Building a pipeline for automated CRISPR construct design in *Drosophila*

Sebastian Kittelmann, Korneel Hens, Saad Arif

skittelmann@brookes.ac.uk, khens@brookes.ac.uk, sarif@brookes.ac.uk

Centre for Functional Genomics

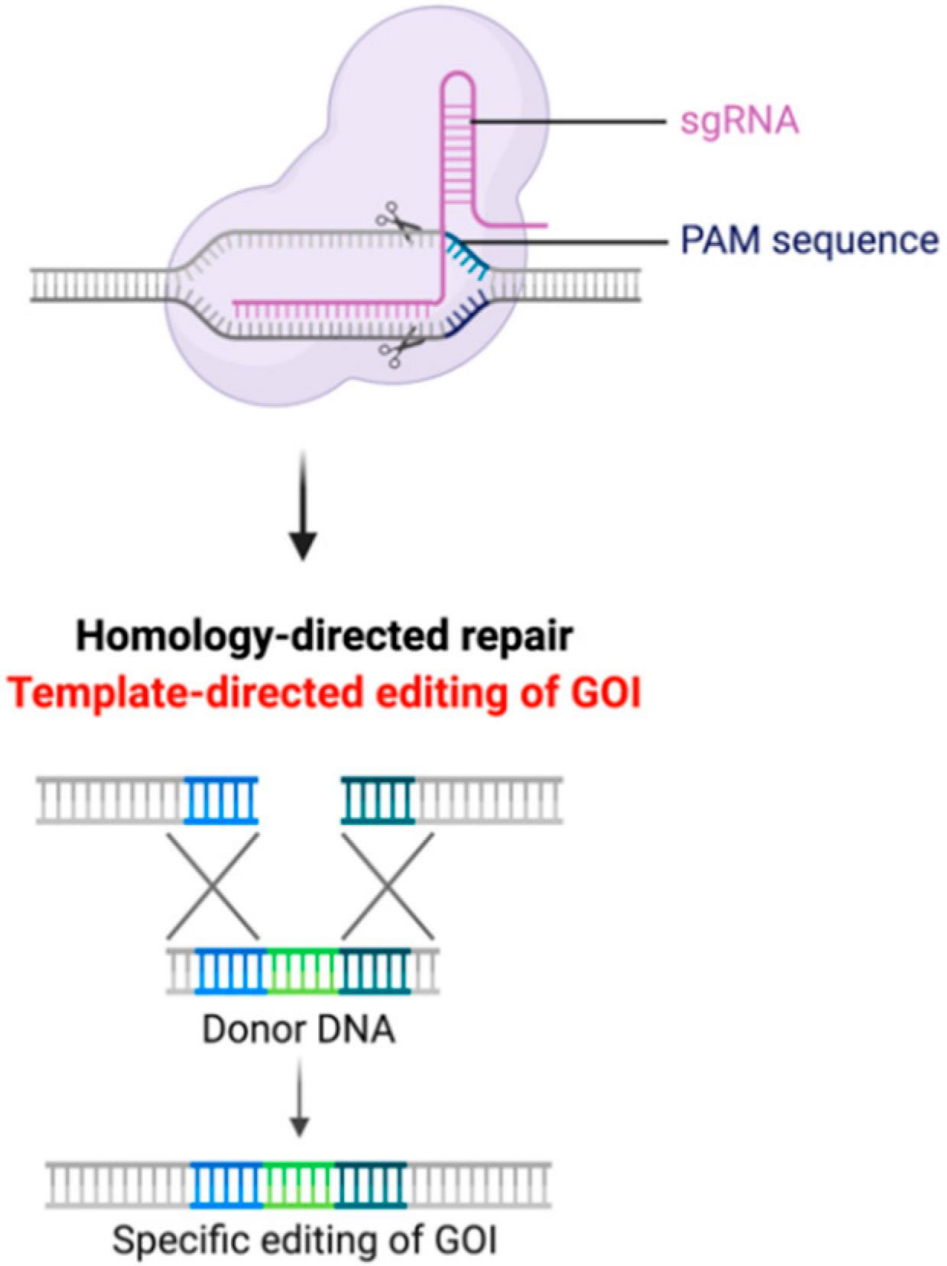
Oxford Brookes University

TFTag: a library of tagged transcription factors

- Make transgenic fly lines with tagged transcription factors (TFs)

