nature portfolio

https://doi.org/10.1038/s41596-022-00741-3

Supplementary information

Enzyme-free targeted DNA demethylation using CRISPR–dCas9-based steric hindrance to identify DNA methylation marks causal to altered gene expression

In the format provided by the authors and unedited

Supplementary Methods for:

Enzyme-free targeted DNA demethylation using CRISPR/dCas9-based steric hindrance to identify DNA methylation marks causal to altered gene expression

Sapozhnikov & Szyf

This supplementary information contains Supplementary Methods detailing an example workflow for gRNA design.

Example workflow for gRNA design:

After identifying a target CpG of interest, we typically use IGV (https://software.broadinstitute.org/software/igv/) to extract the sequence, including approximately 50-100 base pairs on either side.

For example, cg25896754 is the Illumina Infinium MethylationEPIC BeadChip probe ID of a CpG in the human TP53 promoter (Extended Data Figure 1). Using IGV and genome build hg38, we navigate to the coordinates of cg25896754 by inputting them in the search field: chr17:7687559-7687560.

Next, to aid in visualization, we click Tools -> Find Motif and enter CG as the search pattern, followed by clicking OK. This highlights all CpGs in the region in a new track.

Next, we zoom in/out until a region of ~100-200 base pairs (bp) is visible: in this example, we chose a view that covers 98 bp (CAGTCTTGAGCACATGGGAGGGGAAAACCCCAATCCCATCAACCCCTGCGAGGCTCCTG GCACAAAGCTGGACAGTCGCCATGACAAGTAAGGGCAAG) (Extended Data Figure 1).

It is important not to manually pan the viewer, as the view is currently centered at the target CpG. Next, we press Ctrl+R. This produces a red bar near the top of the viewer: we right-click on the red bar and click “Copy sequence” to copy the visible sequence to the clipboard.

Next, we navigate to the gRNA design tool of choice: for us, it is CRISPick (https://portals.broadinstitute.org/gppx/crispick/public). Here, we select the following options: “Human GRCh38” as the “Reference Genome”, “CRISPRko” as the “Mechanism”, and “SpyoCas9 (NGG)” as the “Enzyme”. If using the recommended plenti-gRNA-puro plasmid for gRNA expression, select "Hsu (2013) tracrRNA": the scaffold encoded by this plasmid begins with GTTTT. If using an alternative plasmid (not recommended) check whether the scaffold begins with GTTTT or GTTTV and select the appropriate option, as explained in the CRISPick tool. Note that while gene knockout (“CRISPRko”) is not the true application in this protocol, it is necessary to choose this option for gRNA design in this workflow: the other two options (CRISPRa and CRISPRi) expect gene IDs rather than raw sequence and are therefore not appropriate. We also select “Bulk” in order to upload the target sequence and paste the sequence copied from IGV in the field that appears below. It is critical to check the box titled “Report unpicked sequences” as otherwise the tool will show only a reduced selection of gRNAs based on an intended application of gene knockout. After submission and typically <1 minute processing time, several results files are available for download. We download the first results file, titled “Picking Results”.

We then open this tab-delimited file with Microsoft Excel for simple visualization. The gRNA sequences can be found in column “T” and are already ranked from best to worst by the algorithm (e.g., by on-target and off-target activities). “Pick Order” and “Picking notes” are not relevant to gRNAs for demethylation. The only critical columns are:

1. T: gRNA sequence.
2. S: the gRNA “cut site”, which for dCas9 demethylation is simply useful for navigation purposes. Target cg2589675 is at the middle of the 98-bp region or at 49-bp (for the C in

CpG). If the cut site is indicated as, for example, 49, this means that the CG is near the 3' end of the gRNA: the cleavage site of the catalytically active Cas9 protein is 3 base pairs upstream of the PAM (i.e., at the third to last base in the gRNA sequence). This is useful as the gRNA with a cut site of 89 will clearly be too far from the CpG and will not sterically interfere with DNMT1 at the CpG. The physical spacing of all gRNAs relative to the target CpG can be simply visualized by copying each gRNA sequence from column "T" into the Tool->Find Motif query in IGV (Extended Data Figure 1).

However, columns with additional relevant information include:

1. R: orientation of the gRNA relative to the input sequence.
2. AB thru AQ: numbers of predicted off-targets as a factor of sequence similarity.
3. AS: On-target efficacy score.

The general recommendation here is to pick the highest-ranking gRNA that overlaps with the target CpG. If possible, it is better to include the CpG in the 3' end of the gRNA (near the "cut-site") and to have an on-target efficacy score >0.5. If no gRNA overlaps with the target CpG, it is possible to achieve demethylation with a gRNA for which the CpG is positioned within 5-bp from the 5' end and 10-bp from the 3' end (or 7-bp from the end of the PAM). If this is not possible, or if a user wishes to err on the side of caution and include the target CpG within the gRNA sequence, it may be necessary to use an orthologous dCas9 protein with different PAM requirements (such as S. aureus dCas9, N. meningitidis dCas9, F. novicida dCas12a, etc.). However, this will require different gRNA design, different plasmids, and, likely, a modified experimental design that are beyond the scope of this protocol.

In this particular example, we would elect to proceed with the gRNA sequence of CAATCCCATCAACCCCTGCG as it is ranked 2nd, positions the target CpG within the gRNA sequence, has a high on-target efficacy score, and importantly, is placed such that 5' and 3' adjacent CpGs would almost certainly fall outside of the steric interference footprint of dCas9 (the nearest CpG – seen in the "All CpGs" track – is >20-bp from the 3' end) and would thus not be predicted to be demethylated by this gRNA (Extended Data Figure 1); this gRNA would be predicted to drive specific demethylation of only the single target CpG.

The top-ranked gRNA – gRNA1 – will be useful to target the adjacent CpG if the researcher wishes to compare the contribution to gene expression of the methylation of these two CpGs or, more specifically, if the specific contribution of methylation of only the target CpG and not of the adjacent CpG is important to distinguish.

Multiple gRNAs can and should be picked for the development of an efficient and specific demethylation protocol for the target CpG: in this case, the 4th, 5th, 6th, and 7th ranked gRNAs would all be suitable for testing.