SapTrap Builder Instructions

Introduction

The SapTrap Builder facilitates designing reagents for SapTrap assemblies, and for CRISPR editing in general. After entering a raw genomic DNA sequence of an intended edit region, the user is guided through each element of CRISPR edit site design, including specifying the tag insert site, selecting the guide RNA, and introducing point mutations into the repair template. The program outputs guide RNA and repair template sequences in formats for direct ordering or further *in silico* manipulation.

GUI Overview

Here is a brief overview of the GUI. Read further for detailed descriptions of each program element.
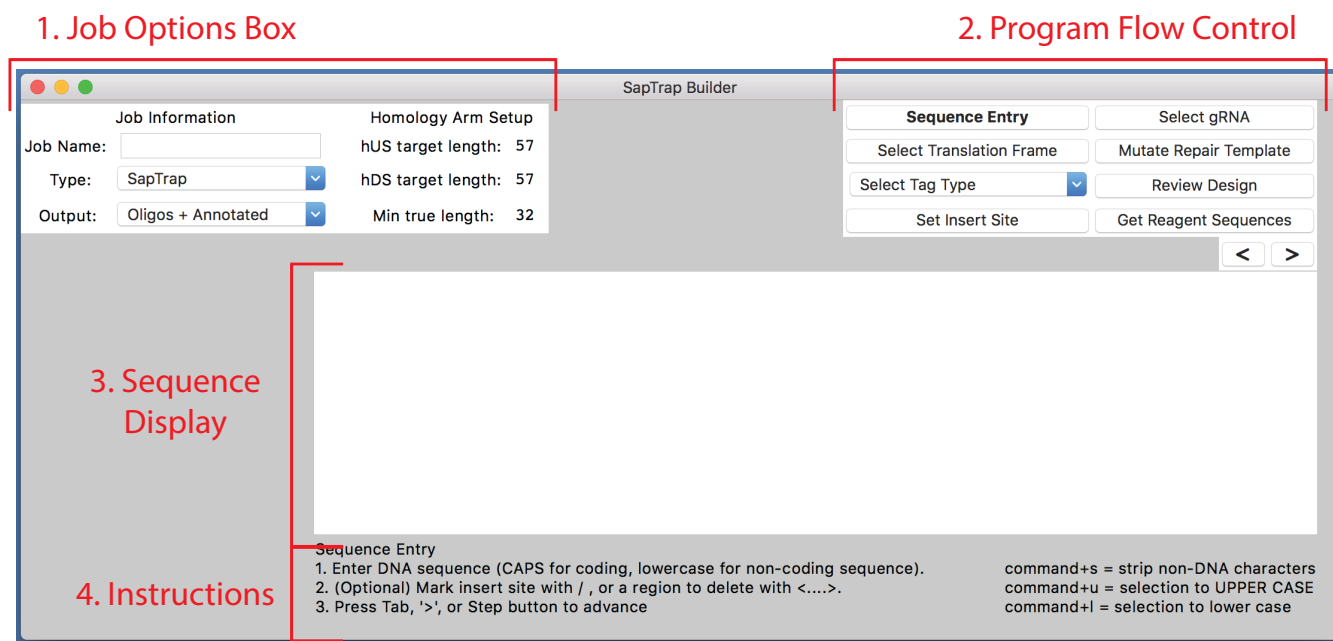
1. Job Options Box: allows the user to adjust parameters of the program output, such as homology arm length and output format.

2. Program Flow Control: allows the user to transition between the various steps of CRISPR reagent design.

3. Sequence Display: used to input the initial raw genomic DNA sequence, to display sequence information during each step of the design process, and to display reagent sequences.

4. Instructions: contains brief instructions relevant to the active step of reagent design.

Job Options

The job options box allows the user to adjust parameters of the program output, such as homology arm length and output format. By default, the options are set appropriately for annealed oligo based SapTrap assemblies. If using this paradigm, no adjustments are necessary.

Job Information		Homology Arm Setup	
Job Name:		hUS target length:	57
Type:	SapTrap	hDS target length:	57
Output:	Oligos + Annotated	Min true length:	32

The job options can be adjusted at any point during design; they do not need to be adjusted prior to commencing with reagent design to be able to affect program output.

