

Supplementary

1. gRNA orientation

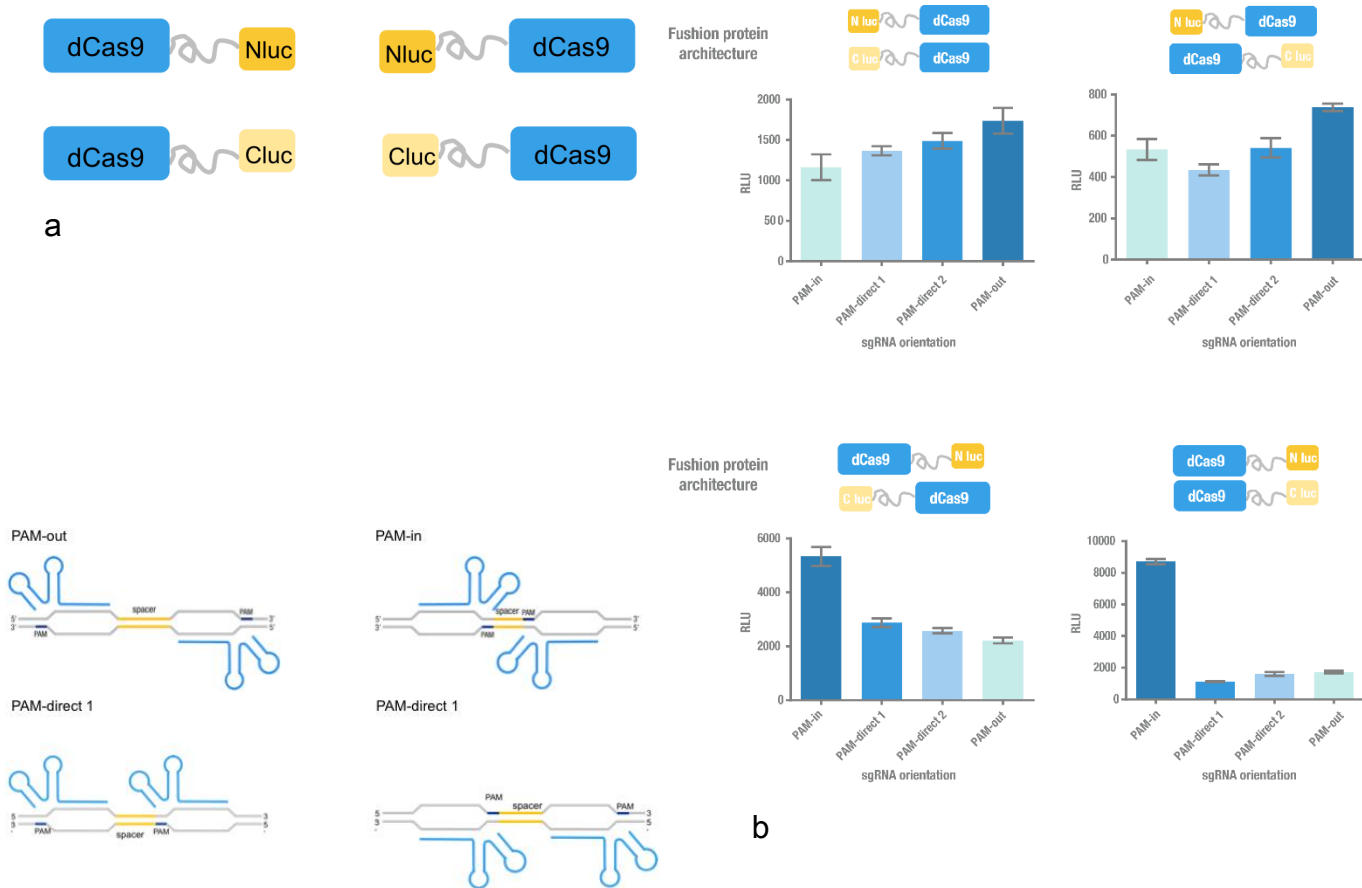


Figure 1. Construction and optimization of PC reporter system. (a) 4 different split luciferase-dCas9 fusion strategies. (b) 4 different sgRNA orientation settings. In orientation PAM-out, the pair of PAM sequences are distal from the spacer sequence, with the 5' end of the sgRNA adjacent to the spacer; in orientation PAM-in, the pair of PAM sequences are adjacent to the spacer sequence, with the 3' end of the sgRNA in proximity to the spacer; in orientation PAM-direct 1 and PAM-direct 2, one PAM sequence is adjacent to and another distal from the spacer. (c) The experimental results of PC reporter

systems using 16 different (4 different constructs X 4 different sgRNA orientation settings) configurations.

As it is shown in the Figure 1c, besides the “PAM-out” orientation, other orientation can also be chosen to search paired specific guide sequences of gRNA in genome. In genome, the paired target sites of Paired dCas9 Reporter System with “PAM-out” orientation, “PAM-in” orientation, “PAM-direct1” orientation, “PAM-direct2” orientation are remarked as “5'-CCNN20-...-N20NGG-3'”, “5'-N20NGG-...-CCNN20-3'”, “5'-CCNN20-...-CCNN20-3'”, “5'-N20NGG-...-N20NGG-3'”, respectively. It's obvious that all of the four kinds of target sites are assembled by only two kinds of single gRNA target site, 'CCNN20' and 'N20NGG'. So it is very convenient to adjust our method for guide sequences design with other orientations. And you can easily pair specific left and right gRNA target sites with the orientation as you like after the second step of SSPD method.

