Application Note: Generating GFP-Tagged Human CD81 Tetraspanin Protein Using SBI's PrecisionX SmartNuclease System And HR Tagging Vectors

Table of Contents:

I. Background

The recent discovery of the CRISPR/Cas9 complex has provided researchers an invaluable tool to target and modify any genomic sequence with high levels of efficacy and specificity. The system, consisting of a nuclease (Cas9) and a DNA-directed guide RNA (gRNA) allows for sequence-specific cleavage of target sequence containing a protospacer adaptor motif "NGG" (Fig. 1). By changing the gRNA target sequence, virtually any gene sequence upstream of a PAM motif can be targeted by the CRISPR/Cas9 system, enabling the possibility of systematic targeting of sequences on a genomic scale.
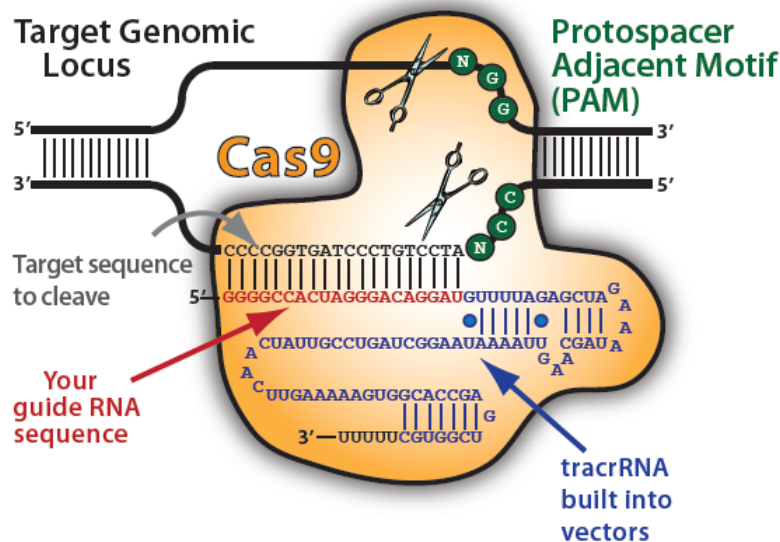

Fig. 1. Illustration of the CRISPR/Cas9 Heterocomplex

This application note is designed for first-time and experienced users of the CRISPR/Cas9 system to learn how to create an endogenous GFP-tagged version of human CD81 tetraspanin protein using a combination of SBI's PrecisionX Cas9 SmartNuclease and Homologous Recombination (HR) Tagging vectors. The CD81 tetraspanin protein has been found to be preferentially localized to extracellular microvesicles (EMVs) such as exosomes, and may be utilized as a marker for tracking the location of exosomes for *in vitro* and *in vivo* applications.

The protocol is designed for use of the CAS940A-1 [CMV-hspCas9-H1-gRNA] cloning vector and HR120PA-1 [EcoRI-copGFP-loxP-EF1a-RFP-T2A-Puro-LoxP-MCS] HR Tagging vector to generate a C-terminal fusion of CD81 protein to GFP. Other combinations of Cas9 SmartNuclease Vectors and HR tagging vectors may be substituted as well.