Job Name: Text entered here will be prepended to final reagent sequence names. This field can be left blank if generic reagent sequence names are acceptable.

Type: By default, the SapTrap Builder generates reagent sequences for SapTrap assemblies. Sequences will have cohesive end and SapI recognition sequences appended as necessary. If this option is switched to 'Generic,' raw gRNA and repair template sequences without SapTrap elements will be generated. This allows the program to be of use to those opting for alternative DNA assembly strategies, or for alternative CRISPR schemes including coCRISPR.

Output: For Saptrap assemblies, site-specific DNA reagents can be either 1) oligos annealed with single-stranded 5' overhangs to produce cohesive ends for ligation or 2) double-stranded DNA flanked by SapI sites, so that SapI digests produces cohesive ends for ligation. The SapTrap Builder can output sequences corresponding to either form of DNA.

>*Oligo* outputs sequences to be ordered as single-stranded oligonucleotides, which are then annealed in pairs to produce double-stranded DNA to be fed into SapTrap reactions. The oligo output appends cohesive end sequences for SapTrap assemblies, but does not append SapI recognition sequences.

>*Synthetic DNA* outputs sequences that can be ordered as synthetic double-stranded DNA fragments. The synthetic output appends both cohesive end sequences and SapI recognition sequences to the output. The synthetic DNA output can also be used to design PCR primers for producing larger SapTrap reagents by PCR amplification from genomic DNA. Additionally, this output is useful for *in silico* production of vector files for plasmids made with any source of input DNA.

>The program can output these two basic forms in a variety of combinations:
>   1. Oligo + Annotated (*default*): Oligo sequences (5' – 3'), followed by an annotated synthetic DNA sequence.

2. Oligos: Only oligo sequence, 5' – 3'.
3. Synthetic DNA: Raw synthetic DNA sequence, with no annotation.
4. Annotated Synthetic DNA: Only an annotated synthetic DNA file is output.
5. Alignment: Oligo sequences are depicted as annealed, double-stranded DNA (as they will appear in a SapTrap reaction).

In 'Generic' mode, the program will only output raw gRNA binding site and homology arm sequences without any appended SapI cohesive end or recognition site sequences.

hUS target length: This is the target length, in bases, of the upstream (5') homology arm. Provided enough sequence is supplied to the program, the homology arm produced will be at least this long.

hDS target length: This is the target length, in bases, of the downstream (3') homology arm. Provided enough sequence is supplied to the program, the homology arm produced will be at least this long.

Min true length: This is the minimum length of the true homology arm portion of either hUS or hDS homology region. The 'true' homology arm is the portion of the homology arm that extends past the Cas9-induced cut in the genomic DNA. For example, if Cas9 is programmed to cut the genomic DNA 10 bases from the 5' end of the hUS, only the 5' 10 bases of the hUS serve as a homology arm for the repair machinery. In this case, the program will add additional bases to the hUS so that the number of bases extending past the cut site matches the value provided in the Min true length field.
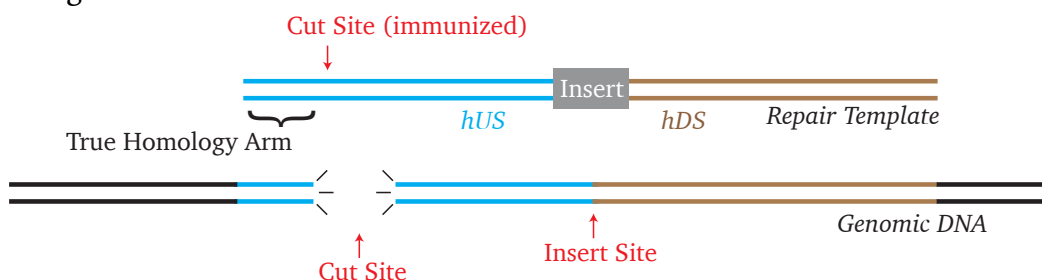
Program Operation

The program divides design into discrete steps. Users can freely navigate between the steps. During operation, the active step is bolded. The effect of keyboard input into the program and the behavior of the sequence display box changes depending on which step of the program is active.

