Jens's guide to CRISPR/Cas9 mediated genome editing

CRISPR/Cas9 mediated genome editing allows the precise insertion of DNA into the human genome. In most instances we use it to introduce tags (for example the HaloTag, or other fluorescent proteins) at the endogenous loci of genes coding for proteins that we would like to study. But it can be used for gene-knockout or the introduction of specific mutations as well. This guide is a step-by-step tutorial to carry out CRISPR/Cas9 mediated genome editing in human tissue culture cells.

CRISPR/Cas9 mediated genome editing requires the introduction of a double stranded break into the human genome, ideally very close to the site at which one wants to insert foreign DNA. This double stranded break is made by the endonuclease Cas9 and is guided by the sgRNA, which is complementary to site that it Cas9 will cut. After the introduction of the double stranded break, foreign DNA is introduced by homology directed repair (HDR) using a repair template containing the foreign DNA, which is provided in trans. For more information on HDR please read http://blog.addgene.org/crispr-101-homology-directed-repair. This tutorial will take you through designing and making plasmids that will serve as the HDR template and that will express Cas9 and the sgRNA. In addition, this guide includes information on the protocols we use to generate clonal genome edited cell lines.

1. Selecting the appropriate tag

The tag constructs we use in the lab all include a SV40 promoter driven puromycin resistance gene, which is flanked by LoxP site, allowing removal of the cassette after selection is complete. Selection with puromycin is extremely efficient in enriching for cells that have integrated the HDR donor. For N-terminal tagging of the protein of interest the puromycin resistance cassette is contained within the 3xFLAG-HaloTag (Fig. 1A) and after removal by CRE-recombination provides a short linker between the 3xFLAG and HaloTag. Thus, the promoter region of the tagged gene is completely untouched by the genome editing procedure. A downside of this approach for N-terminal tagging is that, due to the insertion of an expression cassette within its coding sequence, the protein of interested will be non-functional while the puromycin resistance cassette is present. This N-terminal tagging strategy is therefor only recommended for non-essential proteins. C-terminal tagging in the other hand is compatible with essential genes since the resistance marker cassette is inserted

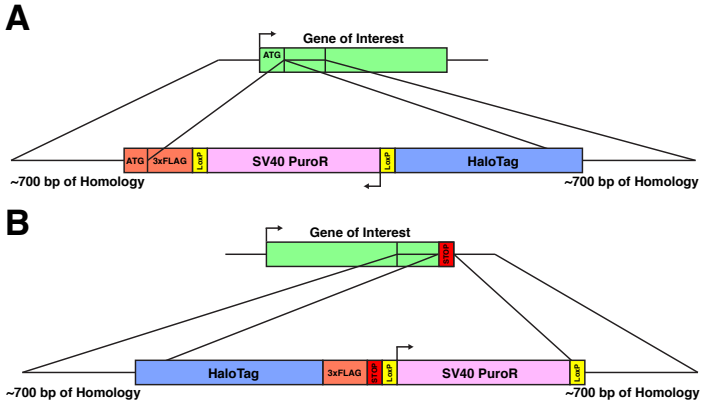

Figure 1. HaloTag design for (A) N-terminal and (B) C-terminal fusion proteins. SV40 driven puromycin resistance markers is contained within the tag coding sequence for the N-terminal tag cassette and after the STOP codon for the C-terminal tag cassette. ~700 bp of homology to the gene of interest can be introduced my PCR amplifying the tag cassette with primers that include ~60 bp of homology to the endogenous locus of the gene that will be tagged.

downstream of the STOP codon of the protein of interest (Fig. 1B). Both the N-terminal and C-terminal tags contain a linker sequence, including a TEV-protease cleavage site, that should facilitate flexible fusion with many different proteins, ideally without affecting functionality of the tagged protein. Functionality of the fusion-protein should be verified after tagging with essays appropriate for the respective protein.

It is also possible to introduce a fluorescent tag without including a selectable marker. I would only suggest going this route if the protein of interest is highly expressed and it is reasonable to assume that cells that have undergone HDR can easily by identified using FACS to detect the expression of the fluorescently tagged protein.