II. Analysis of Gene Target

Human CD81 Tetraspanin Genomic Sequence covering Exon 6,7,8 and 3'UTR region

```
agCTTGACTG CTGTGGCTCC AGCACACTGA CTGCTTTGAC CACCTCAGTG   2417146
CTCAAGAACA ATTTGTGTCC CTCGGGCAGC AACATCATCA GCAACCTCTT   2417196
CAAGgtgcgc gaggccggtg gggccgcgcc tgaccccccg catgtcccgc   2417246
ccctgggtgg ggtcctaggg gtgggcaggt cacacggcag ccccacaggg   2417296
agcgaccaca ctgggtggca tggcccctgt cagggctgct ctgctgggag   2417346
ggttgggggtg ggaccgcatc tggcccacga ggaaggcagg cgccctgtgc   2417396
tgcgcattcc gggtgaagaa ggtggaggct ctgggggggtg ggaactcacc   2417446
tgcacccccca gctccacgtg tgcactcgtg ggtgtggacg cccctgacag   2417496
cctgtagctg gcagggcctg caggccatat agtgccctgt ggaagtttcc   2417546
tgctgaggcc tcagtggaag tcgtcatcag tgatgcttta ggggtctagt   2417596
gacaccaatg accgtgatct cagtggaaaa gggcacagtg tgtcccaggc   2417646
atttcgcgtt tatgttaaaa cgggtggaag atagcaagcc ggcagaggcc   2417696
gggccgctgc acccgcctgt tccgaggtgg gtagggggtg gggggctgtt   2417746
cccaggattc ccctctacgc tttctgtggt gaccacggat tactgcgtga   2417796
caacgggaag ccgggagccg aggcccggtc cctgaccacg cgtgcctggc   2417846
caccccctgca gGAGGACTGC CACCAGAAGA TCGATGACCT CTTCTCCGGG   2417896
AAGCTGTACC TCATCGGCAT TGCTGCCATC GTGGTCGCTG TGATCATGgt   2417946
gagcgggcgg gggcggaggg cctgctctct gggctgcccc ttccgcgggg   2417996
ccttgtgctg actgcgcccc ccaccaccct cctgcagATC TTCGAGATGA   2418046
TCCTGAGCAT GGTGCTGTGC TGTGGCATCC GGAACAGCTC CGTGTACTGA   2418096
GGCCCCGCAG CTCTGGCCAC AGGGACCTCT GCAGTGCCCC CTAAGTGACC   2418146
CGGACACTTC CGAGGGGGCC ATCACCGCCT GTGTATATAA CGTTTCCGGT   2418196
ATTACTCTGC TACACGTAGC CTTTTTACTT TTGGGGTTTT GTTTTTGTTC   2418246
TGAACTTTCC TGTTACCTTT TCAGGGCTGA CGTCACATGT AGGTGGCGTG   2418296
TATGAGTGGA GACGGGCCTG GGTCTTGGGG ACTGGAGGGC AGGGGTCCTT   2418346
CTGCCCTGGG GTCCCAGGGT GCTCTGCCTG CTCAGCCAGG CCTCTCCTGG   2418396
GAGCCACTCG CCCAGAGACT CAGCTTGGCC AACTTGGGGG GCTGTGTCCA   2418446
CCCAGCCCGC CCGTCCTGTG GGCTGCACAG CTCACCTTGT TCCCTCCTGC   2418496
CCCGGTTCGA GAGCCGAGTC TGTGGGCACT CTCTGCCTTC ATGCACCTGT   2418546
CCTTTCTAAC ACGTCGCCTT CAACTGTAAT CACAACATCC TGACTCCGTC   2418596
ATTTAATAAA GAAGGAACAT CAGGCATGCT ACCAGGCCTG TGCAGTCCCT   2418646
CAGtgccagt ggtgtctgag acctagggggt tggccggagg gcaggggaat   2418696
ctgacatcgg tggggcttgg ctctgtggac tctgtggggt ccagggtgag   2418746
ggt
```
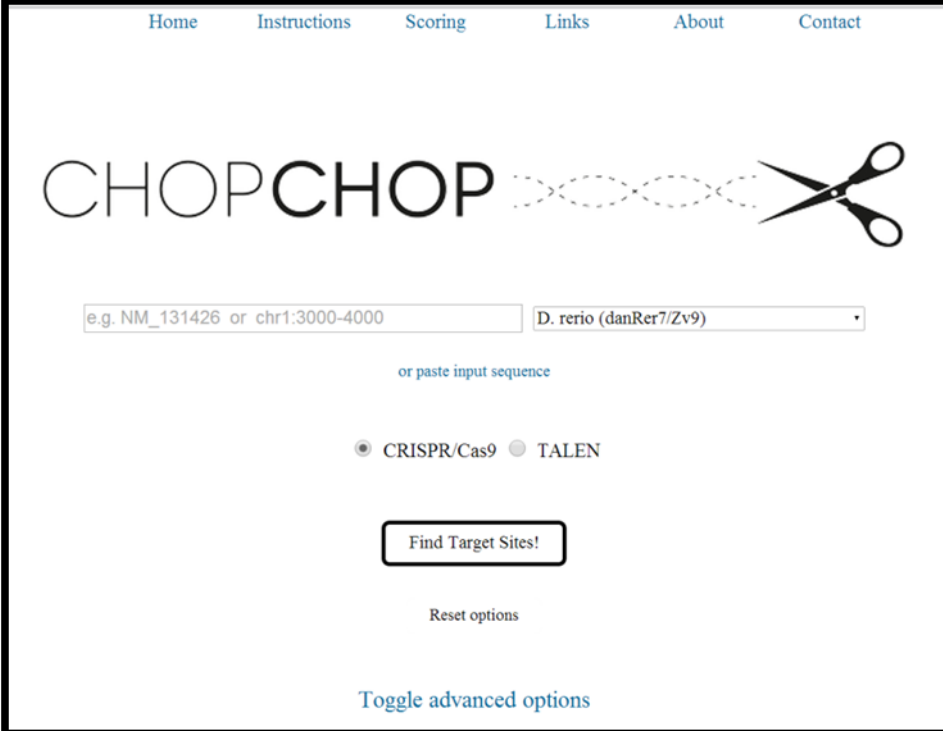
Exons colored blue, introns in black, 3' UTR in orange