Sequence Entry	Select gRNA
Select Translation Frame	Mutate Repair Template
Select Tag Type ⌄	Review Design
Set Insert Site	Get Reagent Sequences
	< >
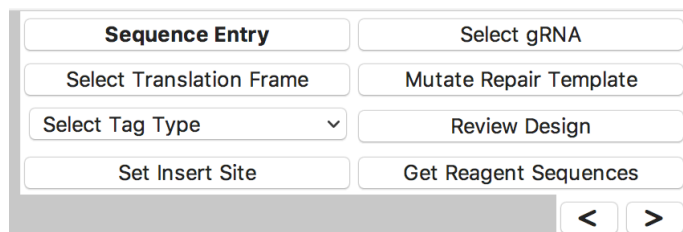
Users navigate among program steps by directly clicking on the labeled step button, by clicking on the '<' and '>' arrow keys to move through the steps in default order, or by pressing 'Tab' and 'Shift-Tab' to move forward and backward through the default order. With the exception of the 'Sequence Entry' step, all other steps can be freely visited and revisited in any desired order. A valid genomic DNA sequence must be entered into the program before visiting any downstream steps, and returning to the sequence entry step resets the program and deletes the working sequence.

The sequences that are displayed in the sequence display box are added to the display on an 'as-needed' basis. Sequence lines are labeled in the grey space to the left of the sequence display box.

There are no 'committal' actions required to lock in a change. The sequence as displayed is stored into memory, and proceeding to the next step will 'lock in' any changes made during the previous step.

Below are detailed descriptions of the functionality of each program step.

Sequence Entry

The program opens in the Sequence Entry step. In this step, the sequence display box accepts sequences pasted (or typed) by the user. Any character can be input into the entry box, but the program will not allow you to proceed further if invalid characters are present. The program requires that sequences be formatted with non-coding DNA expressed as lowercase letters (a, c, g, t) and coding DNA expressed as uppercase letters (A, C, G, T). Other valid characters in the sequence entry are:

    returns (enters) – these are ignored by the program
    '/'   : a single backslash can be placed in the sequence to mark the desired tag insert location
    '<' and '>'   : If you are creating a large deletion, you can mark the intended deletion region with <…..>.

To facilitate formatting DNA sequences for input into the program, a few simple processing features are available:

command + s: strips non-DNA characters from the sequence in the display box.
additionally, if bases are highlighted, will format the sequence with line-lengths equivalent to the length of the highlighted region

command + u: converts highlighted bases to uppercase

command + l: converts highlighted bases to lowercase

Once the sequence in the display box is correctly formatted, you can proceed to the next step of the program by pressing Tab, the '>' button, or by pressing the button of the next desired step. Note you cannot proceed directly to the 'Review Design' or 'Get Reagent Sequences' steps as these steps require the tag type and guide RNA to be selected.

Select Translation Frame

Available controls: Left and Right arrow keys.
Function: Shift translation frame -1 or +1.

The program makes an initial guess at the translation frame, and displays the protein translation above uppercase DNA sequence. Single letter amino acid designators are placed above the first nucleotide in each codon. The program assumes that stretches of lowercase DNA between coding regions are introns, and translates the spliced product. If the translation frame is incorrect, correct it with the appropriate arrow key press.

Select Tag Type

Available controls: Up and Down arrow keys, or mouse based direct selection.
Function: Select the SapTrap tagging type. Selection is irrelevant if using 'Generic' mode.

The cohesive ends appended to each homology arm are different depending on the type of tagging (see Appendix II). The options are:

N-Tag: For adding a tag at the ATG start codon of a gene.

C-tag: For adding a tag immediately prior to the STOP codon of a gene.

Internal – with linkers: For adding a tag in the middle of a gene coding sequence, and appending SapTrap linkers to each side of the tag.

Internal, no linkers: For adding a tag in the middle of a gene coding sequence, and splicing the tag sequence directly to the gene coding sequence without additional linkers.

Native Intron: For point mutations, for dropping unc-119+ rescue cassette into a native intron.

Syntron: For point mutations, for creating a new unc-119+ syntron within a coding reagion.

If using 'Generic' mode, a selection must be made but it will not affect the output.