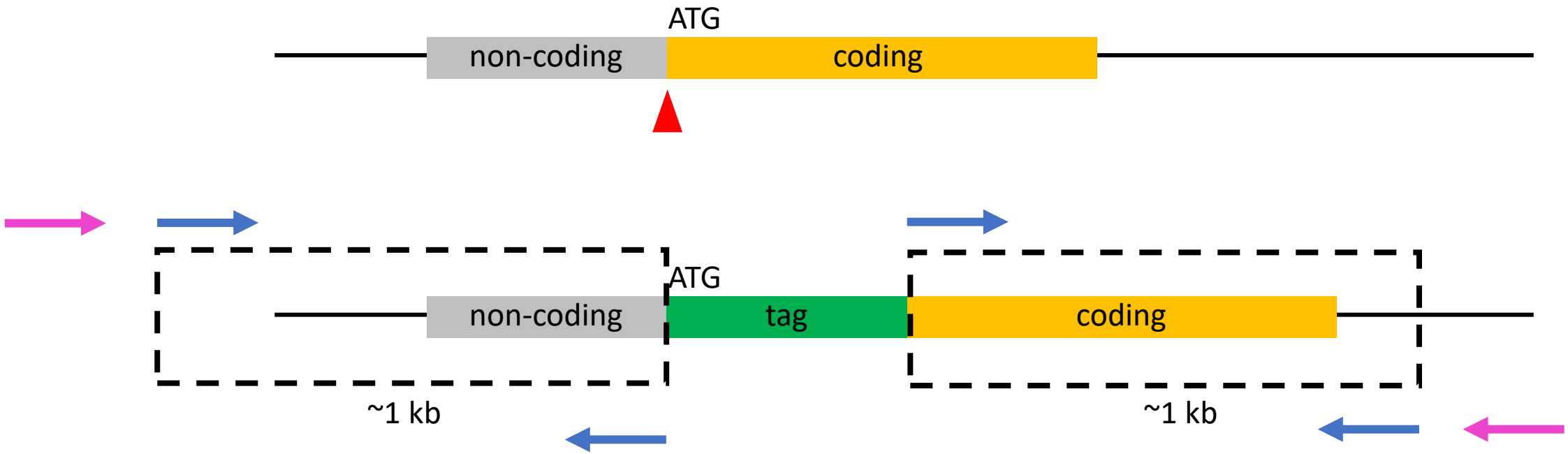
- Tag all TFs in the genome (~750) at all annotated termini (i.e., at every start and stop codon)

- Use CRISPR/Cas9 for tagging



- Cas9 protein: expressed in transgenic flies
- sgRNA: expression plasmid
- HDR template: plasmid

Reuven *et al.*, PMID: 33916763



- For each locus we need:
  - sgRNA sequence
  - Sequences of left and right homology arms (HAL & HAR)

    →One forward and one reverse primer for each
  - Primers to validate the insertion