Stop Site

III. Guide RNA Design

A single gRNA targeting the 3'UTR region of Human CD81 has been selected using the CHOPCHOP CRISPR/gRNA algorithm (https://chopchop.rc.fas.harvard.edu/), which identifies and ranks gRNA targets using an evidence-based scoring algorithm incorporating the # of potential off-target hits as well as GC-content.
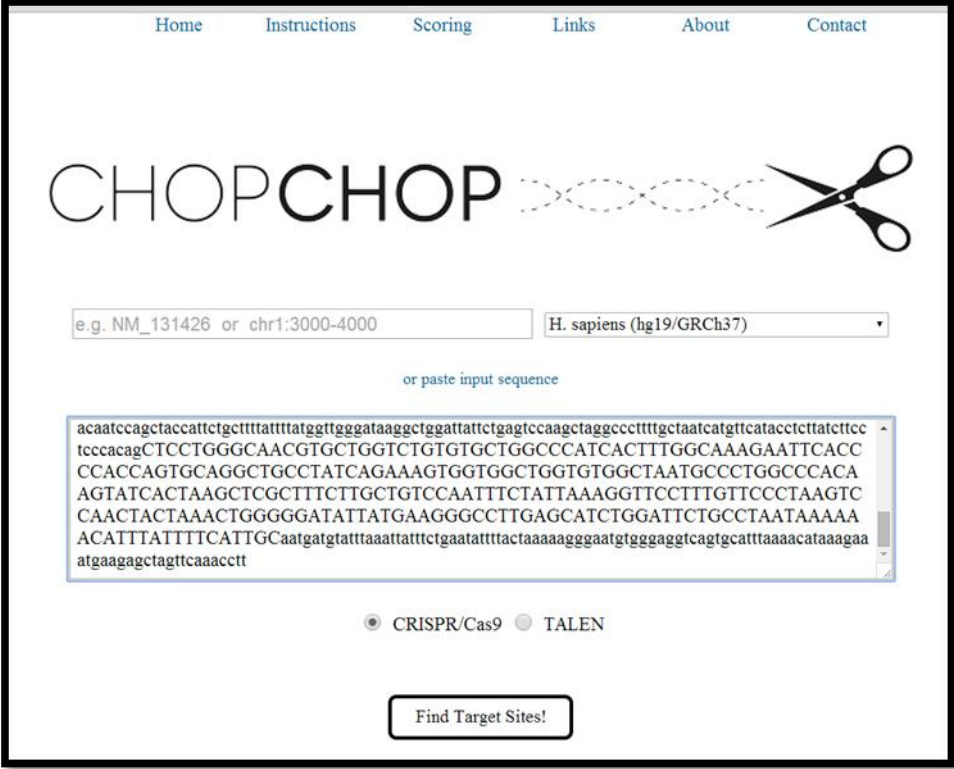

Fig. 1. Screenshot of CHOPCHOP CRISPR/gRNA design utility

1. Open the CHOPCHOP CRISPR/gRNA design tool (https://chopchop.rc.fas.harvard.edu/)

2. Grab the genomic DNA sequence of human CD81 gene from Exon 6-8 to the end of the 3' UTR region (include introns) using genome browser utilities such as UCSC Genome Browser (http://genome.ucsc.edu/). Gene ID or RefSeq accession numbers may also be used for the next step, but we find it is helpful to have a copy of the reference genomic sequence available for gRNA and HR tagging vector design.