Set Insert Site

Available controls: Left and Right arrow keys (± Shift), 'r', 'c'
Function: Indicate location of the insert sequence (i.e. GFP, unc-119+ cassette, etc)

During this step, the text |-INSERT-| will appear to mark the location of the desired insertion. If a '/' was used during sequence entry to indicate the insert location, the insert will be located at the position of the '/'. Otherwise, the insert is by default positioned at the midpoint of the sequence.

To move the insert site, press the arrow keys to move in single base steps, or press Shift + arrow to move in 9 base steps. Pressing 'r' will position the insert back at the original position. If a guide RNA has already been selected, pressing 'c' will move the insert site to the cut site of the guide RNA.

The text 'INSERT' is in uppercase when the insert is positioned in frame with the coding sequence, and is lowercase (|-insert-|) when positioned out of frame with the coding sequence.

Select gRNA

Available controls: Left and Right arrow keys.
Function: Chose a single guide RNA from candidates within the sequence.

```
    Wild Type Translation:
          Wild Type DNA: agag|-INSERT-|tttcccctcatcccccagccatctcctttct
   Guide RNA binding site:      |-INSERT-|    cccctcatcccccagccatctcc
                    PAM:                      ccc  /\              (9)
```

The program automatically identifies all available gRNA binding sites in the sequence and sorts them based on the distance between the cut site for the individual guide and the insert site specified for the genomic region. At any given time, only a single guide RNA candidate is displayed. The initial guide candidate is the available guide with the shortest distance between its cut site and the insert site. To select a different guide, simply use the arrow keys to select the next closest guide candidate in the indicated direction.

Two strings are displayed for a guide RNA candidate: the binding site string shows the binding site and PAM sequence aligned to the genomic DNA sequence. The PAM string shows only the PAM sequence. Additionally, the PAM string contains an arrowhead '/\' marker to mark the

location of the cut site, and shows the distance between the cut site and the marked insert site in parentheses under the binding site.

```
Wild Type Translation: F  A  P  F  S  F  E  K  I  L  A  E  E  R  E  A  E  E  N  L  *
        Wild Type DNA: GTTCGCTCCATTCTCATTCGAAAAAATCCTTGCTGAAGAGCGTGAAGCTGAGGAGAATCTCTAA|-INSERT-|gatcacctcggccacttcaaacagtgtgacatcgacgttcgaca
Guide RNA binding site:                                TGCTGAAGAGCGTGAAGCTGAGG        |-INSERT-|
                  PAM:                                (17)           /\  AGG
                                        gRNA sequence contains SapI site. Sequence incompatible with SapTrap assembly.
```

The program will display flags in red text beneath the guide RNA binding site if the guide sequence contains problematic sequences. Potentially problematic sequences that are flagged include:

> Presence of a SapI site: guides containing SapI sites cannot be used in SapTrap assemblies. The DNA will be cleaved.
> Presence of a polIII terminator: guides with polIII terminators cannot be transcribed by polIII.
> Presence of a C base upstream of the PAM: We have found that guides with a 'C' base immediately preceding the 'NGG' PAM for Cas9 exhibit extremely low activity, and we do not use them.

The program will not prevent you from using flagged sequences; flags merely serve as warnings.

Mutate Repair Template

Available controls and functions:

> **Left and right arrow keys**: move the active base (cursor) left and right by a single base.
> **Shift + left and right keys:** move cursor by 9 base increments.
> **Up and down arrow keys:** in coding regions, cycle through synonymous codons for the codon including the selected base.
> **Shift + up and down keys:** in coding regions, cycle through non-synonymous codons for the codon including the selected base.
> **a, c, g, t:** mutate the active base to the entered base.
> **'-' :** delete the active base.
> **A, C, D, …:** Mutate the amino acid coded for by the active base to the entered amino acid.
> **r:** reset any changes from the current session (all changes since entering the mutator)
> **z:** reset all mutations back to wild-type.

