737	-	15	-4	11
35	GACAGCTACGCGCACATGAA	Sox2	NM_011443	1	3	34650892	-	15	-4	11
36	GATAAGTACACGCTTCCCGG	Sox2	NM_011443	1	3	34650787	-	15	-4	11
37	GATCAGCAAGCGCCTGGGCG	Sox2	NM_011443	1	3	34650630	-	15	-4	11
38	GCTGTTCTTCTGGTTGCCGC	Sox2	NM_011443	1	3	34650512	+	15	-4	11

39	GGCGGGAACAGCATGGCGAGC	Sox2	NM_011443	1	3	34650823	−	15	−4	11
40	GTGCCCGGCACGGCCATTAA	Sox2	NM_011443	1	3	34651327	−	15	−4	11
41	GTTCATGTAGGTCTGCGAGC	Sox2	NM_011443	1	3	34651055	+	15	−4	11
42	TACTCGCAGCAGGGCACCCC	Sox2	NM_011443	1	3	34651099	−	15	−4	11
43	TGCGGGCGTGAACCAGCGCA	Sox2	NM_011443	1	3	34650870	−	15	−4	11
44	TGGTCATGGAGTTGTACTGC	Sox2	NM_011443	1	3	34651035	+	15	−4	11
45	TGTGCCGTTAATGGCCGTGC	Sox2	NM_011443	1	3	34651334	+	15	−4	11
46	AAAGTTTCCACTCCGCGCCC	Sox2	NM_011443	1	3	34650645	+	15	−5	10
47	AAGGATAAGTACACGCTTCC	Sox2	NM_011443	1	3	34650784	−	15	−5	10
48	AGCTCGCAGACCTACATGAA	Sox2	NM_011443	1	3	34651054	−	15	−5	10
49	ATCCTTCTTCATGAGCGTCT	Sox2	NM_011443	1	3	34650770	+	15	−5	10
50	ATTATAAATACCGGCCGCGG	Sox2	NM_011443	1	3	34650740	−	15	−5	10

C. Top 50 sgRNAs design of mouse gene *Sox2* for gene repression.

ID	sgRNA sequence	Target Gene	Transcript ID	Distance to TSS	Chromosome	Location	Strand	E score	S score	E+S score
1	AAAACTAATCACAACAATCG	Sox2	NM_011443	165	3	34650169	–	20	-4	16
2	ACTAATCACAACAATCGCGG	Sox2	NM_011443	168	3	34650172	–	20	-4	16
3	CGAGACGGGCGAAGTGCAAT	Sox2	NM_011443	224	3	34650228	+	20	-4	16
4	CTCCGCCAGATCTCCGCGCA	Sox2	NM_011443	267	3	34650271	–	20	-4	16
5	GAGACGGGCGAAGTGCAATT	Sox2	NM_011443	223	3	34650227	+	20	-4	16
6	GGGGGAAGCGGAGCTCGAGA	Sox2	NM_011443	239	3	34650243	+	20	-4	16
7	TCTCCGCCAGATCTCCGCGC	Sox2	NM_011443	266	3	34650270	–	20	-4	16
8	TTTGGAAAGGCTTAAGCCTC	Sox2	NM_011443	147	3	34650151	+	20	-4	16
9	ACAACAATCGCGGCGGCCCG	Sox2	NM_011443	175	3	34650179	–	20	-5	15
10	AAACAGGCGCTCTCCTCCTC	Sox2	NM_011443	194	3	34650198	+	20	-6	14
11	ATCTGGCGGAGAATAGTTGG	Sox2	NM_011443	258	3	34650262	+	20	-6	14
12	GAAGAGAGTGTTTGCAAAAA	Sox2	NM_011443	43	3	34650047	–	20	-6	14
13	GATCTGGCGGAGAATAGTTG	Sox2	NM_011443	259	3	34650263	+	20	-6	14
14	GGAGAGAAGTTTGGAGCCCG	Sox2	NM_011443	128	3	34650132	–	20	-6	14
15	AGATCTGGCGGAGAATAGTT	Sox2	NM_011443	260	3	34650264	+	20	-7	13
16	TCTGGCGGAGAATAGTTGGG	Sox2	NM_011443	257	3	34650261	+	20	-7	13
17	TGCCTCTTTAAGACTAGGGC	Sox2	NM_011443	79	3	34650083	–	20	-7	13
18	ATGGAGCCCAGCGCCATACC	Sox2	NM_011443	1116	3	34651120	+	12	0	12
19	CATGGAGCCCAGCGCCATAC	Sox2	NM_011443	1117	3	34651121	+	12	0	12
20	CGGGCCGCTCTGGTAGTGCT	Sox2	NM_011443	1303	3	34651307	+	12	0	12
21	CTCGCCATGCTGTTCCCGCC	Sox2	NM_011443	819	3	34650823	+	12	0	12
22	CTGACGTCGTAGCGGTGCAT	Sox2	NM_011443	1005	3	34651009	+	12	0	12
23	GAGATCTGGCGGAGAATAGT	Sox2	NM_011443	261	3	34650265	+	20	-8	12
24	GCAGGGCGCTGACGTCGTAG	Sox2	NM_011443	1013	3	34651017	+	12	0	12
25	GCTCGCCATGCTGTTCCCGC	Sox2	NM_011443	820	3	34650824	+	12	0	12
26	GGTACATGCTGATCATGTCC	Sox2	NM_011443	1229	3	34651233	+	12	0	12
27	GGTGGGCGAGCCGTTCATGT	Sox2	NM_011443	1063	3	34651067	+	12	0	12
28	GTAGGTCTGCGAGCTGGTCA	Sox2	NM_011443	1045	3	34651049	+	12	0	12
29	GTGCCGTTAATGGCCGTGCC	Sox2	NM_011443	1329	3	34651333	+	12	0	12
30	TAATGGCCGTGCCGGGCACC	Sox2	NM_011443	1322	3	34651326	+	12	0	12
31	TCGCCATGCTGTTCCCGCCG	Sox2	NM_011443	818	3	34650822	+	12	0	12
32	TGGAGCCCAGCGCCATACCG	Sox2	NM_011443	1115	3	34651119	+	12	0	12
33	TTAATGGCCGTGCCGGGCAC	Sox2	NM_011443	1323	3	34651327	+	12	0	12
34	ACAGCCCGGACCGCGTCAAG	Sox2	NM_011443	523	3	34650527	–	15	-4	11
35	ATGGGCCTCTTGACGCGGTC	Sox2	NM_011443	531	3	34650535	+	15	-4	11
36	CACGGCCCTGCGCGGAGATC	Sox2	NM_011443	275	3	34650279	+	15	-4	11
37	CAGGGGCAGTGTGCCGTTAA	Sox2	NM_011443	1339	3	34651343	+	12	-1	11
38	CGAGTAGGACATGCTGTAGG	Sox2	NM_011443	1081	3	34651085	+	12	-1	11