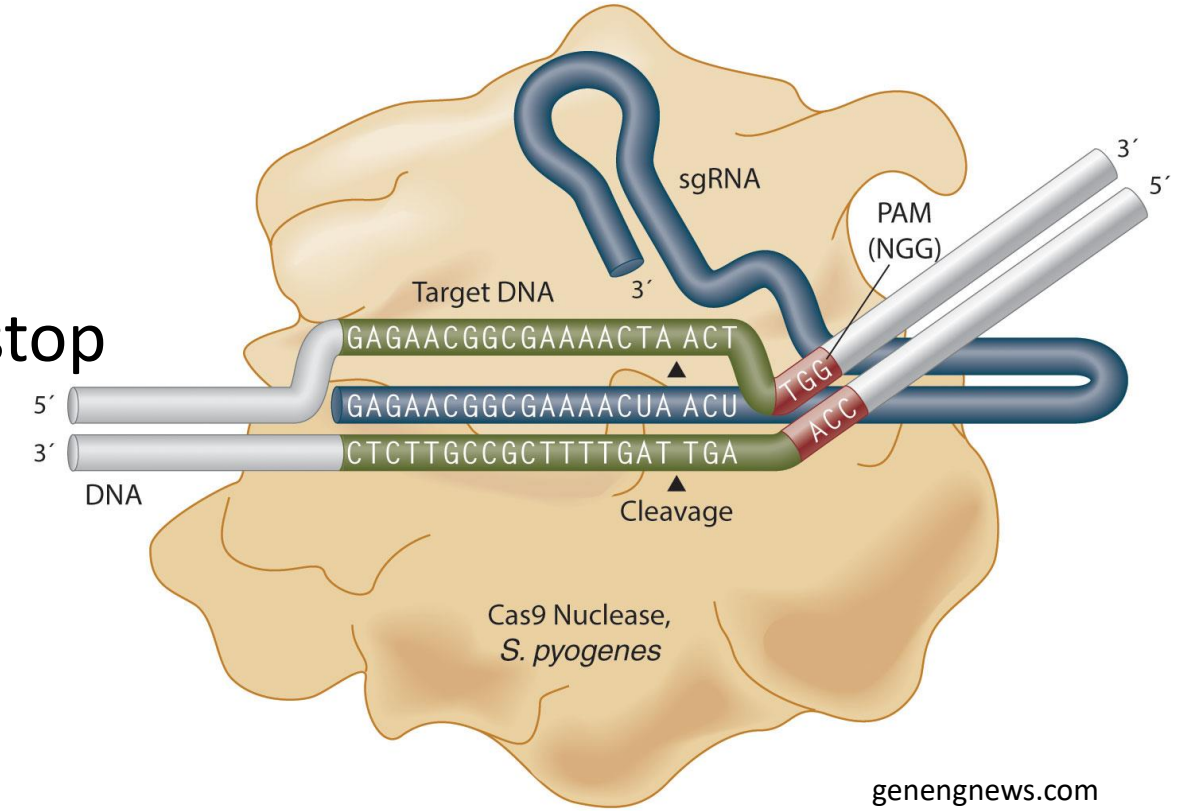
sgRNA

- Primary Adjacent Motif (PAM; NGG)

  should be as close to the start/stop as possible

- sgRNA: 20 bp upstream of PAM

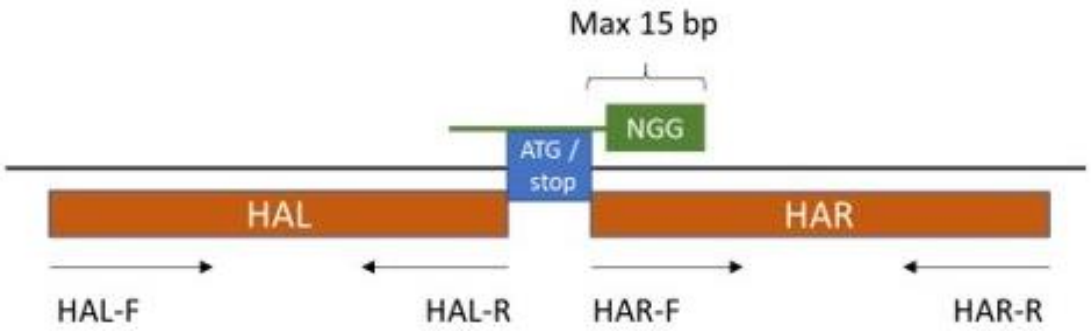
- Database with possible sgRNAs

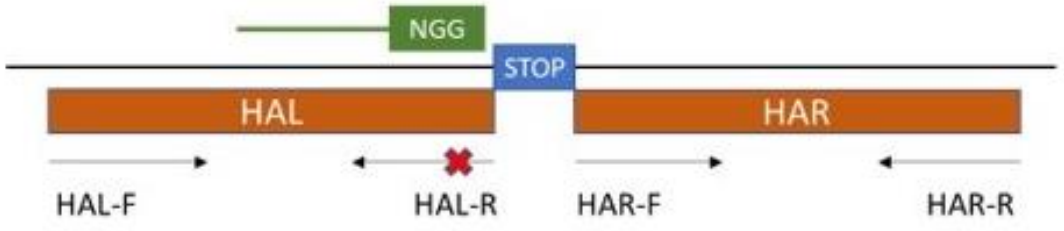


genengnews.com

sgRNA contd.