Purpose: The purpose of the mutator is introduce needed and desired mutations into the repair template sequence. Needed mutations include mutations within the guide RNA binding site that serve to immunize the repair template against Cas9 mediated cleavage with the chosen guide. Additionally, the repair template cannot contain a SapI site if assembled by SapTrap. SapI sites must be removed by mutation. Desired mutations include point mutations of research interest. For general tagging, you will want to introduce silent rather than coding changes. The up and down arrow keys cycle through silent mutations on a per amino acid basis. Synonymous codons are not cycled continuously: the program arranges the codons based on usage in *C. elegans*. The most used codon is the 'highest' codon in the cycle (accessed by pressing 'up'), and the least used codon is the 'lowest' codon in the cycle (accessed by pressing 'down'). Thus, to access all possible codons for a given amino acid, you may need to press both the up and down keys.

```
      Wild Type Translation:  L  A  E  E  R  E  A  E  E  N  L              *
                                                   *
 Repair Template Translation:  L  A  E  E  R  E  T  E  E  N  L              *
                                                   |
     Repair Template DNA: TCCTTGCTGAAGAACGTGAAACAGAGGAGAATCTC|-INSERT-|TAAgatcacctcggccacttca
                                        *        * *
           Wild Type DNA: TCCTTGCTGAAGAGCGTGAAGCTGAGGAGAATCTC|-INSERT-|TAAgatcacctcggccacttca
       Guide RNA binding site:                      GAGAATCTC|-INSERT-|TAAgatcacctcgg
                      PAM:                           (8)                /\  cgg
                SapI Sites:
```

By pressing Shift + Up or Shift + Down, you can cycle through non-synonymous codons in alphabetical order. This list will cycle continuously. Alternatively, you can change a codon to that of another amino acid by pressing the uppercase single letter code for the amino acid. Whenever a coding change is enacted, the codon applied is the codon with highest usage in *C. elegans*. After applying a codon change, you can cycle to other codons for the changed amino acid using the up and down arrow keys without the shift button depressed.

The mutator displays alignments of the Wild Type DNA with the repair template DNA, and of the Wild Type translation with the repair template translation. Differences are noted with an asterisks (*). An additional string titled SapI sites highlights any SapI sites present in the repair template sequence, as these need to be removed for SapTrap assembly.

Review Design

The design reviewer will display an alignment of all sequences pertinent to the design. The repair template DNA sequence will be trimmed to show only the DNA that will be incorporated in the final repair template. At this stage, the program checks for numerous common design errors, and alerts the user to any found errors, as well as to any potentially unexpected design modifications. Errors include not properly immunizing the repair template against the selected

guide RNA. Unexpected modifications include extension of a homology arm beyond the target length so as to satisfy the minimum true homology arm length specification. These notifications will not prevent you from proceeding and obtaining reagent sequences, they are simply warnings and notifications so that you can correct the design if desired. See the appendix for a full description of all warning flags.

Get Reagent Sequences

Once the design is acceptable, pressing 'Get Reagent Sequences' will display the designed reagents in the selected format. See above for descriptions of the available formats. Note that this step does not represent a final state of the program; you can still go back to other steps and modify the design without starting over.

During all other phases of the design, the sequence display box is locked from user input / selection. During this step, the box is unlocked so that you can type additional text into the sequence display, and select and copy sequences from the display box. Oligo sequences can be copied and pasted directly into excel or a text editing program. Annotated sequences can be copied and pasted into an Ape window to view the annotated file.

Specific Cases
Here, special considerations for specific types of edits are considered.

Tagging

The program is designed with tagging in mind. Following the program steps in the default order will provide the most logical progression through the process of designing a protein tag insertion. During sequence entry, include a '/' character in the sequence at the location of the desired tag insert to bypass manually selecting the translation frame and manually positioning in the insert in the sequence.

SapTrap tags start with an ATG and do not contain a stop codon. The program will detect the native ATG and remove it for N-terminal tags, as this prevents creating a double ATG (native + SapTrap tag). For C-terminal tagging, position the Insert just before the native stop codon, so that the native stop codon is included in the hDS homology arm immediately downstream of the tag.