2. Downloading the genomic sequence of your target gene from the genome browser

After choosing the tag you want to introduce into the human genome based on the nature of your target protein, you have to download the genomic sequence that codes for your protein of interest. Go to UCSC genome browser website (https://genome.ucsc.edu), select genome browser and enter the name of the protein you want to tag into the search prompt. The gene will appear as a set of boxes and lines. Boxes indicate exons, lines indicate introns. Narrow boxes are untranslated regions (5' and 3' UTRs) and wide boxes are the coding region (Fig. 2). Zoom into the START codon (transition from narrow to wide box), to identify the position of the insertion of the N-terminal tag. To download the genomic sequence of the gene, click on the gene name next to the top line of the genome browser (Fig. 2, red arrow). This will bring up a page with a more detailed description of the genomic element selected. To access the sequence, click on the genomic sequence description (Fig. 3, left, red arrow). On the next page make sure to include at least 1000 bp upstream of the transcriptional start site by checking the corresponding box (Fig. 3, right, top arrow), and then click submit (Fig. 3, right, top arrow). This will open a new page that displays the genomic sequence in FASTA format, with all exons in capital letters and introns in lower case letters. Copy this sequence into your favorite cloning tool (SnapGene, ApE, Benchling, etc.), and annotate the START codon of your gene. To avoid making mistakes due to different isoforms or splice variants,

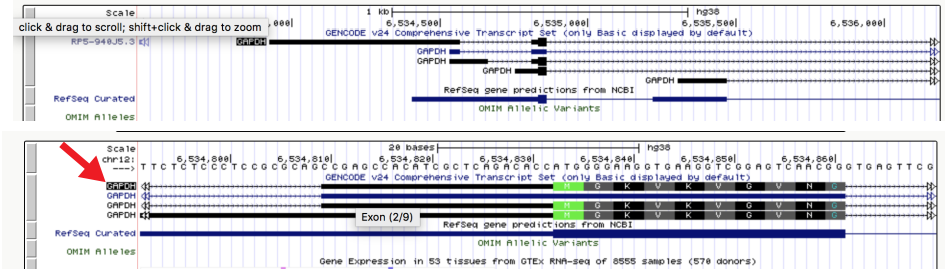

Figure 2. Using the genome browser to inspect the genomic region of your protein.

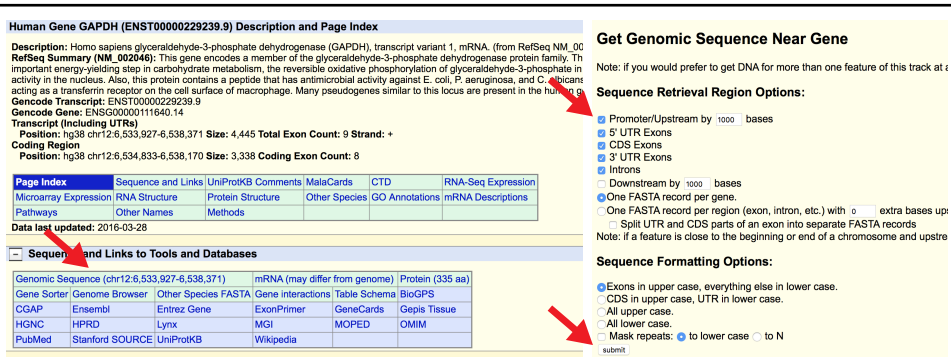

Figure 2. Downloading the genomic sequence of your protein of interest.

double check the protein sequence with an independent tool, for example the UNIPROT data base (http://www.uniprot.org). For introducing a C-terminal tag annotate the STOP codon of your gene.

3. Designing the HDR donor template

To introduce HaloTags by genome editing we generate plasmids including the tag, as well as ~700 bp homology arms flanking both sided of the tag. This can be achieved by Gibson assembly following PCR amplification of the tag and the homology arms from genomic DNA of the cell lines that one wants to edit.

Gibson assembly is a technique that can be used to paste together linear pieces of DNA that contain small regions of overlapping sequences at their ends. For more information please read the extensive protocols available elsewhere. We assemble the tag sequence as well as the ~700 bp homology arms into a vector backbone in a single step. The homology arms are amplified from genomic DNA isolated from the cell line we wish to edit, to assure any cell line specific genomic alterations are maintained. To streamline the process, we have one set of primers for the tag and include the overlapping sequence with the tag and the vector backbone in the primers for the homology arms.