2. Spacer Length Range

In our Paired dCas9 Reporter System, the optimal spacer length is ~20 bp for split luciferase. However, a variety of reporter proteins have been rationally dissected into two pieces, including ubiquitin, dihydrofolate reductase(DHFR), β -Lactamase, β -galactosidase, GFP. In view of the diverse structures of reporter proteins, the optimal spacer length may vary

from one kind to another. And it is also adjustable for other split enzymes by simply changing the spacer length range we choose, which can be easily accomplished by just changing parameters in user interface after the specific test result has been saved.

3. Oligo Generator

We can easily find the reliable target sites on genome using SSPD method. However, designing multiple target sites into oligonucleotides sequences for following sgRNA construction manually can be laborious. Thus we developed a supplementary program to facilitate oligo sequence generation, which is combined with our sgRNA generator.

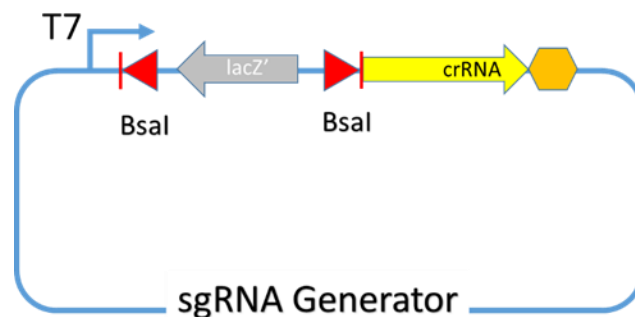


Fig. 2 Schematic of sgRNA Generator plasmid, containing T7 promoter, lacZ' and lacZ promoter flanked by two Bsal Enzyme cutting site sites, crRNA sequence and T7 terminator.

There are two Bsal Enzyme cutting sites with sticky ends of TAGG and AAAC respectively in the sgRNA Generator. In order to undertake Golden Gate Cloning to insert the target 20 bp, we need to append the sticky ends

to the appropriate end of target sites. As it is showed in the Fig. 3, the guide sequence (5'-3') of left gRNA is reverse complementary to the sequence of left target site on the 5'-3' strand in the genome, while is the same as the insert 20bp fragments on coding strand (5'-3') in the left sgRNA plasmid. Thus we append 5'-TAGG-3' to the 5' end of the reverse complement sequence of left target sites on the 5'-3' strand in the genome, and 5'-AAAC-3' to the 5' end of left target site sequence on the 5'-3' strand in the genome. As for right gRNA, we can get the conclusion through similar analysis that we append 5'-AAAC-3' to the 5' end of the reverse complement sequence of left target sites on the 5'-3' strand in the genome, and 5'-TAGG-3' to the 5' end of left target site sequence on the 5'-3' strand in the genome.

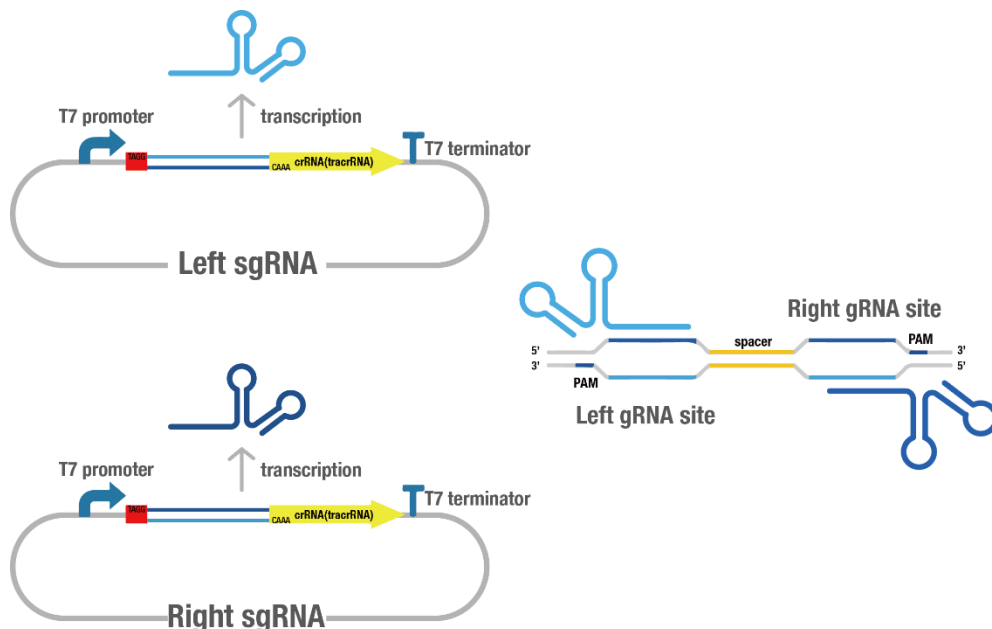


Fig. 3 Schematic of left gRNA and right gRNA transcription and binding to the corresponding target site. The 5'-3' strand of the target sites is colored dark blue,

while the complementary strand is colored light blue. And the two strand of inserted 20 bp target in the sgRNA plasmid is colored by the same color as the strand of the target sites with the same sequence.

4. Human Oral Meta-Genome (HOMG)

The Human Oral Meta-Genome (HOMG) used in our project is downloaded from <http://hmpdacc.org/HMRGD/>, which has 22394 sequences in total.

5. Model: Wilcoxon Rank Sum Test of Block Design

Furthermore, we use our model to analyze each PCR fragment. Here is our result.

	F1**	F2	F3**	F4**	F5*	F6**	F7**	F8**	F9**
W	101	45	115	120	96	120	96	103	120
W _{BD}	2.62	-4.70	4.78	5.55	1.85	5.55	3.71	2.93	5.55
p-value	0.0048		8.76e-07	1.43e-08	0.0312	1.43e-08	1.0e-04	0.0018	1.43e-08

So F2 is not statistical significant, F5 is significant, while the other 7 fragments are all highly statistical significant. In all, it is highly significant.