

sgRNA Design Protocol:

1. Select target region and acquire the corresponding DNA sequence. For better sgRNA selection near the ends of the target region, add 100 bp on either side.
2. Run this region through the following sgRNA design software.
 - a. MIT sgRNA design program at <http://crispr.mit.edu/>. This script only accepts 250 bp regions of DNA. As a result, you may need to break your target sequence into 250 bp segments. To ensure sgRNA design at the junctions, overlapping these regions is recommended.
 - b. BROAD sgRNA design tool at <http://www.broadinstitute.org/rnai/public/analysis-tools/sgrna-design-v1> (Doench et al., 2014). This is the older version of the tool. While the newer version gives more complex output, it lacks the simple sgRNA score that we used in our design process. As a result, we have continued to use the older tool.
 - c. The sgRNAs9 scripts at <http://www.biotoools.com/> (Xie et al., 2014). We use version 2.0.10. We installed this on a Dell N5110 running Windows 7 (64-bit) with an Intel Core i5-2430M processor and 6 GB of RAM. We expect that this program can be run on other personal computers that are able to execute Perl scripts. Before use, you will need to download and format a reference genome for the script to use. Once ready, the program is run from the command line with a command such as the following:

```
sgRNAs9_2.0.10.pl -i "target_region_FASTA.txt" -g target_genome_FASTA.fa -o b -t s -v w
```

Note: this program took several hours to run for each target region.

3. In Microsoft Excel combine the lists of potential sgRNA sites from all three programs and eliminate duplicates in the MIT program output.
 - a. Be sure to combine both the CRISPR.targets_A (antisense strand) and CRISPR.targets_S (sense strand) generated by sgRNAs9 script.
 - b. The BROAD tool output contains the spacer sequence without the PAM and an extended sequence that includes several bases on either side of the spacer. For use with the VLOOKUP function below, we extract the middle of the extended sequence that includes the PAM.
4. Use the online Venn diagram tool Venny (<http://bioinfogp.cnb.csic.es/tools/venny/index.html>) to identify sgRNA target sites identified by all three programs.
5. Copy the list of common sgRNA into a separate tab in Excel.
6. Use the VLOOKUP function to match the BROAD score and MIT score to each sgRNA sequence.
7. The BROAD score is a rough prediction of activity for each sgRNA so sort the sgRNA list by the BROAD score. Eliminate any sgRNA with BROAD score < 0.2. The MIT score is based on off-target binding sites and can be used to select more specific sgRNA.
8. Use the BLAT tool from the UCSC genome browser to quickly identify the genome coordinates for each sgRNA.
 - a. Set the BLAT tool Output Type to psl.
 - b. Copy the output text into a text editor to replace the space characters with tab characters so that it will copy directly into Microsoft Excel.
9. Copy over the sgRNA sequences and identifiers into the matching row of the BLAT output. Sort the resulting table based on genomic position.
10. Select sgRNA according to the needs of your experiment. To tile sgRNA across a locus, we aimed to design sgRNA every 50 to 100 bp and did not consider strand in our selection. When sgRNA are very close to each other or overlapping, we chose the one with the fewer off target sites (higher MIT score).
11. Order oligonucleotides for each sgRNA (we order from IDT).
 - a. DO NOT include the PAM sequence with the rest of the spacer sequence. Only clone the remaining 20 bp into the sgRNA expressing plasmid.
 - b. The spacer sequence that was entered into the BLAT tool but without the PAM sequence is the forward oligo. To this sequence, append ACAC to the 5' end and G to the 3' end. Total 25 nt.
 - c. Determine the reverse complement of the 20 bp spacer sequence. This is the reverse oligo. Append AAAAC to the 5' end of the spacer sequence. Total 25 nt.