Deletions

To make large-scale deletions (for example, delete an entire coding sequence), use the <...> syntax during the sequence entry step. Leave the genomic DNA sequence between the <...>
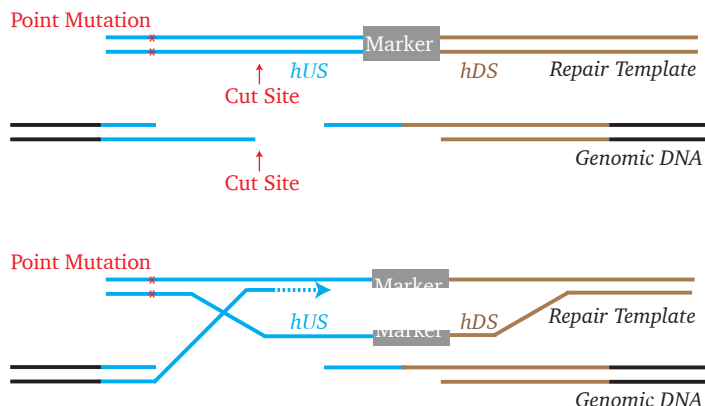
brackets. Upon entering a sequence with the <…> bracket notation, the sequence display box will show the complete Wild Type sequence aligned to the repair template DNA sequence, in which the deleted bases are represented with -'s.

```
Wild Type Translation:                      V   E   F   Q   S   K   F   Y   G   G   L   G   Y   E   F
   Repair Template DNA: attcaataattttcag-------------------------------------------------
         Wild Type DNA: attcaataattttcagGGTCGAGTTCCAATCAAAGTTCTATGGAGGGCTTGGATATGAGTT(
```

For large-scale deletions, the default insert site is near the 5' edge of the deletion. If you select a guide RNA that is partially or entirely within the deletion region, no further mutation is required to immunize the repair template against the gRNA.
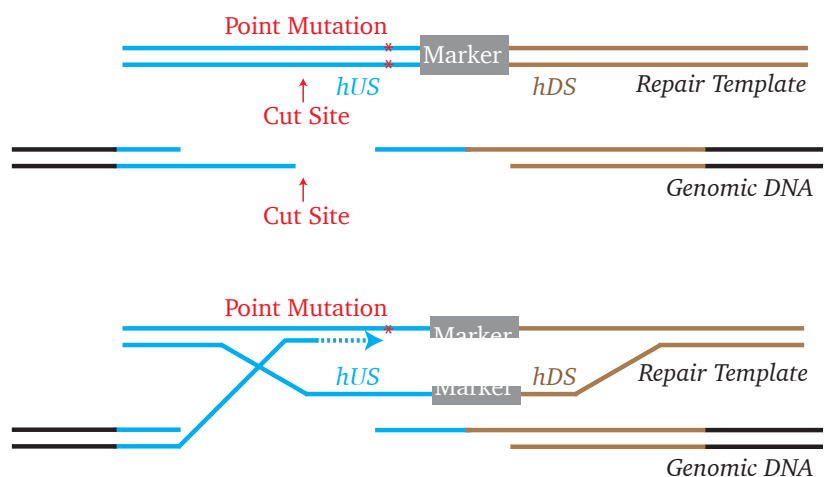
Point Mutations

Inserting tags and deleting sequences requires little consideration of the mechanism of homology dependent repair; one must simply flank the tag with homology arms of the target

and the tag will be inserted. Introducing point mutations by CRISPR requires further mechanistic consideration. Homology directed repair of a double stranded break proceeds as follow: First, the 5' ends of the genomic DNA are resected, leaving single stranded 3' ends. The single stranded 3' ends can invade duplex DNA of the repair template, where they anneal to matching sequences and serve as substrates for new DNA synthesis that proceeds across the break. The intertwined DNAs are resolved by a variety of complex molecular pathways. The important point for CRISPR design is that the location within the homology arm where strand invasion and binding occurs is stochastic; regions of the arm distal to the cut are less frequently incorporated during repair. If putative inserts are identified by incorporation of a selectable marker, and the Cas9 cut site sits between the selectable marker and the desired point mutation, repair events that incorporate the selectable marker but not the desired point mutation can occur.

To ensure that a point mutation is incorporated during repair, the point mutation should be positioned between the cut site and the selectable marker. With this arrangement, events identified by the incorporation of the selectable marker will generally also incorporate the desired point mutation.

