Supplementary Information

1. gRNA orientation

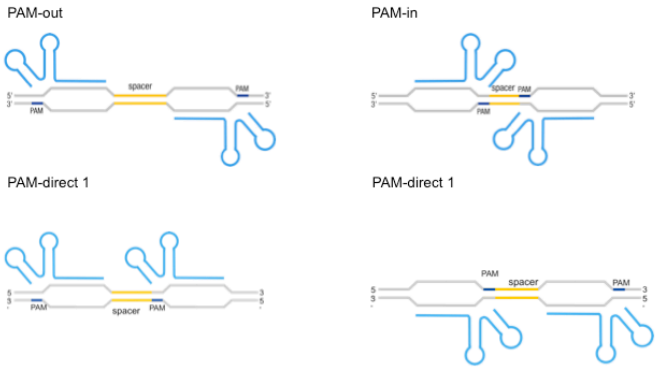

Figure 1. Four different sgRNA orientation settings. In orientation PAM-out, the pair of PAM sequences are distal from the spacer sequence, with the 5' end of the sgRNA adjacent to the spacer; in orientation PAM-in, the pair of PAM sequences are adjacent to the spacer sequence, with the 3' end of the 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.

Besides the “PAM-out” orientation, other orientation can also be used for MTB specific markers identification. 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 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

The programmability of PC reporter system allows for a variety of split protein to be used as reporter, such as split DHFR, split β-Lactamase. In view of the diverse structures of reporter proteins, the optimal spacer length may vary from one to another. Our SSPD algorithm is adjustable for different split reporter by simply changing the spacer length range. It is accomplished by changing parameters in user interface after the specific test result being saved.

3. Oligo Generator

Our SSPD algorithm is efficient to select all MTB specific markers for MTB multi-marker array, however constructing multiple sgRNA can be laborious. Here we designed Oligo Generator to automatically generate an oligo sequence list from selected MTB specific targets. Together with our sgRNA generator, it facilitates sgRNA construction through Golden Gate Cloning.

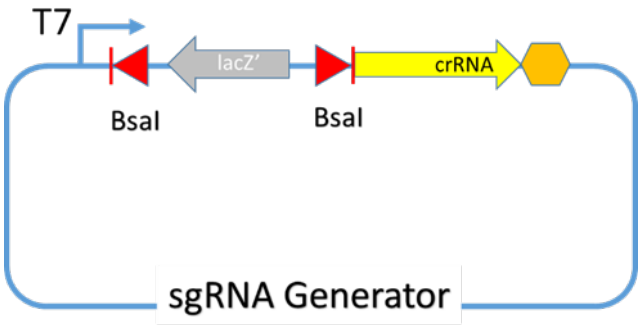

Fig. 2 Schematic of sgRNA Generator plasmid, which contains T7 promotor, lacZ' and lacZ promoter flanked by two BsaI Enzyme cutting site sites, crRNA sequence and T7 terminator.

There are two BsaI Enzyme cutting sites with sticky ends of TAGG and AAAC respectively in the sgRNA Generator. In order to undertake Golden Gate Cloning to replace LacZ with 20 bp guide sequence, we need to append the sticky ends to the appropriate end of target sites. As is shown 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 come to 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.

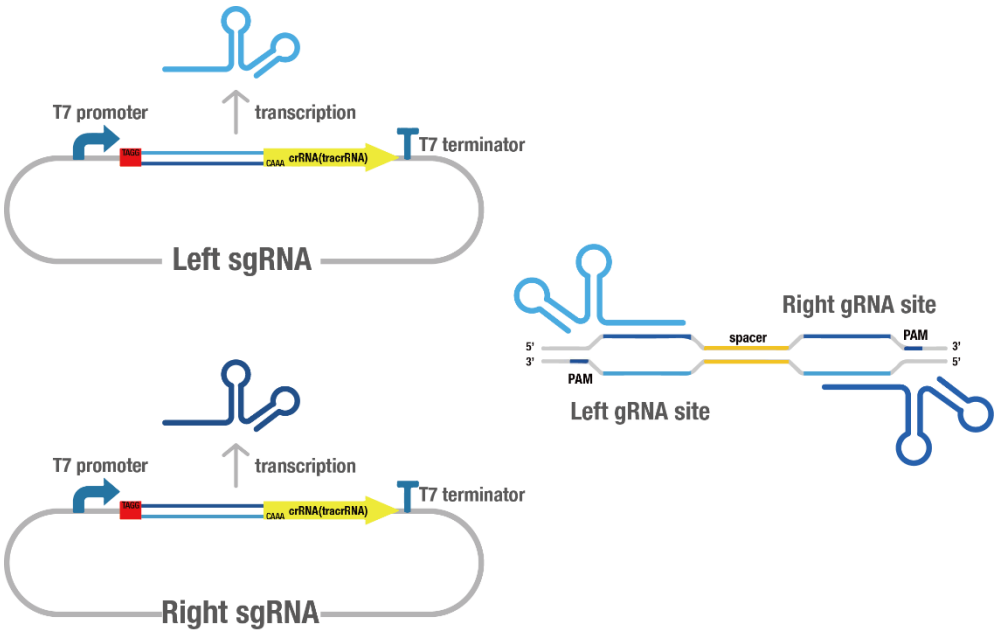

Fig. 3 Schematic of left gRNA and right gRNA transcription and binding to the corresponding target site. The 5'-3' strand of the target sites is colored dark blue, while the complementary strand is colored light blue. And the two strand of inserted 20 bp target in the sgRNA plasmid is colored by the same color as the strand of the target sites with the same sequence.

4. BLAST

However, searching for potential off-target sites in human oral meta-genome (HOMG) requires a huge calculation. Therefore, we adopted a BLAST-based 2-step filter approach. *NCBI blast-2.2.31+* local program was used to run the BLAST, and *BioPython-1.65* was used to find specific sequences. As the BioPython has advantages in processing '*.fasta', '*.gb' nucleotide sequences files and blast out files, we use it as a powerful tool to organize the program.

5. 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.

6. 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.