CRISPR DESINGNING YOUR Grna FOR CAS 9 GENOME EDITING USING Dna SENSORS TO DETECT CEREBROSPINAL FLUID MORBIDITIDITY INVITRO VIRAL GENE TRANSMISSION

Authors:

Dr.nisha mithal

Dr.sriram varahan

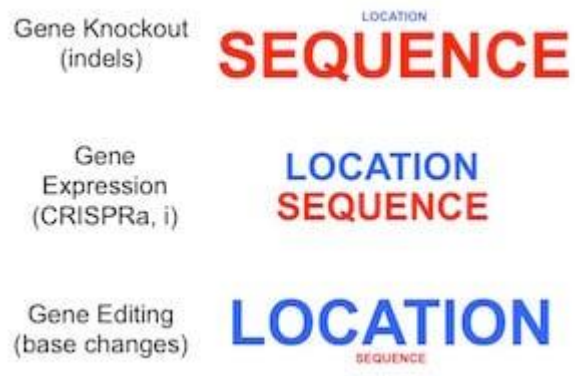
Dr.nivedha

INRODUCTION :

CRISPR technology has made it easier than ever both to engineer specific DNA edits and to perform functional screens to identify genes involved in a phenotype

interest. We will discuss differences between these approaches, and provide updates on how best todesign gRNAs. The hammer, the jigsaw, and the wrench are all great tools, but which one you use, of course, depends on what you are trying to do – there's no "best" tool among them. While this seems obvious, it is important to remember that the same is true when designing gRNAs for using CRISPR technology – the "best" gRNA depends an awful lot on what you are trying to do: gene knockout, a specific base edit, or modulation of gene expression. Location and sequence are important considerations for designing your gRNAs. For indels, it's not so important what location in the gene you target, but it is important that your gRNA sequence is designed to be highly active and reduce off targets. For CRISPRa and CRISPRi, these considerations are of roughly equal importance (target should be near the TSS but you can worry less about sequence optimality because you generally have fewer sequences to choose from). Finally, for HDR, location is much more important because you have to target within ~30 nt of your proposed edit, which means there are so few gRNAs to choose from that sequence preferences must largely be ignored.

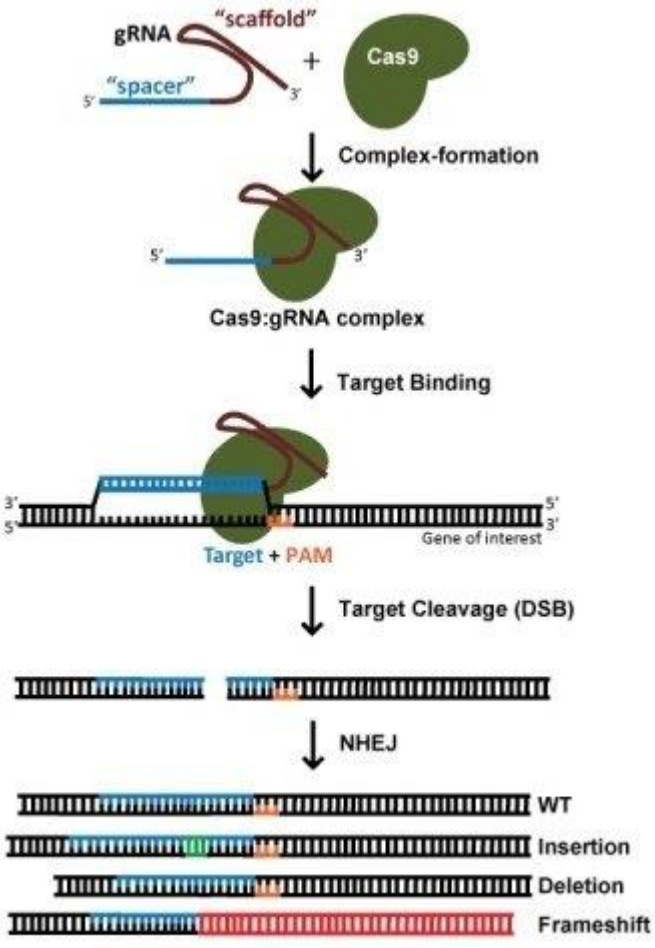
Picture 1:



The hammer: Gene knockout by NHEJ:

Gene knockout with CRISPR technology is usually accomplished by Cas9-mediated dsDNA breaks: following a cut, the error-prone nature of non-homologous end joining (NHEJ) often leads to the generation of indels and thus frameshifts that disrupt the protein-coding capacity of a locus. When using *S. pyogenes Cas9* (SpCas9), potential target sites are both [5'-20nt-NGG] and [5'-CCN-20nt], as it is equally efficacious to target the coding or non-coding strand. As a rule of thumb, we avoid targetarget sites are both [5'-20nt-NGG] and [5'-CCN-20nt], as it is equally efficacious to tt sites that code for amino acids near the N' terminus of the protein, in order to mitigate the ability of the cell to use an alternative ATG downstream of the annotated start codon. Likewise, we avoid target sites that code for amino acids close to the C' terminus of the protein, to maximize the chances of creating a non-functional allele. For a 1 kilobase gene, since potential target sites occur ~1 in every 8 nucleotides, restricting gRNAs to 5 – 65% of the protein coding region will still result in many dozens of gRNAs to choose from. With so many possibilities, picking a gRNA with an optimized sequence is of primary importance .

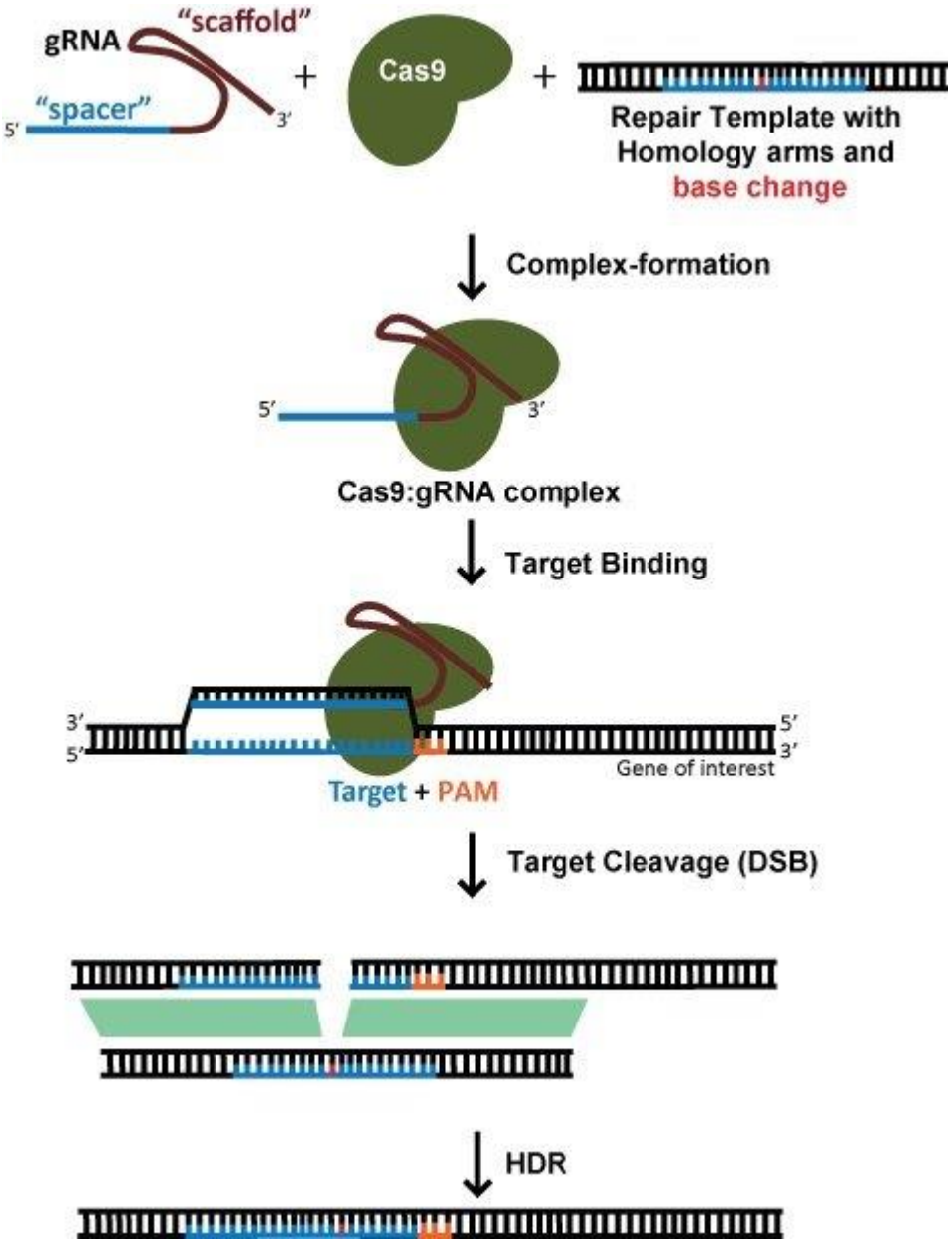
Picture 2



For a specific edit, such as the insertion of a fluorescent tag or the introduction of a specific mutation, one generally relies on homology directed repair (HDR) to incorporate new information into DNA. This also requires an exogenous DNA template. HDR, however, is a very low-efficiency process, and usually involves the need for single cell cloning and subsequent screening for successful edits. This is a very time consuming process and should not be undertaken lightly! Indeed, truly achieving the gold standard requires not one but *two* rounds of single cell cloning – as a control, one should revert the edit back to the original in order to prove that the phenotype was really due to the intended edit rather than some passenger variant that came along with the single cell clone (although this is rarely done).

When targeting a dsDNA break for HDR, the choice of target site is far more constrained by the desired location of edit; efficiency decreases dramatically when the cut site is >30nt from the proximal ends of the repair template (Yang et al., 2013). This means that, for gene editing, there are usually few potential gRNAs. While SpCas9, with a PAM preference of NGG, is still the most widely-used Cas enzyme, the development of SaCas9, NmeCas9, Cas12a enzymes, and engineered variants thereof offers additional PAM options that can greatly expand gRNA options.