We have recently created a new set of SapTrap vectors to facilitate generating point mutations with *unc-119+* selection. These vectors contain a reverse-oriented, floxed *Cbr-unc-119* selectable marker embedded in a partial syntron. The 5' and 3' splice site sequences of the syntron have been removed. When the marker (Insert site) is positioned in a native intron and 'Native Intron' tagging type is selected, the program will generate homology arms that ligate the native intron sequences directly to the splice site devoid unc-119+ cassette. When the marker is positioned in a coding region and 'Syntron' tagging type is selected, the program will generate homology arms that add 5' and 3' sequences to the partial syntron, allowing the syntron to function putatively within any coding sequence. If a desired point mutation lies near a native intron, the unc-119 marker can be added to the native intron. If a desired point mutation is located deep within an exon, the unc-119 marker can be added within a new syntron near to the mutation location. In both cases, the unc-119 marker can be removed by Cre expression following isolation of the mutant strain.

```
        Wild Type Translation:                         V   E   F   Q   S   K   F   Y   G   G   L   G
                                                                            *
    Repair Template Translation:                       V   E   F   Q   S   A   F   Y   G   G   L   G
                                                                            |
        Repair Template DNA: tcaataatttt|-insert-|cagGGTCGAGTTCCAATCAGCTTTCTATGGAGGGCTTGGA
                                                                          ***
            Wild Type DNA: tcaataatttt|-insert-|cagGGTCGAGTTCCAATCAAAGTTCTATGGAGGGCTTGGA
        Guide RNA binding site:               |-insert-|     GAGTTCCAATCAAAGTTCTATGG
                    PAM:                                      (24)               /\   TGG
            SapI Sites:
```
Example of 'Native Intron' unc-119 cassette positioning for K>A mutation.

```
    Wild Type Translation: E   F   Q   S   K   F   Y   G   G   L   G   Y   E               F   A   P
                                                                           *
Repair Template Translation: E   F   Q   S   K   F   Y   G   G   L   G   A   E               F   A   P
                                                                   *  *      **
            Wild Type DNA: GAGTTCCAATCAAAGTTCTATGGAGGGCTTGGATATGA|-insert-|GTTCGCTCCA
        Guide RNA binding site:         TCAAAGTTCTATGGAGGGCTTGG         |-insert-|
                    PAM:                (12)                /\   TGG
            SapI Sites:
```
Example of 'Syntron' unc-119 cassette position for Y>A mutation. Additional silent mutations are included to disrupt guide RNA binding to repair template. Note that insert does not need to be in frame, as the insert will be a complete, functional synthetic intron.

Appendix I – Program Flags

Guide RNA selection:

gRNA binding site ends in C base. Such guide exhibit low activity in C. elegans.

> Guides containing a C immediately prior to the NGG Cas9 PAM are nearly inactive in our experience

gRNA sequence contains SapI site. Sequence incompatible with SapTrap assembly.

> Guides cannot contain a SapI site. The guide RNA sequence must match the genome exactly, so there is not possibility of mutating the SapI site for assembly, as is the case for SapI sites elsewhere in the repair template. SapI sites are only incompatible with SapTrap assembly and do not reduce activity if the gRNA construct is produced by other means.

gRNA sequence contains polIII terminator. Sequence incompatible with production in vivo.

> gRNAs are expressed *in vivo* with a polIII promoter. If the binding site contains a polIII terminator (6 t's in a row), the scaffold portion of the guide will not be transcribed, and functional gRNA or sgRNA will not be produced.

Review Design:

gRNA binding site intact. Repair template can be cut by Cas9 + guideRNA.

> The repair template contains a completely intact gRNA + Cas9 binding site, and will be subject to cleavage, both before and after incorporation into the genome. If you disrupt the PAM by deletion, the program will check of the adjacent bases form a new PAM, rendering the apparently immunized repair template subject to cleavage.

gRNA binding weakly disrupted. Consider mutations nearer to PAM.

> Cas9 is most sensitive to mismatches between the guide RNA and the target DNA within the first 8 bases nearest the PAM. Mutations outside this region are better tolerated, and might render the repair template partially sensitive to cleavage.

X homology arm extended beyond target length because cut site within minimum distance of edge.

> The homology arm needed to be extended beyond the target length specification in order to satisfy the minimal true length requirement. This does not impair function, but may affect pricing of oligo reagents.