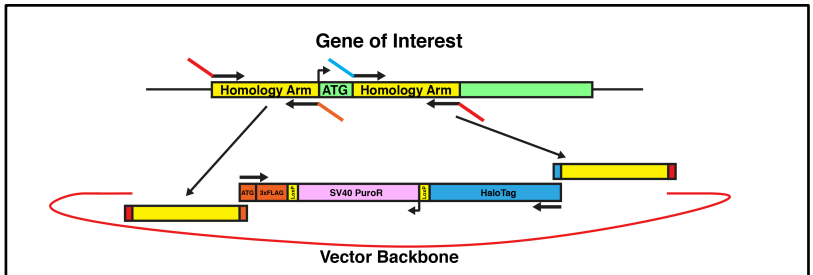

Figure 3. Gibson Assembly of HDR Donor

For inserting the C-terminal tag outlined in figure 1, the overall approach is the same. Be sure to include a STOP codon at the correct position. One consideration is whether or not to include a poly-adenylation signal after the tag, to assure immediate expression of the tagged protein. It is not necessary to excise the selectable marker after C-terminal tagging, but be aware that any regulation mediated by the 3'-UTR of the tagged gene is likely lost.

4. Designing the sgRNA and cloning of the Cas9/sgRNA plasmid

For designing the sgRNA to introduce the double stranded break that will trigger HDR at the gene you want to tag, there are a few important considerations. First, the double stranded break should be as close as possible to the site of integration, especially when using HDR repair templates with only 60 bp of homology. The second consideration is whether any untagged alleles of the gene you are targeting should be rendered non-functional, or whether they must stay intact. For N-terminal tagging, if the sgRNA is designed to cut within the coding region of the gene immediately downstream of the ATG, alleles that are not repaired by HDR using the donor template, will likely be knocked out, due to short insertions or deletions (indels) leading to frame shifts (Fig. 4). If an essential gene must be

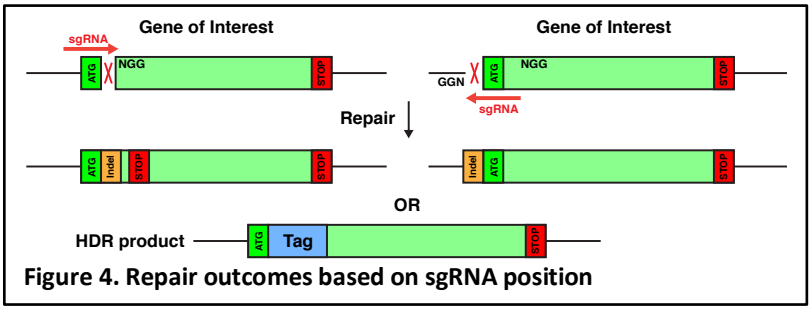

Figure 4. Repair outcomes based on sgRNA position

tagged at the N-terminus, the sgRNA has to be designed to cut upstream of the START codon of the gene (Fig. 4). Small indels upstream of the ATG will not change the coding sequence of your gene of interest but might lead to expression level changes. Importantly, this approach will result in **both tagged and untagged protein being expressed**.

Most Cas9 variants use a 20 bp recognition sequence, which is immediately upstream of the protospacer adjacent motif (PAM), a three-nucleotide sequence that is invariant for a particular Cas9 variant. We typically use the pX330 (Addgene #42230) from Feng Zhang's group to express the sgRNA and Cas9. This version of the Cas9 nuclease uses an NGG as its PAM, meaning the sgRNA will be the 20 bases immediately upstream of the NGG. Cas9 will cut 3-4 bases upstream of the NGG, which must be taken into consideration when determining whether the Cas9 cut will occur before, or within the coding region of the gene (Fig. 4). When identifying potential sgRNA sequences, don't forget about looking at the antisense strand as an option. A PAM on the antisense strand is a CCN sequence on the sense strand. If possible, we design the sgRNA to overlap the integration site of the tag (Fig. 5). That way the sgRNA sequence will be split apart by the integration and Cas9 will not cut the genome edited alleles once it has been generated by HDR. Alternatively, a silent mutation can be introduced in the PAM sequence of the HDR donor, to prevent the genome edited allele from being targeted by Cas9. Once you have identified a suitable sgRNA sequence, two oligonucleotides have to be ordered and ligated into the pX330 vector. General cloning instructions for pX330 can be found Feng Zhang's Addgene site (https://www.addgene.org/crispr/zhang/). We follow the general design principle they outline, and we always include a G upstream of the 20-bp target sequence (i.e. the forward primer will be CACC**G**-20-bp target, and the reverse primer AAAC-reverse complement of 20-bp target-**C**). We have found that digesting pX330 with BbsI followed by gel purification of the linearized product yields low background backbone plasmid even without treating the backbone with CIP. And we use unphosphorylated oligos for the ligation. Be aware that, if you treat the backbone with CIP, you must either phosphorylate the oligos to be ligated with PNK, or order oligos with a 5'-phosphate, which adds cost. Once you have cloned the sgRNA/Cas9 plasmid generate a transfection quality stock of it (ideally Maxiprep DNA).