3. Paste the sequence into a plasmid utility program (such as the ApE plasmid editor, http://biologylabs.utah.edu/jorgensen/wayned/ape/) to remove any non-nucleotide characters and copy the resulting sequence file as input into the CHOPCHOP program by clicking the "or paste input sequence" link, or enter the target Gene ID or mRNA accession number.

Note: The CHOPCHOP program provides options for changing parameters to specify general sequence and algorithm restrictions for identifying gRNAs under the "Toggle Advanced Options" link. For the purposes of this application note, the default settings can be used.

Fig. 2. Screenshot of pasting genomic DNA sequence into CHOPCHOP program

4. Click on "Find Target Sites" to get a ranked list of suitable gRNA targets (Fig. 3 on next page). We chose a guide RNA targeting the 3' UTR region with the highest possible ranking to limit potential off-target cutting and which is consistent with our tagging strategy to generate a C-terminal fusion (more information in Section III).
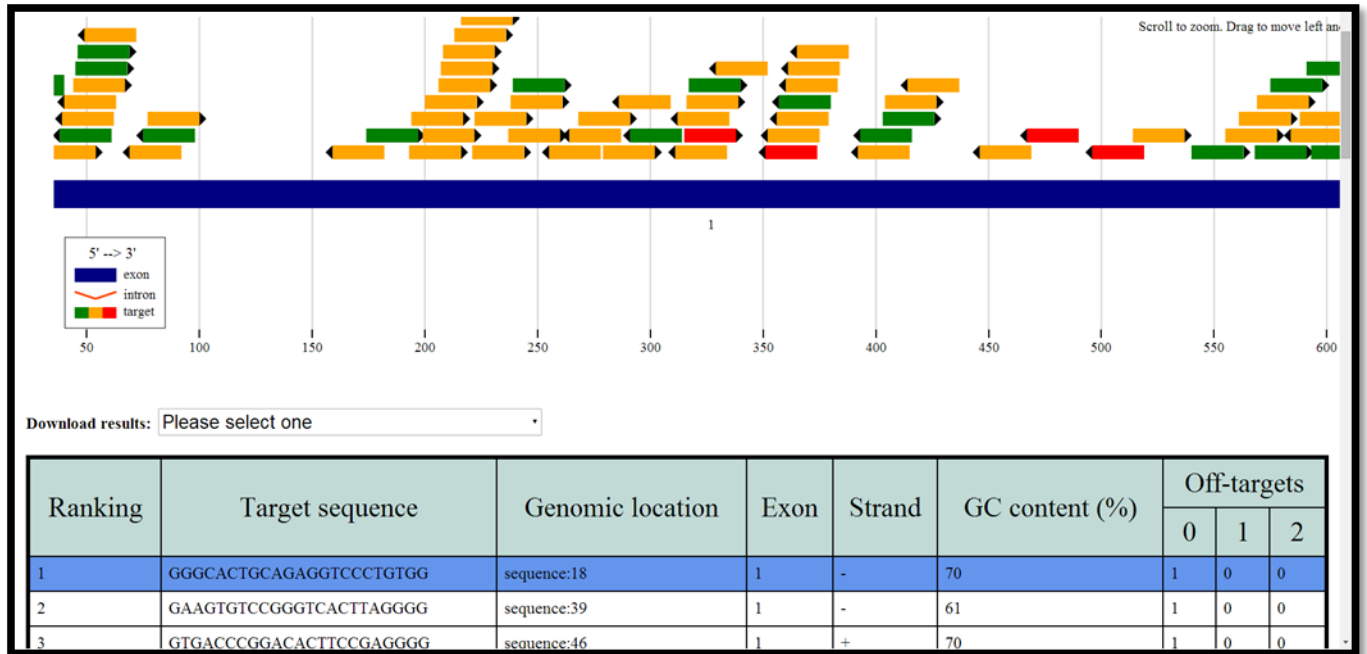
Fig. 3. Screenshot of CHOPCHOP gRNA output

gRNA ID	Location of gRNA	Target Sequence	Protospacer Adaptor Motif
gRNA-CD81-3'UTR	chr11:2418113	5' GGGCACTGCAGAGGTCCCTG 3'	TGG