X homology arm extended beyond target length minimum arm length > target length.

> If you set the minimum true length to be greater than the target homology arm length, the program will extend the arms to meet the minimum true length requirement. This can only happen if you

Insufficient sequence provided for desired length X homology arm.

> The original genomic DNA sequence entered into the program is too short to calculate a homology arm satisfying the target length or minimum true length requirement. You need to input more genomic DNA sequence and run the design again.

SapI site in repair template. Incompatible with SapTrap assembly.

There is at least one SapI recognition sequence in the repair template. If you try to assemble using SapTrap, this site will be cut and correct assembly is unlikely. You should introduce a silent mutation in the SapI recognition sequence to obliterate the site.

For point mutagenesis only: Mutations marked by ~ have a lower probability of incorporation in genome. Position critical mutations between the cutsite and the insert site.

We have found that mutations located between the cut site and the unc-119+ insert site are almost always incorporated in the genome during edits, but mutations outside of this region are incorporated at a lower frequency. Thus, it is a good design practice to select a gRNA cut site and unc-119+ insert site that flank any critical mutations, such as desired non-synonymous coding changes.

This flag is only triggered when 'Native Intron' or 'Syntron' tagging types are selected, or when a non-synonymous change is detected between the repair template and the original genomic DNA. If you are specifically trying to add silent mutations and not using the typical point mutation tagging strategy, this warning will not show up even if your design violates the rule.
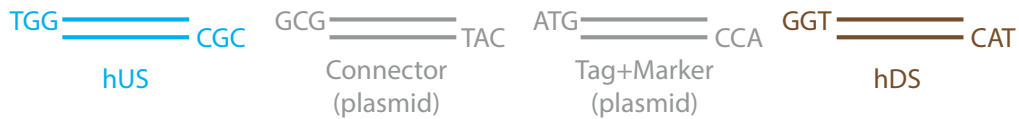
Appendix II – SapTrap Overhangs
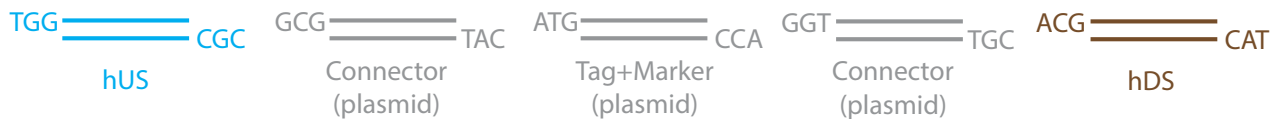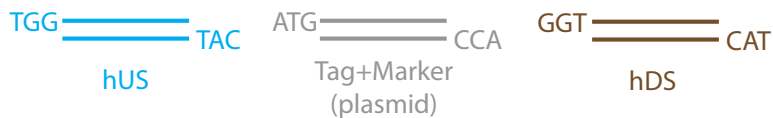
gRNA sequence (all tagging modalities)

TTG — CAA
gRNA seed

N-Tag

TGG — TAC
hUS

ATG — CCA
Tag+Marker
(plasmid)

GGT — TGC
Connector
(plasmid)

ACG — CAT
hDS

C-Tag

TGG — CGC
hUS

GCG — TAC
Connector
(plasmid)

ATG — CCA
Tag+Marker
(plasmid)

GGT — CAT
hDS

Internal - with linkers

TGG — CGC
hUS

GCG — TAC
Connector
(plasmid)

ATG — CCA
Tag+Marker
(plasmid)

GGT — TGC
Connector
(plasmid)

ACG — CAT
hDS

Internal - no linkers

TGG — TAC
hUS

ATG — CCA
Tag+Marker
(plasmid)

GGT — CAT
hDS

Native Intron

TGG — tca
hUS

agt — aaa
Syntron+Marker
(plasmid)

ttt — CAT
hDS

Syntron

TGG — gta
cattca
hUS

agt — aaa
Syntron+Marker
(plasmid)

tttcag
gtc — CAT
hDS

Depiction of the cohesive ends added to each fragment (hUS, hDS, sgRNA) designed by the program for various tagging modalities. In grey are depictions of connector and tag+Marker cassettes that are donated to SapTrap assemblies by plasmids from the SapTrap kit.