Two newer technologies offer an alternative to HDR for introducing edits. The same locational constraints are even more exquisite for the so-called base editor Cas9, which makes DNA changes in the absence of dsDNA breaks (Rees et al., 2018). For C>T and A>G base editors, the intended edit must be in a 5 - 10 nt window relative to the PAM, and bystander edits are possible if there is another target C or A in the window. Another technology, prime editing (reviewed in Anzalone et al., 2020). is not limited to single nucleotide transitions but still requires a nearby PAM, although these are still early days for this technology, and the user may need to optimize numerous parameters to generate the desired edit.



The wrench: Gene activation and inhibition by CRISPRa and CRISPRi

Finally, for modulating gene expression at the level of transcription – CRISPRa (activation) and CRISPRi (inhibition) technologies – a nuclease-dead Cas9 (dCas9) is directed near the promoter of a target gene. Here, the target window is not quite as broad as for knockout via CRISPR cutting. For CRISPRa, it is most-efficacious to target a ~100nt window upstream of the transcription start site (TSS), while for CRISPRi, a ~100nt window downstream of the TSS gives the most activity. Thus, a given gene will only have a dozen or so gRNAs to choose from in the optimal location. It is also important to have good information on the exact location of the TSS. Different databases annotate the TSS in different ways, and it has been shown that the FANTOM database, which relies on CAGE-seq to directly capture the mRNA cap, provides the most accurate mapping (Radzisheuskaya et al., 2016). In this case,

location and sequence are of approximately equal importance in design – an optimized sequence will do little if it is in the wrong place, but because the target window is more-narrow, there are fewer gRNA to choose from, and thus an optimal sequence may not be available.

Predicting gRNA efficacy

We and others have examined the ability to use sequence-based and other features to nominate gRNAs that are likely to be active, not only for SpCas9 but also for some other Cas enzymes. It seems to be the case that there is no universal scoring system for selecting a gRNA, as the method of producing the guide (synthetic, in vitro transcription, or lentiviral delivery) can affect the accuracy of a predictive score, as well as dynamic aspects of the target (e.g. accessibility due to chromatin status). No computational prediction is ever perfect, but this can decrease the number of guides one needs to test in the lab.

Importantly, for any modification of interest, it would be unwise to make conclusions on the basis of the activity of a single gRNA, and thus diversity of gRNAs across a gene should be examined whenever possible when using knockout or transcriptional modulation approaches.