39	GATCAGCAAGCGCCTGGGCG	Sox2	NM_011443	626	3	34650630	−	15	-4	11
40	GCTGTTCTTCTGGTTGCCGC	Sox2	NM_011443	508	3	34650512	+	15	-4	11
41	GGCCCTGCGCGGAGATCTGG	Sox2	NM_011443	272	3	34650276	+	15	-4	11
42	GTGCAGTCTACTGGGCGCAG	Sox2	NM_011443	1276	3	34651280	+	12	-1	11
43	TCCCAGCCCTAGTCTTAAAG	Sox2	NM_011443	84	3	34650088	+	20	-9	11
44	TCTCCATCATGTTATACATG	Sox2	NM_011443	410	3	34650414	+	15	-4	11
45	TGCAGTCTACTGGGCGCAGC	Sox2	NM_011443	1275	3	34651279	+	12	-1	11
46	TTTTGGAAAGGCTTAAGCCT	Sox2	NM_011443	148	3	34650152	+	15	-4	11
47	AAAACAGGCGCTCTCCTCCT	Sox2	NM_011443	195	3	34650199	+	20	-10	10
48	AAAGTTTCCACTCCGCGCCC	Sox2	NM_011443	641	3	34650645	+	15	-5	10
49	ACAATCGCGGCGGCCCGAGG	Sox2	NM_011443	178	3	34650182	−	15	-5	10
50	CAGCCCTCACATGTGCGACA	Sox2	NM_011443	1357	3	34651361	+	12	-2	10

