Apr 01, 2022

🌐 How to select the best gRNA(s) for frameshift knockouts in zebrafish

Francois Kroll[1]

[1]University College London, University of London

👤 **Francois Kroll**
Sorbonne Université

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

[Create free account]

Protocol status: Working
Routinely using this approach since 2021.

Created: April 01, 2022

Last Modified: April 01, 2022

Protocol Integer ID: 60181

Keywords: frameshift knockouts in zebrafish, frameshift knockout, frameshift mutation, unique grna, zebrafish, using crispr, stable knockout line, grna, f0 knockout, cas9

Abstract

Remember to cite the protocol if it is helpful to you!

This protocol describes my approach (as of April 2022) to select gRNAs to generate frameshift knockouts in zebrafish using CRISPR-Cas9. It assumes you are selecting 3 gRNAs to generate F0 knockouts, following the method we described in doi: **10.7554/eLife.59683** and protocol **10.17504/protocols.io.bs2rngd6**.

Are you making a stable knockout line carrying a frameshift mutation? You can also follow this protocol to select a unique gRNA.

Get in touch for questions/suggestions
twitter – @francois_kroll
email – francois@kroll.be

Troubleshooting

1   Choose the exon(s) you want to target
_____

We will use the gene *appa* as an example.

Open the Ensembl page for your gene of interest, e.g.
https://www.ensembl.org/Danio_rerio/Gene/Summary?
db=core;g=ENSDARG00000104279;r=1:617620-656693

Look at the transcripts in the Genome Browser track, e.g. appa-203, appa-205, etc.

Pay attention to the direction of the gene. e.g. *appa* is reverse, because the transcripts are written as **<** appa-203. If it was forward, they would be written as **>** appa-203.

If the gene is forward, left-most exon is exon 1, i.e. read the transcript from left to right.

If the gene is reverse, right-most exon is exon 1, i.e. read the transcript from right to left.

Choose a few exons that you may want to target. If making F0 knockouts using the method in **10.7554/eLife.59683** you will (ideally) target three separate exons. Try to find 3–5 exons that make good targets.

Try to target (roughly in order of priority):

1. Exons that are common to all/most transcripts. Ignore transcripts that are not protein-coding (e.g. retained intron or processed transcript). Transcripts do not always overlap so it is not always possible to target exons common to all transcripts, but do the best you can. If doing F0 knockouts, try to have each transcript targeted by at least 1 gRNA.
2. Early exons. Make sure to read to the transcript in the right direction for this (see above).
3. Asymmetrical exons (see below).
4. Not exon 1, as there may be alternative start codons towards the beginning of the transcript.

It is rare to find more than 2 exons that fulfill all criteria, but follow as guidance to prioritise exons.

___

About 3. asymmetrical exons:

I originally read this idea in Tuladhar et al., 2019 doi: **10.1038/s41467-019-12028-5**.

An exon is *symmetrical* if its length is a multiple of 3, i.e. it has a round number of codons.
An exon is *asymmetrical* if its length is not a multiple of 3, i.e. it does not have a round number of codons.

Exon skipping is a common compensatory mechanism where the cell splices out an exon which contains a mutation. This could cancel your frameshift mutation by skipping the mutated exon altogether.

Skipping a symmetrical exon keeps the reading frame intact. For example, skipping a 300-bp exon deletes exactly 100 codons, it does not create a frameshift.

However, skipping an asymmetrical exon shifts the frame of the mRNA after the skipped exon. For example, skipping a 299-bp exon deletes 99 codons and 2 bp of the next codon, so shifts the reading frame.

Therefore, targeting asymmetrical exons is a strategy that nullifies exon skipping as a possible compensatory mechanism. For example, say you generate a 2-bp deletion in an asymmetrical (299-bp long) exon. If the exon is *not* skipped, the 2-bp deletion shifts the reading the frame; if the exon is skipped, the 299-bp deletion shifts the reading frame. Either way, you created a frameshift mutation.

This criterion is not essential when generating F0 knockouts with 3 gRNAs. As another two exons should also carry mutations, it is unlikely that skipping one mutated exons is sufficient to make an mRNA that encodes a functional protein. However, it is more important to target an asymmetrical exon when generating a stable knockout line with a unique gRNA.

In practice:

Choose the longest protein-coding transcript as reference for numbering the exons. For this, see length of the protein in the transcript table, e.g. transcript appa-202 codes for the longest protein, it is 682-amino acid long.

In the Genome Browser, you can also click on an exon to get its number.

Open the page of the transcript you chose as reference; e.g.

https://www.ensembl.org/Danio_rerio/Transcript/Summary?
db=core;g=ENSDARG00000104279;r=1:617620-656693;t=ENSDART00000166786

Go to *Sequence > Exons*, on the left column

Scroll down to the table. ENSDARE... are exon IDs. Look at the *Length* column.

Is the length of the exon a multiple of 3? Then the exon is **symmetrical**.
Is the length of the exon not a multiple of 3? Then the exon is termed **asymmetrical**.