Avoiding off-target effects

The off-target activity of gRNAs is important to consider. While the basic landscape of mismatches that can nevertheless still lead to activity has been established, and can be used to identify sites that are likely to give rise to an off-target effect, there's not enough data to fully predict which sites will and will not show appreciable levels of modification. Whole-genome sequencing of cells modified by CRISPR indicates that the consequences of off-target activity, at least for the experimental conditions used, led to no detectable mutations (Veres et al., 2014). When working with single-cell clones, the authors note that "clonal heterogeneity may represent a more serious obstacle to the generation of truly isogenic cell lines than nuclease-mediated off-target effects." Further, large-scale datasets of hundreds of genetic screens using genome-wide libraries have shown high concordance between different sequences targeting the same gene, suggesting that off-target effects did not overwhelm true signal in these assays (Dempster et al., 2019). Again, the experimental strategy is clear: for any gene of interest, one should require that multiple gRNAs of different sequences give rise to the same phenotype in order to conclude that the phenotype is due to an on-target effect.



One advantage to using the CRISPR/Cas system for genome engineering is the fact that Cas9 can be easily programmed to make a DNA double strand break (DSB) in the genome wherever the user chooses. After the initial cut, the next steps in the process involve repairing chromosomal DSBs. It is important to know that cells possess two major repair pathways – Non-Homologous End Joining (NHEJ) and Homology Directed Repair (HDR) – and how these pathways work, as this could be relevant when planning your experiment. Here, we'll discuss NHEJ, and how it impacts what happens to Cas9-mediated DSBs in the genome.

CRISPR-Net: A Recurrent Convolutional Network Quantifies CRISPR Off-target Activities with Indels and Mismatches

This repository includes a recurrent convolutional neural network named CRISPR-Net for predicting the off-targets activities with insertions, deletions and mismatches in CRISPR/Cas9 gene editing. There are two command-line tools that can be used to quantify the off-target activities induced by CRISPR guide RNA. One is **CRISPR_Net.py** which can predict the off-target activities with indels and mismatches, the other is **CRISPR_Net_Aggregate.py** which aggregates the gRNA-target scores from CRISPR-Net into a single consensus off-target score.

PREREQUISITE

CRISPR-Net was conducted by Python 3.6 and Keras 2.2.4 (using TensorFlow 1.12 backend)

Following Python packages should be installed:

- scipy
- numpy
- pandas
- scikit-learn
- TensorFlow

Usage

CRISPR-Net can run with:

```
python CRISPR_net.py [-h] input_file
```

positional arguments: input_file

optional arguments: -h, --help show the help message and exit

Example input file

```
GAGT_CCGAGCAGAAGAAGAATGG,GAGTACCAAGTAGAAGAAAAATTT
GTTGCCCCACAGGGCAGTAAAGG,GTGGACACCCCGGGCAGGAAAGG
GGGTGGGGGGAGTTTGCTCCCGG,GTGTGGGGTAAATTTGCTCCCAG
GGGTGGGGGGAGTTTGCTCCAGG,AGGTGGGGTGA_TTTGCTCCAGG
GATGGTAGATGGAGACTCAGNGG,GGTAGGAAATGGAGAGGCAGAGG
GAGTCCGAGCAGAAGAAGAAAGG,GAGTTAGAGCAGAAGAAGAAAGG
```

Save it as 'input.txt'.

Now you can run CRISPR-Net as following:

```
$> python ./CRISPR_net.py input.txt
...
```

Then output will be generated:

- The first column of the output file indicates the on-target sequence,
- The second column of the output file indicates the off-target sequence,
- The third column is the off-taget score predicted by CRISPR-Net

and saved to ./CRISPR_net_results.csv:

```
on_seq,off_seq,CRISPR_Net_score
GAGT_CCGAGCAGAAGAAGAATGG,GAGTACCAAGTAGAAGAAAAATTT,0.000000e+00
GTTGCCCCACAGGGCAGTAAAGG,GTGGACACCCCGGGCAGGAAAGG,1.490116e-07
GGGTGGGGGGAGTTTGCTCCCGG,GTGTGGGGTAAATTTGCTCCCAG,6.078471e-01
GGGTGGGGGGAGTTTGCTCCAGG,AGGTGGGGTGA_TTTGCTCCAGG,3.451956e-01
GATGGTAGATGGAGACTCAGNGG,GGTAGGAAATGGAGAGGCAGAGG,5.029583e-05
GAGTCCGAGCAGAAGAAGAAAGG,GAGTTAGAGCAGAAGAAGAAAGG,9.269089e-01
```