- Should span the start/stop codon so that max. 15 bp of the 3’ end (incl. PAM) are on one side
  - This should minimise the risk of the HDR template being cut by CRISPR
- Best case: G of ATG or TAG is part of PAM
- If this is not possible, the PAM must be mutated in the HDR plasmid (i.e., in the primer)
  - Ideally, a silent mutation in the CDS that mutates one of the Gs of the PAM





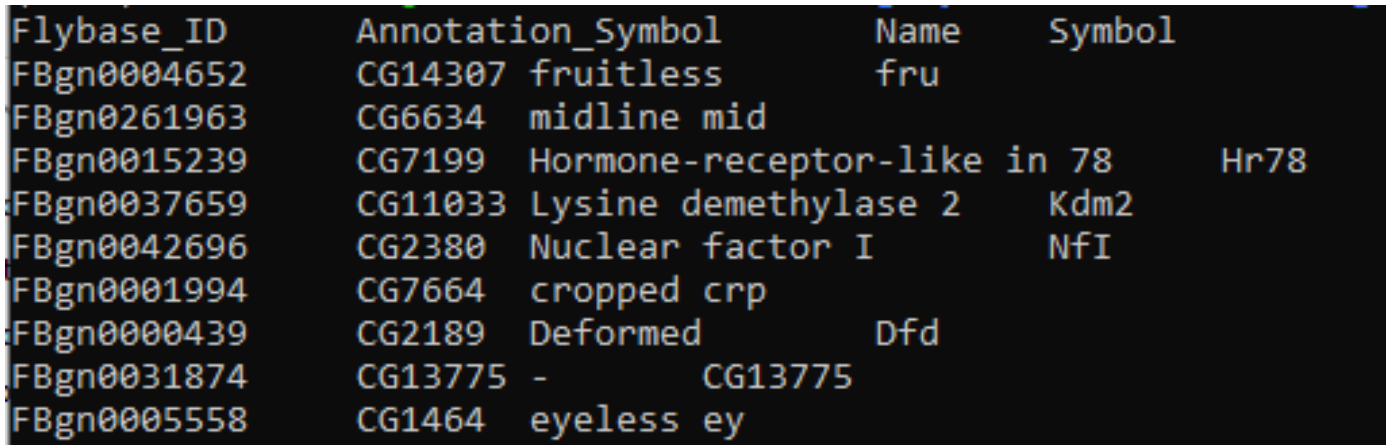
Desired result

Symbol	Flybase ID	Crispr cut site	Strand	tag location	isoforms tagged	Readthrough isoform	disrupts isoforms / other genes	sgR_F	sgR_R	sgRNA seq	HAL_F	HAL_R	HAL refseq	HAL len	gRNA site mutated relative to start of HAL	HAR_F	HAR_R	HAR refseq	HAR len	gRNA site mutated relative to start of HAR	val_F	val_R
oc	FBgn0004102	X:8632680..8632681	-	C	C,D,E,F,G	no	H	GTCGCAGGATAAGTACGCGAATA	AAACTATTCGCGTACTTATCCTG	GCAGGATAAGTACGCGAATA	CAACAACAATTCGGCAGCAG	CACCATgTTtGCGTACTTATCCTGCG	CAACA…	1012	g1003a, t1006c	AATCTACTCTTCCAATACTGCAGC	CAATCCTCGCATGCATTTACTC	AATCT…	1047	no		

- See Tagginglist.txt/Tagginglist.xlsx

Files

- List with TF names: TFs.txt/TFs.xlsx



Flybase_ID	Annotation_Symbol	Name	Symbol	
FBgn0004652	CG14307	fruitless	fru	
FBgn0261963	CG6634	midline	mid	
FBgn0015239	CG7199	Hormone-receptor-like in 78		Hr78
FBgn0037659	CG11033	Lysine demethylase 2	Kdm2	
FBgn0042696	CG2380	Nuclear factor I	NfI	
FBgn0001994	CG7664	cropped	crp	
FBgn0000439	CG2189	Deformed	Dfd	
FBgn0031874	CG13775	-	CG13775	
FBgn0005558	CG1464	eyeless	ey	