Note; the first and last exon may include a UTR. Symmetrical/asymmetrical is only meaningful for protein-coding exons, so do not count the length of the UTR.

e.g. in transcript https://www.ensembl.org/Danio_rerio/Transcript/Exons?
db=core;g=ENSDARG00000024771;r=18:5213338-5227420;t=ENSDART00000033574
The first exon includes a 5'-UTR (in orange). Its length (217 bp) includes this UTR, so you need to substract the length of the UTR from 217 bp. Here the UTR is 60 bp, so the protein-coding part of this exon is 217 − 60 = 157 bp, i.e. the exon is asymmetrical.

You do not want to count the length of the UTR, as the ribosome does not read it.
E.g. here: https://www.ensembl.org/Danio_rerio/Transcript/Exons?
db=core;g=ENSDARG00000104279;r=1:617620-656693;t=ENSDART00000167331
Exon 5 has a UTR (orange), but it is counted in the length. So copy-paste just the protein-coding (blue) sequence and count the number of characters.

—

Following the criteria above, select at least one (for stable knockout) or at least three (for F0 knockout) exons that make good targets.

e.g. following appa-202 as reference transcript, I would select exon 6/19, exon 8/19, exon 15/19. Note that there is no exact answer.

2   **Select gRNA(s) in CHOPCHOP**
━━

Go to **CHOPCHOP**.

Target = your gene name, e.g. *appa*

In = Danio rerio (danRer11/GRCz11)

Using = CRISPR/Cas9

For = knock-out

In *Options*, under *Efficiency score*, select
Moreno-Mateos et al. 2015 - only for NGG PAM

Leave the rest as default.

Click *Find Target Sites!*

——

Why Moreno-Mateos et al. 2015?

This is the algorithm used by CHOPCHOP to calculate the efficiency (on-target) score. CRISPRScan (Moreno-Mateos et al. 2015 – **10.1038/nmeth.3543**) is the only tool which correlated with in-vivo results. See Uribe-Salazar et al., 2022 **10.1186/s12864-021-08238-1** Figure 2B. Even if CRISPRScan appears to be the best performing algorithm, note that it remains imprecise (correlation Illumina sequencing vs CRISPRScan: r = 0.27–0.31)

——

CHOPCHOP results: top of the page is the transcript map. Note that CHOPCHOP always shows the 5'–3' transcript, i.e. exon 1 is always the left-most. If the gene is reverse, this will not match the map in Ensembl, which shows the 5'–3' genome.

Zoom on each exon you selected in the previous step. It is easy to look at the wrong exon if the transcript is reversed compared to Ensembl. CHOPCHOP may also not use the same transcript as reference. To make sure you are looking at the correct exon, compare its positions on the CHOPCHOP map versus its coordinates in Ensembl (e.g. **appa-202**).

For each exon you want to target, open the best gRNAs (usually those in green) in new tabs. The CHOPCHOP ranking considers the efficiency score (column *Efficiency*) calculated by CRISPRScan so by doing so you will only select gRNAs with good efficiency scores.

Do not select gRNAs that are very close to the intron boundary as you risk making mutations in the intron, which would not lead to a frameshift. Consider the mutations will be centered around the double-strand break site (see **10.7554/eLife.59683** Figure 2–

figure supplement 1), which is at the PAM (the N of NGG) minus 4 bp, i.e. 4 bp into the gRNA binding site, starting from the arrow tip.

In the details of each gRNA (new tabs), look under *Shen et al. 2018 predictions of repair profile - statistics > Frameshift frequency*. This is predicted percentage of frameshift mutations. It is calculated using an algorithm called inDelphi (**10.1038/s41586-018-0686-x**). The algorithm was trained on data from mouse embryonic stem cells, not zebrafish embryos, but it achieves some accuracy in zebrafish embryos (see **10.1038/s41598-020-71412-0**).

Note; a gRNA which makes indels of random lengths will achieve 66% frameshift frequency**. Accordingly, any frameshift frequency below 66 is worse than random. Try to select gRNAs with high predicted frameshift frequencies, ideally > 80%.

Proceed by elimination and select 3 (for F0 knockouts) or 1 (for stable knockout line) gRNA matching as many criteria as possible. Consider also the exons from your selection. For example, I would probably select a gRNA with very high (> 85) predicted frameshift frequency and high efficiency score (> 0.60) even if it is in an asymmetrical exon. Alternatively, if you absolutely need to target a specific exon so each protein-coding transcript is targeted, you may have to choose a gRNA without great frameshift frequency for this exon.

** Consider: a 1-bp deletion is a frameshift, a 2-bp deletion is a frameshift, a 3-bp deletion is *not* a frameshift, and so on.

——

About off-targets:

In the CHOPCHOP results, MM0 is the number of off-targets with 0 mismatch with the gRNA sequence (i.e. same as the on-target), MM1 is the number of off-targets with 1 mismatch with the gRNA sequence, etc.

In **10.7554/eLife.59683** Figure 2F; we sequenced the top three off-targets of three gRNAs, for a total of 9 off-targets. Only one was significantly mutated (first off-target of gRNA D). Incidentally, this off-target was the only one with 2 mismatches, while the other 8 had 3–4 mismatches with gRNA sequence.