CRISPR-Net-aggregate can run with:

```
python CRISPR_net_aggregate.py gRNA_offTargets.csv
```

positional arguments: gRNA_offTargets.csv

optional arguments: -h, --help show the help message and exit

Example input file:

```
off_target,Gene_mark,on_target
GAACTAGCCTTGTATCCCAGGGA,RP4-669L17.10,GACCTTGCATTGTACCCGAGGGG
GAACTAGCCTTGTATCCCAGGGA,RP5-857K21.4,GACCTTGCATTGTACCCGAGGGG
GTGTTTGCAATGTACCCGTGTTG,TTLL10-AS1,6,1173076,GACCTTGCATTGTACCCGAGTGG
GACCTGTGGTTGTTCCTGAGAGG,,GACCTTGCATTGTACCCGAGAGG
GCCCTTGGATTGGCCGCGAGGGC,,GACCTTGCATTGTACCCGAGGGG
```

Save it as 'gRNA_offTargets.csv'.

Now you can run CRISPR-Net-Aggregate as following:

```
$> python ./CRISPR_Net_aggregate.py gRNA_offTargets.csv
...
```

Then off-target aggregate score will be generated.

You can try CRISPR-Net-Aggregate on the our example file (input_example/aggregate_example_GACCTTGCATTGTACCCGAG.csv) by:

```
python CRISPR_net_aggregate.py ./input_example/aggregate_example_GACCTTGCAT
TGTACCCGAG.csv
```

CONTAINS:

- code/CRISPR_Net.py : Python script to run CRISPR-Net for predicting off-target activities with indels and mismathces
- code/CRISPR_Net_Aggregate.py : Python script to run CRISPR-Net-Aggregate that aggregates the gRNA-target scores from CRISPR-Net into a single consensus off-target score.
- code/evaluate_CRISPR_Net.py : Python script to retrain CRISPR-Net on any dataset in /data and evaluate its performance.
-

code/evaluate_CRISPR_Net_Aggregate.py : Python script to train and test CRISPR-Net-Aggregate

Non-homologous end joining steps

Unlike HDR, NHEJ is active throughout the cell cycle and has a higher capacity for repair, as there is no requirement for a repair template (sister chromatid or homologue) or extensive DNA synthesis. NHEJ also finishes repair of most types of breaks in tens of minutes – an order of magnitude faster than HDR. NHEJ is consequently the principle means by which CRISPR/Cas9-introduced breaks are repaired.

The following factors are required for NHEJ repair regardless of end structure, and dictate the major events of the pathway:

1. Broken ends are recognized by loading of the **Ku70/Ku80** heterodimer

2. Ku then acts as a scaffold for recruitment of a kinase (**DNA-PKcs)** and a two subunit DNA ligase **(XRCC4-ligase IV)**; together with some accessory factors (**PAXX, XLF**), this complex holds a pair of DNA ends together, forming a paired end complex
3. The paired end complex then ligates compatible DNA ends together, thus repairing the break.

This is a simplified, streamlined version of this pathway and does not consider the missing or damaged nucleotides that are common to biological sources of DSBs, and which need to be processed. Processing occurs prior to ligation as incompatible DNA ends interfere with that step. Accordingly, NHEJ has a vast toolbox of processing factors, including polymerases (Pol μ and Pol λ), nucleases (Artemis), and structure-specific end cleaning enzymes (Aprataxin, Tdp2) that function to make ends better substrates for ligation. Although we do not describe these steps here, the processing of DNA ends tends to be the point where mutations are introduced.

Repair of Cas9-induced breaks by NHEJ

As illustrated below, NHEJ-mediated repair of Cas9-generated breaks is useful if the intent is to make a knockout in your gene of interest, as it is prone to generating indel errors. Indel errors generated in the course of repair by NHEJ are typically small (1-10 bp) but extremely heterogeneous. There is consequently about a two-thirds chance of causing a frameshift mutation. Of some importance, the deletion can be less heterogeneous when constrained by sequence identities in flanking sequence ("microhomologies").