Column 1 (Flybase_ID) and column 4 (Symbol) are important and should show up in the results table

Files

- *Drosophila* genome in fasta format: dmel-all-chromosome-r6.48.fasta

- Genome annotation file: dmel-all-r6.48.gtf

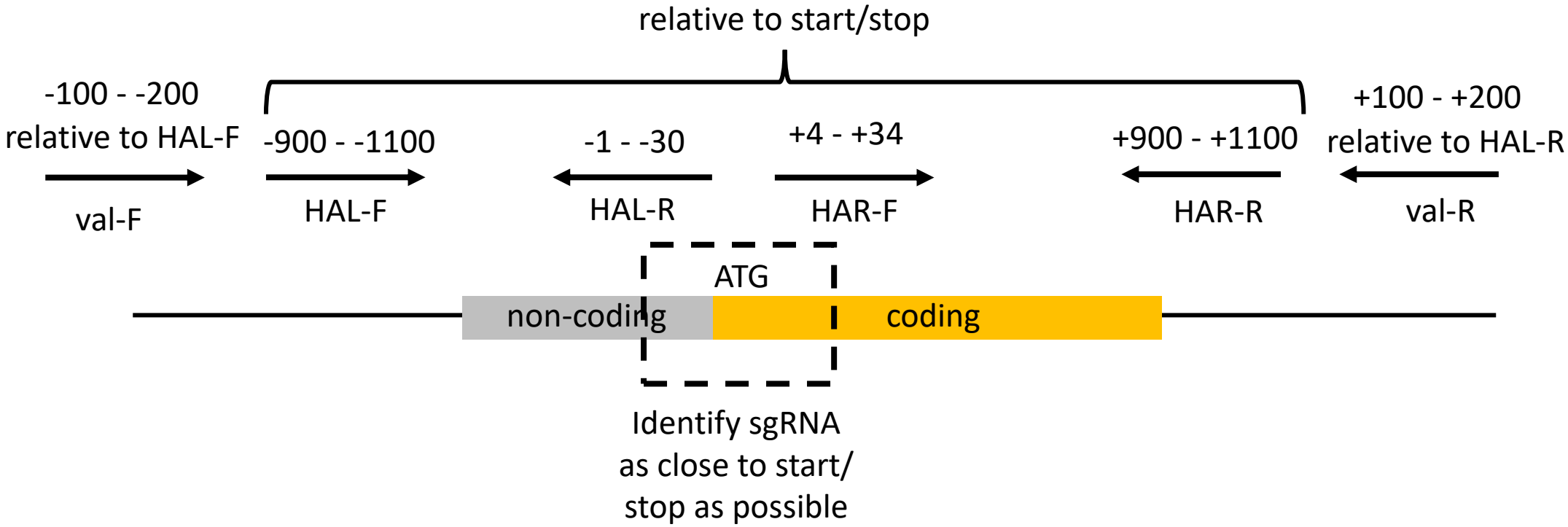
```
X       FlyBase gene    19961297        19969323        .       +       .       gene_id "FBgn0031081"; gene_symbol "Nep3";
X       FlyBase mRNA    19961689        19968479        .       +       .       gene_id "FBgn0031081"; gene_symbol "Nep3"; transcript_id "FBtr0070000"; transcript_symbol "Nep3-RA";
X       FlyBase 5UTR    19961689        19961845        .       +       .       gene_id "FBgn0031081"; gene_symbol "Nep3"; transcript_id "FBtr0070000"; transcript_symbol "Nep3-RA";
X       FlyBase exon    19961689        19961845        .       +       .       gene_id "FBgn0031081"; gene_symbol "Nep3"; transcript_id "FBtr0070000"; transcript_symbol "Nep3-RA";
X       FlyBase exon    19963955        19964071        .       +       .       gene_id "FBgn0031081"; gene_symbol "Nep3"; transcript_id "FBtr0070000"; transcript_symbol "Nep3-RA";
X       FlyBase exon    19964782        19964944        .       +       .       gene_id "FBgn0031081"; gene_symbol "Nep3"; transcript_id "FBtr0070000"; transcript_symbol "Nep3-RA";
X       FlyBase exon    19965006        19965126        .       +       .       gene_id "FBgn0031081"; gene_symbol "Nep3"; transcript_id "FBtr0070000"; transcript_symbol "Nep3-RA";
X       FlyBase exon    19965197        19965511        .       +       .       gene_id "FBgn0031081"; gene_symbol "Nep3"; transcript_id "FBtr0070000"; transcript_symbol "Nep3-RA";
X       FlyBase exon    19965577        19966071        .       +       .       gene_id "FBgn0031081"; gene_symbol "Nep3"; transcript_id "FBtr0070000"; transcript_symbol "Nep3-RA";
X       FlyBase exon    19966183        19967012        .       +       .       gene_id "FBgn0031081"; gene_symbol "Nep3"; transcript_id "FBtr0070000"; transcript_symbol "Nep3-RA";
```

```
X       FlyBase start_codon     19963955        19963957        .       +       0       gene_id "FBgn0031081"; gene_symbol "Nep3"; transcript_id "FBtr0070000"; transcript_symbol "Nep3-RA";
X       FlyBase start_codon     19963955        19963957        .       +       0       gene_id "FBgn0031081"; gene_symbol "Nep3"; transcript_id "FBtr0307554"; transcript_symbol "Nep3-RB";
X       FlyBase start_codon     19963955        19963957        .       +       0       gene_id "FBgn0031081"; gene_symbol "Nep3"; transcript_id "FBtr0307555"; transcript_symbol "Nep3-RC";
X       FlyBase start_codon     20094398        20094400        .       +       0       gene_id "FBgn0062565"; gene_symbol "Or19b"; transcript_id "FBtr0070003"; transcript_symbol "Or19b-RA";
X       FlyBase start_codon     20133990        20133992        .       +       0       gene_id "FBgn0031088"; gene_symbol "CG15322"; transcript_id "FBtr0301569"; transcript_symbol "CG15322-RB";
X       FlyBase start_codon     20133990        20133992        .       +       0       gene_id "FBgn0031088"; gene_symbol "CG15322"; transcript_id "FBtr0343166"; transcript_symbol "CG15322-RC";
X       FlyBase start_codon     20143186        20143188        .       -       0       gene_id "FBgn0041626"; gene_symbol "Or19a"; transcript_id "FBtr0070029"; transcript_symbol "Or19a-RB";
X       FlyBase start_codon     20110474        20110476        .       -       0       gene_id "FBgn0040784"; gene_symbol "karr"; transcript_id "FBtr0301572"; transcript_symbol "karr-RA";
X       FlyBase start_codon     20092383        20092385        .       -       0       gene_id "FBgn0031086"; gene_symbol "fd19B"; transcript_id "FBtr0070032"; transcript_symbol "fd19B-RA";
```