D. Top 50 sgRNAs design of mouse gene *Sox2* for gene activation

ID	sgRNA sequence	Target Gene	Transcript ID	Distance to TSS	Chromosome	Location	Strand	E score	S score	E+S score
1	AGGAGCCGGCGCTCGCTGAT	Sox2	NM_011443	-74	3	34649930	−	20	-4	16
2	CACGGCGCACGCCTGTTCGA	Sox2	NM_011443	-272	3	34649732	+	20	-4	16
3	CGTGACCTGTTGCTGAAAAC	Sox2	NM_011443	-256	3	34649748	−	20	-4	16
4	GCACGCCTGTTCGAAGGAAG	Sox2	NM_011443	-278	3	34649726	+	20	-4	16
5	GGGAATAAATGGGTTTCCGG	Sox2	NM_011443	-48	3	34649956	+	20	-4	16
6	GGGATACAAAGGTTCCCCAG	Sox2	NM_011443	-221	3	34649783	−	20	-4	16
7	TACAAAGGTTCCCCAGCGGC	Sox2	NM_011443	-217	3	34649787	−	20	-4	16
8	TGACCTGTTGCTGAAAACGG	Sox2	NM_011443	-254	3	34649750	−	20	-4	16
9	CACGCCTGTTCGAAGGAAGT	Sox2	NM_011443	-279	3	34649725	+	20	-5	15
10	GACCTGTTGCTGAAAACGGG	Sox2	NM_011443	-253	3	34649751	−	20	-5	15
11	TCATGCAAAACCCTCTGGCG	Sox2	NM_011443	-111	3	34649893	−	20	-5	15
12	TTCGGGGCACAGCGCTCTGC	Sox2	NM_011443	-169	3	34649835	−	20	-5	15
13	TTTACCCACTTCCTTCGAAC	Sox2	NM_011443	-286	3	34649718	−	20	-5	15
14	CTGTTGCTGAAAACGGGGGG	Sox2	NM_011443	-250	3	34649754	−	20	-6	14
15	GCAAAACCCTCTGGCGAGGC	Sox2	NM_011443	-107	3	34649897	−	20	-6	14
16	CAAAACCCTCTGGCGAGGCT	Sox2	NM_011443	-106	3	34649898	−	20	-7	13
17	CCGTGACCTGTTGCTGAAAA	Sox2	NM_011443	-257	3	34649747	−	20	-7	13
18	GTGACCTGTTGCTGAAAACG	Sox2	NM_011443	-255	3	34649749	−	20	-7	13
19	GCACAGCGCTCTGCTGGGCT	Sox2	NM_011443	-163	3	34649841	−	20	-8	12
20	AAACAATGGCACACCACCTC	Sox2	NM_011443	-471	3	34649533	−	15	-4	11
21	AATGCCACCTCCGCGCCTCG	Sox2	NM_011443	-985	3	34649019	−	15	-4	11
22	AGTCGCCCTGAACCACCCAT	Sox2	NM_011443	-417	3	34649587	−	15	-4	11
23	CGAGGCGCGGAGGTGGCATT	Sox2	NM_011443	-985	3	34649019	+	15	-4	11
24	CGCCTCCCCCGCGCGGTTCG	Sox2	NM_011443	-185	3	34649819	−	15	-4	11
25	CGCTGTGCCCCGAACCGCGC	Sox2	NM_011443	-175	3	34649829	+	15	-4	11
26	CGTTGATGCTTTCGTCCCCA	Sox2	NM_011443	-903	3	34649101	−	15	-4	11
27	CTTGGGGACGAAAGCATCAA	Sox2	NM_011443	-901	3	34649103	+	15	-4	11
28	GCCGACTCAGGTGTGGCTCA	Sox2	NM_011443	-724	3	34649280	+	15	-4	11
29	GCGCTGTGCCCCGAACCGCG	Sox2	NM_011443	-174	3	34649830	+	15	-4	11
30	GCGGATGAGCGCAGAAACAA	Sox2	NM_011443	-485	3	34649519	−	15	-4	11
31	GCTGTGCCCCGAACCGCGCG	Sox2	NM_011443	-176	3	34649828	+	15	-4	11
32	GGCGGCCAATCAGCGAGCGC	Sox2	NM_011443	-66	3	34649938	+	15	-4	11
33	GGGGGCTGTCAGGGAATAAA	Sox2	NM_011443	-37	3	34649967	+	20	-9	11
34	GGTCTAACTTCTCGTCTGTA	Sox2	NM_011443	-575	3	34649429	−	15	-4	11
35	TCCCCCATGCTACGGAATAT	Sox2	NM_011443	-329	3	34649675	+	15	-4	11
36	TCGGGGCACAGCGCTCTGCT	Sox2	NM_011443	-168	3	34649836	−	20	-9	11
37	TGCGACAGAGCGCAGTGCCG	Sox2	NM_011443	-504	3	34649500	−	15	-4	11
38	TGGCTTCCGTGTCATCGGCT	Sox2	NM_011443	-620	3	34649384	+	15	-4	11
39	TGGGAACCGAGCCGATGACA	Sox2	NM_011443	-629	3	34649375	−	15	-4	11

40	ACCTGAGTCGGCCTAGGAAA	Sox2	NM_011443	-716	3	34649288	-	15	-5	10
41	AGAGCCAATATTCCGTAGCA	Sox2	NM_011443	-336	3	34649668	-	15	-5	10
42	AGCCAATATTCCGTAGCATG	Sox2	NM_011443	-334	3	34649670	-	15	-5	10
43	AGCCACACCTGAGTCGGCCT	Sox2	NM_011443	-722	3	34649282	-	15	-5	10
44	CAGTCGCCCTGAACCACCCA	Sox2	NM_011443	-418	3	34649586	-	15	-5	10
45	CATAAGGGTGGATGGGGCGC	Sox2	NM_011443	-366	3	34649638	+	15	-5	10
46	CCCTTTCATGCAAAACCCTC	Sox2	NM_011443	-116	3	34649888	-	20	-10	10
47	CGCCCCCCGTTTTCAGCAAC	Sox2	NM_011443	-248	3	34649756	+	15	-5	10
48	CGTCTGTACGGTGAGAAGAG	Sox2	NM_011443	-563	3	34649441	-	15	-5	10
49	CTCGTCTGTACGGTGAGAAG	Sox2	NM_011443	-565	3	34649439	-	15	-5	10
50	CTGTGCCCCGAACCGCGCGG	Sox2	NM_011443	-177	3	34649827	+	15	-5	10

Reference

Kent, WJ *et al*. (2002), The human genome browser at UCSC. *Genome Res.* **12**, 996-1006.

Kent, WJ. (2002), BLAT - the BLAST-like alignment tool. *Genome Res.* **12**, 656-64.

Langmead, B *et al*. (2009), Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**,R25.