

Stem Cells & Gene Editing Workshop

Workshop webpage: <https://www.allencell.org/scge-workshop-2019.html>

Introduction

Welcome workshop attendees! The first workshop session will be an interactive seminar on endogenous gene tagging with a focus on the design of crRNAs and donor templates for CRISPR/Cas9 knock-in. **Bring a laptop** with a fully charged battery (a limited number of power outlets will be available) and **complete the following short tasks (5-10min) before arriving at the Allen Institute for the workshop:**

1. You will need software for viewing, annotating, and editing sequences. You are welcome to use any software that you already have and are experienced with. We will be using Geneious throughout the workshop. If you want to follow along with us and use Geneious (recommended), **download the Geneious software and get a free 14 day trial license.**
[Download Geneious](https://www.geneious.com/download/) (<https://www.geneious.com/download/>)
[Geneious trial license](https://www.geneious.com/free-trial/) (<https://www.geneious.com/free-trial/>)
2. Open Geneious, and under Local, **create a new folder** called GE_workshop (right click → New Folder)
3. **Download sequences** below from workshop webpage (Right click → Save Link As). Both sequences are in genbank format (file ending with .gb; eg. nup153.gb).
[NUP153 sequence](#)
[mEGFP sequence](#)
4. **Import NUP153 and mEGFP** sequences into Geneious. Select GE_workshop folder → File → Import → From File (or Ctrl+i). Keep sequences separate.

Workshop Design Instructions

We will design **crRNAs** and a **donor template** for tagging the N-terminus of NUP153 with mEGFP (linker = YSDLE).

crRNA Design

Section 1: Editing design pre-requisite

1. Download NUP153 sequence from NCBI
 - Go to [NCBI gene](https://www.ncbi.nlm.nih.gov/gene) (<https://www.ncbi.nlm.nih.gov/gene>) and search for **NUP153**
 - On the search results page, select the *Homo sapiens* NUP153 (entrez id 9972) by **clicking directly on the gene symbol**.
 - Scroll down the page to the section titled **Genomic regions, transcripts, and products**
 - On the right side, there are several options for accessing the nucleotide sequence. Find **Go to nucleotide: Graphics FASTA GenBank** and click on the **GenBank** option
 - Locate **Change region shown** section on the right side. Change the selected region to

from 17613035 to 17708834 and click on Update View. Wait for view to load and then download by clicking on **Send to → Complete Record → Choose Destination (File) → Format (GenBank) → Create File**

2. If you have not downloaded mEGFP sequence, download it now: [mEGFP sequence](#)
3. View NUP153 and mEGFP sequences and annotations
 - Open Geneious and under Local, create a new folder called GE_workshop (right click → New Folder)
 - **Import NUP153 and mEGFP into Geneious:** select GE_workshop folder → File → Import → From File (or Ctrl i).
 - **Locate and annotate desired insertion site** at N-terminus of NUP153.
4. Check for WTC specific variants near insertion site and in homology arms (1kb on each side of insertion site). UCSC track of WTC variants is available [here](https://www.allencell.org/genomics.html) (https://www.allencell.org/genomics.html):
 - Click on link for **WTC whole genome** to load UCSC WTC variant track
 - GRCh38 genomic coordinates for NUP153 insertion site are chr6:17705384-17707384; **copy and paste genomic coordinates** (underlined) into UCSC search bar
 - Locate WTC parental line variant track

Section 2: Find crRNAs with CRISPOR design tool

1. Download CRISPOR results file (genbank format) and import into Geneious [CRISPOR results](#)
2. We will look for crRNAs (w/ NGG PAM) in sequence 50bp on each side of desired insertion using GRCh38 genome
3. Identify candidate guides near the insertion site with high specificity scores
 - CRISPOR identifies crRNAs using their position and strand in the query sequence (ex. 2/rev). crRNAs near 50 will be closest to the desired insertion site.
 - Hover over a crRNA in Geneious to see the MIT specificity score (higher score means more specific). Ideally, specificity score should be > 50. Try clicking on the CRISPOR results url in notes.
 - Annotate 5 guides in Geneious. You will use these guides in the donor template design section. These are the 5 guides we will use for the donor template:
 - i. CR1: TCCGACTCCTCCGGCTCCCG
 - ii. CR2: TCCGGCTCCCGAGGCCATGG
 - iii. CR3: AAGCGGCGGAGGCTCCGCCA
 - iv. CR4: GGAGGCTCCGCCATGGCCTC
 - v. CR5: GCCATGGCCTCGGGAGCCGG

Donor Template Design

Section 1: Finish Annotations on NCBI File

1. Locate file (from NCBI or Ensembl) that contains the locus of interest and is annotated with cell line-specific SNPs (if available), and annotated with potential crRNA targeting regions in your software of choice (referred to as “the NCBI file” for this checklist). If

you did not complete the crRNA design portion, the annotated file is below:

[Annotated NUP153 NCBI sequence](#)

2. Annotate the desired insertion site in the NCBI file (if not already done)
3. Annotate the 5' (upstream) homology arm in the NCBI file – recommended 1000 bp
4. Annotate the 3' (downstream) homology arm in the NCBI file – recommended 1000 bp