Design of gRNAs

- Annotations of sgRNAs: Hu.2019.8.28.sgRNA_designs.tar.gz
  - Unpacks into several .gff files with different stringencies and off-targets
  - Cycle through lists; if the most stringent one has no suitable sgRNA, check the next level down etc.

- Check if sgRNA can target the same site in the *Drosophila* strain we want to inject into
  - Sequenced genomes for “nos-Cas9 on 2” (will be used for injection if targeted TF is on chromosome 3) & “nos-Cas9 on 3” (will be used for injection if targeted TF is on chromosome X, 2, or 4)
  - BLAST sgRNA in these genomes to verify that there is no mutation that blocks it from cutting



- HAL-R and HAR-F must have their 5’ ends at -1 and +4, respectively
- Primers should be 18-25 bp
- Primer design with Pimer3 (https://primer3.org/)

Python function for Primer3

- You should be able to recycle this from someone else, e.g. https://pypi.org/project/primer3-py/
- HAL-R and HAR-F must start directly before and after the start/stop codon
- Manually, one would copy 30 bp up (HAL-R) or downstream (HAR-F) of the start or stop into primer3 and see if any of the primers start at position 1 (HAR-F) or position 30 (HAL-R).
- If not, start relaxing various parameters until you get one.
- If none come up, just take a primer that doesn't start at the correct position and manually elongate it.
- In other words, cycle through different versions with decreasing stringency until a primer is found.
- If the primer needs to be mutated, it should be longer to compensate for the mismatch (i.e., the mutation shouldn't be at the 3' end).

Questions?