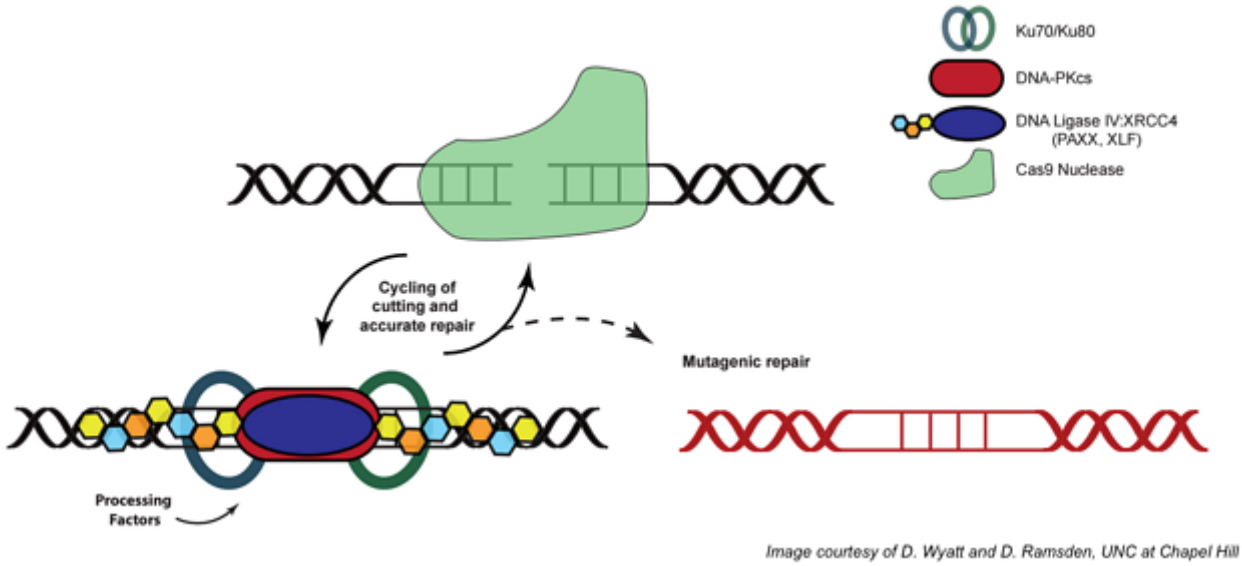


Image courtesy of D. Wyatt and D. Ramsden, UNC at Chapel Hill

NHEJ doesn't obligatorily introduce indels. Given the end structure of the Cas9 DSB (blunt or near-blunt ends without nucleotide damage) such products are rare, probably accounting for less than 5% of repair events. However, the products of

accurate repair are easily re-cleaved while indel products aren't, so repeated cycles will favor accumulation of the latter products. As noted above, a single cycle of cleavage and accurate repair should take less than an hour, thus a population of cells constitutively expressing a targeted Cas9 should possess indels in the majority of their chromosomes within a day. Another factor expected to impact repair is that the Cas9 protein doesn't immediately release from the broken end after cleavage, which may interfere with loading of Ku and normal NHEJ activity. Other factors can also influence NHEJ activity. For example, repressing RecA and overexpressing the NHEJ machinery improved NHEJ accuracy in *M*. *smegmatis* (Yan et al., 2020).

NHEJ can also be engaged by variants of the canonical Cas9 approach. A pair of CRISPR guides that flank regions of hundreds of base pairs or more can simultaneously introduce a pair of chromosome breaks, and could result in deletion of the intervening DNA ("pop-out" deletions) if NHEJ joins the distal ends together. Similarly, it may be possible to direct insertion of an exogenous DNA fragment at a Cas9 targeted break (or pair of breaks) by NHEJ-dependent repair ("pop-in" insertion) provided a template containing compatible overhangs is available. Cas9 can also be altered to generate a targeted single strand break; when two such breaks are introduced near each other, in opposite strands. This "double nickase" strategy vastly reduces breaks and mutations at off-target sites