Section 2: Cut and paste together the new donor template sequence

1. Create a new sequence file, titled as the “donor template”
 - In Geneious, Sequence → New Sequence or Ctl+N
 - In Geneious, you cannot have a blank file, so just type “CAT” as the sequence for now
2. Back in the NCBI file, copy the 5' homology arm (select annotation, hit Ctl+C)
3. Go back to the donor template file and replace “CAT” with the copied 5' homology arm (can use Ctl+V); *you must “Allow Editing” to make any changes to the sequence in this and subsequent steps* – Save file
4. Make sure the translation option is turned on, and the sequence is showing up in the correct frame in the donor template file
 - Under the General (home icon) tab in Geneious, select the “Translation” check box
 - Click on “Options >” next to translation and use the “Frame:” drop down to select the frame to get your sequence properly aligned (should be Frame 2 for our example)
5. Locate the sequence file for mEGFP; copy the annotated sequence *without* a start or stop codon
6. Back in the donor template file, copy the mEGFP sequence to the 3' end of the 5' homology arm – Save file
7. To add the linker sequence, use the codon optimization table to “reverse translate” the amino acid sequence into DNA (see table, below). With “Allow Editing” enabled, you can simply type in the DNA sequence you wish to add - Save file when done
 - e.g. for our example, in human: **Y= TAC, S= AGC, D= GAC, L= CTG, E= GAG**
 - **Cut and paste full linker sequence: TACAGCGACCTGGAG**
 - Avoid rare codons for your species; there are also online tools that can help with this for long linkers, but generally we do this manually
 - [Example codon frequency table for Homo sapiens](#)
(<https://www.genscript.com/tools/codon-frequency-table>) Avoid rare codons (frequency less than 0.1) when making changes to the donor plasmid sequence.
8. Back in the NCBI file, copy the 3' homology arm (select annotation, hit Ctl+C)
9. In the donor template file, place the cursor at the very end of the linker sequence, and paste in the 3' homology arm

Section 3. Introduce synonymous or silent mutations into the donor template homology arm

At this point, the donor template sequence is nearly complete, it should have an mEGFP sequence (no start or stop codon) and a linker flanked by the 5' and 3' homology arms. No bases from the original NCBI sequence should be excluded. The endogenous start or stop codon should be used if the insertion is directly at the N- or C- terminus, respectively. Next, we will check the crRNA target sites for continuity in the donor template sequence (in other words,

whether this crRNA will still be able to “recognize” the target sequence in the donor template). Mutations to the DNA sequence should be introduced to prevent CRISPR/Cas9 recognition of the donor template to avoid degradation of the donor during transfection. These introduced mutations will also prevent CRISPR/Cas9 cutting of the newly inserted sequence in edited cells. If these mutations are not introduced there is a risk of either degrading your donor template, or having NHEJ damage on edited alleles.

1. crRNAs that are disrupted by the insertion of the tag/linker should be checked to make sure there is not still sequence homology present for binding in the disrupted region (CR2, CR4, and CR5 are examples for NUP153)
2. Mutate the target site(s) for crRNAs that bind in the coding region and are not disrupted by the insertion of the tag/linker (CR1 is an example for NUP153)
 - First, can a mutation be made to the PAM? If yes, avoid changing to NAG or NGA, as these may still have some efficacy.
 - *In our example, change the AGG PAM to ATG (on the reverse strand), this preserves the A codon*
 - *Replace the GCC Alanine codon with a GCA Alanine codon (on the forward strand)*
 - If no changes to the PAM can be made, make 3 synonymous mutations in the seed region (10 bp proximal to PAM), avoiding rare codons.
3. Mutate the target site(s) for crRNAs that bind in the non-coding regions and are not disrupted by the insertion of the tag/linker (crRNA3 is an example for NUP153).
 - Before making any changes, check the conservation of sequence in the UCSC Genome Browser
 - i. Go to the UCSC Genome Browser's [BLAT tool](https://genome.ucsc.edu/cgi-bin/hgBlat?command=start) (https://genome.ucsc.edu/cgi-bin/hgBlat?command=start)
 - ii. Highlight a region of sequence you want to check in your template, and use the Blat Tool to search for it
 - iii. Scroll down to the “Comparative Genomics” section and put the “Conservation” track to “full” → refresh
 - If regions of low conservation exist, and changes can be made to the PAM or crRNA binding region, make the changes as described above for coding regions. Synonymous mutations are not important in non-coding regions.
 - If the region is too highly conserved (as is common for N-termini), choose different crRNA(s) for the experiment. *This is the case for our example; we would not be comfortable disrupting this non-coding region; omit CR3 from the experiment.*
4. Confirm that your entire design is in-frame

That's it! This template sequence can now be used to design PCR amplicons, or order oligos that can be sub-cloned into a plasmid backbone, or even submitted directly for plasmid synthesis through a variety of commercial companies.

If you were not able to complete the donor design, a fully annotated donor template sequence is available:

NUP153 donor insert