Accordingly, it seems safe to select gRNAs which only have off-targets with 3 or more mismatches. Even safer if you can avoid having any off-targets with 3 mismatches. In CHOPCHOP, aim to select gRNAs with all 0 in columns MM0, MM1, MM2, MM3.

Part of **SPRINGER NATURE**

3   **Check for SNPs in your target**
     _____

This step is optional, but safer.

It uses a database of zebrafish SNPs: SNPfisher (**10.1242/dev.118786**). Website:
https://snpfisher.nichd.nih.gov/snpfisher/snpfisher.html.

However, the database was built in 2015 so used the Zv9 version of the zebrafish
reference genome. Coordinates from CHOPCHOP refer to danRer11/GRCz11.

To find the coordinates of your target on the Zv9 genome, one solution is simply to
search for that sequence in the Zv9 genome using BLAST.

Get the sequence of your target, including the PAM. e.g.
TATCGGTCCACTGCATCCGG**AGG**
(PAM in bold)

BLAST this sequence in genome Zv9, using

https://blast.ncbi.nlm.nih.gov/Blast.cgi?
PAGE_TYPE=BlastSearch&PROG_DEF=blastn&BLAST_SPEC=Assembly&UID=3943314

The first result in the table is usually the correct one. Click on it. Confirm it is a perfect
match by checking that the entire sequence (23 nucleotides) matches. If you BLAST the
sequence above as example and click on the first result, you will see Identities
23/23(100%). If the sequence does not entirely match, the sequence was not found in
the Zv9 genome and you should skip this step entirely.

The chromosome numbers seem to have stayed the same between Zv9 and
danRer11/GRCz11, so if the sequence matches entirely and your gene of interest is on
that chromosome, it very likely found the correct sequence.

Write down the genomic positions it finds. In the example above, you should get
chr1:504995-505017.

Note, if the gRNA is reverse, you may need to flip the positions so they are increasing.

Search for SNPs in that window in the SNPfisher database.

Go to https://snpfisher.nichd.nih.gov/snpfisher/snpfisher.html.

_____

Type in the genomic window in the box, e.g. chr1:504995-505017.

Click *Look for Strain Differences*.

Does it find any SNP in that window?

If yes – you may want to skip that gRNA and choose another one from the CHOPCHOP results.

Note; what you decide may depend on the allele frequencies in SNPFisher.

Column *FLI_FQ* = allele frequency in *Tg*(*fli1a-eGFP*[y1]) transgenic line
Column *WIK_FQ* = allele frequency in the WIK wild-type strain
Column *TL_FQ* = allele frequency in the *Tüpfel long fin* wild-type strain

For example, if you use TL wild types and the SNP has frequency > 0, I would strongly recommend selecting another gRNA from CHOPCHOP. However, if the SNP was only found in other stains, it might be ok to use this gRNA.

4   **Order your gRNA(s)**

━━

This will be somewhat specific to where you buy your gRNAs. I buy them from Integrated DNA Technologies (IDT).

On the IDT website, go to *CRISPR-Cas9* > *Check your own design*; here: https://eu.idtdna.com/site/order/designtool/index/CRISPR_SEQUENCE

Change *Species* to *Danio rerio*.

Enter your sequence(s) as for example:

> mygene_1
TATCGGTCCACTGCATCCGG

Note; in the gRNA page from CHOPCHOP, the target sequence includes the PAM. For example TATCGGTCCACTGCATCCGG**AGG**. You want to remove the PAM, i.e the last three nucleotides which should always be NGG (any nucleotide / guanine / guanine). The IDT tool only accepts 20-bp sequences so you will receive an error if you forget.

Click *Check*.

It is likely that you get a warning like

"This gRNA may have low on-target performance."

Ignore it. The on-target score from IDT does not achieve any meaningful correlation with in vivo results (see benchmarking by Uribe-Salazar et al., 2022 **10.1186/s12864-021-08238-1** Figure 2B).

—

For instructions on how to prepare the Cas9/gRNA complex, refer to protocol **dx.doi.org/10.17504/protocols.io.bs2rngd6**.

5   **FAQ/troubleshooting**
━━

In CHOPCHOP, all (or great majority of) my gRNAs have an off-target with 0 mismatch. What should I do?

You probably do not have to worry, this off-target is likely fake. Open a few gRNAs in new tabs and look where the off-target is (under section *Off-targets*).
If
1) you see a suspicious chromosome name such as *chr9_KZ115297v1_alt*
2) it is always the same suspicious chromosome across multiple gRNAs
Then it is probably what I call a 'ghost copy' of your gene of interest. The sequence was copied on an alternative chromosome during assembly of the reference genome. In other words, you can ignore the off-target with 0 mismatch, the sequence is not physically present in the genome.

CHOPCHOP cannot find my gene; it says "Error status: 501. Error: Your gene Id: XXX has not been found in our database."

Go to Ensembl and find your gene there. Get the Ensembl ID, e.g. ENSDARG00000013892. Try searching in CHOPCHOP with that, in lieu of the gene name.
If CHOPCHOP still cannot find your gene, you can get the sequence of each exon you want to target and search for gRNAs this way. Just change in CHOPCHOP by clicking *Paste target*.