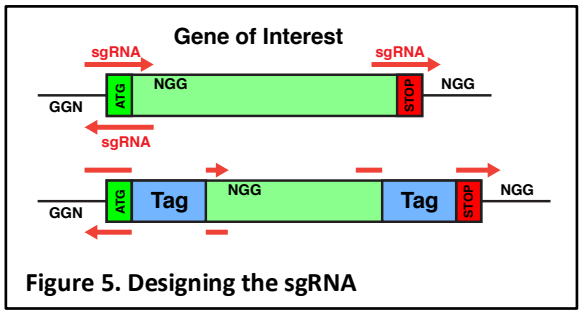

Figure 5. Designing the sgRNA

5. PCR amplification of the HDR donor

A plasmid encoding the N-terminal 3xFLAG-LoxP-SV40-Puro-LoxP-HaloTag is available on Addgene (#86843), for the C-terminal tag vector please contact Jens Schmidt (schmi706@msu.edu). Amplify the tag and homology arms, followed by gel purification of the product. Ideally carry out a large enough volume of PCR to generate a gel purified product with >100 ng/µl concentration, we typically elute in 30 µl of TE. The vector backbone we use (pFastBac) is linearized using a blunt cutting enzyme (HpaI), treated with CIP, and gel purified. We then assemble the vector with the tag and the two homology arms in a single Gibson assembly according to the vendor protocol (NEB).

6. Generating clonal genome edited cell lines

Now that you have both the sgRNA/Cas9 and HDR donors in hand you are ready to edit the genome of your favorite cell line. In most cases we have used HeLa cells, but our approach has also been carried out with U2OS and HEK293T cells, and likely will work with most human cell lines.

A. Transfection and Selection

We usually transfect 1-2x10^6 cells using either Lipofectamine 2000 or Nucleofection (Lonza) with 1 µg of the sgRNA/Cas9 plasmid and 1 µg of the HDR donor. After transfection we culture the cells for 48 hours before starting selection with puromycin (for HeLa cells we use 1 µg/ml, but this might be different for your cell line). To control for non-specific HDR donor integration, we always include a control in which we leave out the sgRNA/Cas9 plasmid. After 4-7 days of selection genome edited cells will have grown up, while all control cells should have perished. Proper integration of the tag can be verified by PCR using a primer within the tag cassette and a second primer well outside of the homology region that is included in the HDR donor (Fig. 6, black or red primer pair), after DNA extraction from the polyclonal population (we use genomic DNA miniprep kits from either Qiagen or Sigma). For the PCR from genomic DNA we typically use Phusion polymerase with 30 seconds amplification per kb of amplicon for 35 cycles. By default we include DMSO and GC-rich buffer.

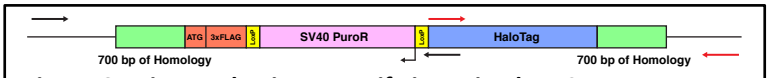

Figure 6. Primer selection to verify insertion by PCR

B. Resistance marker excision and generating single cell clones

To excise the puromycin resistance cassette and to generate single cell clones we transfect a plasmid encoding for a eGFP-CRE fusion protein (Addgene #11923). We usually transfect 2 µg of eGFP-CRE plasmid into 1-2x10^6 cells using Lipofectamine 2000. We culture the cells for 24-48 hours after transfection and then sort single GFP positive cells into 96-well dishes using FACS. The FACS conditions will strongly depend on the cell line you are using. For HeLa cells, we use a 1:1 mixture of fresh and filtered conditioned media (media that HeLa cell have previously been cultured in), and we always spin the plates in the centrifuge at 1000 rpm for 5 min, immediately after cell sorting. In the first well of the plate we add 100 cells, this serves as sorting control, since you can visualize these cells with the microscope right after spinning the plate. We culture the single cell clones for ~2 weeks in the 96 well plates. After 1 week it is possible to screen through the plate and identify wells that harbor single cell clones, which are typically round colonies. If a well contains multiple round colonies or oddly shaped colonies we do not use them, since they are likely derived from more than one cell. Single cell clones can then be expanded and the proper integration of the tag is verified by PCR and sanger sequencing of the amplified products. To analyze whether there are untagged alleles and if so whether they are mutated (Fig. 4) we carry out PCR reactions with a primer set outside of the homology arms (Fig. 6, black forward primer, red reverse primer). Single cell clones that are correct by PCR and sequencing are further analyzed by SDS-page along with fluorescent HaloTag labeling and Western blotting.