The gRNA will target the corresponding sequence in the CD81 3' UTR region as shown below:

agCTTGACTGCTGTGGCTCCAGCACACTGACTGCTTTGACCACCTCAGTGCTCAAGAACAATTTGTGTCCCTCGGG
CAGCAACATCATCAGCAACCTCTTCAAGgtgcgcgaggccggtggggccgcgcctgacccccgcatgtcccgcccctgggtggggtcc
taggggtgggcaggtcacacggcagccccacagggagcgaccacactgggtggcatggcccctgtcagggctgctctgctgggagggttggggtgg
gaccgcatctggcccacgaggaaggcaggcgccctgtgctgcgcattccgggtgaagaaggtggaggctctggggggtgggaactcacctgcaccc
ccagctccacgtgtgcactcgtgggtgtggacgcccctgacagcctgtagctggcagggcctgcaggccatatagtgccctgtggaagtttcctgctg
aggcctcagtggaagtcgtcatcagtgatgctttaggggtctagtgacaccaatgaccgtgatctcagtggaaaagggcacagtgtgtcccaggcat
ttcgcgtttatgttaaaacgggtggaagatagcaagccggcagaggccgggccgctgcacccgcctgttccgaggtgggtaggggtggggggctg
ttcccaggattcccctctacgctttctgtggtgaccacggattactgcgtgacaacgggaagccgggagccgaggcccggtccctgaccacgcgtgcc
tggccacccctgcag**GAGGACTGCCACCAGAAGATCGATGACCTCTTCTCCGGGAAGCTGTACCTCATCGGCATTGCT**
GCCATCGTGGTCGCTGTGATCATGgtgagcgggcgggggcggagggcctgctctctgggctgccccttccgcggggccttgtgctgactg
cgcccccccaccaccctcctgcag**ATCTTCGAGAT**▼**ATCCTGAGCATGGTGCTGTGCTGTGGCATCCGGAACAGCTCCGTG**
TACTGAGGCCCCGCAGCTCTGGCCACAGGGACCTCTGCAGTGCCCCCTAAGTGACCCGGACACTTCCGAGGGGG
CCATCACCGCCTGTGTATATAACGTTTCCGGTATTACTCTGCTACACGTAGCCTTTTTACTTTTGGGGTTTTGTTTT
TGTTCTGAACTTTCCTGTTACCTTTTCAGGGCTGACGTCACATGTAGGTGGCGTGTATGAGTGGAGACGGGCCTG

GGTCTTGGGGACTGGAGGGCAGGGGTCCTTCTGCCCTGGGGTCCCAGGGTGCTCTGCCTGCTCAGCCAGGCCTCTCCTGGGAGCCACTCGCCCAGAGACTCAGCTTGGCCAACTTGGGGGGCTGTGTCCACCCAGCCCGCCCGTCCTGTGGGCTGCACAGCTCACCTTGTTCCCTCCTGCCCCGGTTCGAGAGCCGAGTCTGTGGGCACTCTCTGCCTTCATGCACCTGTCCTTTCTAACACGTCGCCTTCAACTGTAATCACAACATCCTGACTCCGTCATTTAATAAAGAAGGAACATCAGGCATGCTACCAGGCCTGTGCAGTCCCTCAGtgccagtggtgtctgagacctaggggttggccggagggcaggggaatctgacatc
ggtggggcttggctctgtggactctgtggggtccagggtgagggt
Exons in upper case, introns in lower case

Note: gRNA is complimentary to antisense strand

gRNA sequence
PAM motif
3'UTR region
Stop Codon
↓ Nuclease cut site

5. Clone the gRNA insert into SBI's All-in-One PrecisionX SmartNuclease cloning vector (CMV-hspCas9-H1-gRNA, catalog #CAS940A-1) following the recommended protocols in the user manual (http://www.systembio.com/downloads/Cas9-SmartNuclease-user-manual.pdf)

IV. Design and Cloning of Homology Arms for HR Tagging Vectors

SBI has built a line of HR Tagging Vectors (http://www.systembio.com/genome-engineering-precisionx-HR-vectors) that leverage the cell's ability to incorporate large exogenous DNA sequences via homologous recombination (HR) at double-stranded DNA breaks (DSB). The incorporation of fluorescent and/or selection markers via HR and tagging of specific gene sequences allows researchers to easily generate and identify clones that have the desired features.

HR Tagging Vectors are quite similar to our HR Targeting Vectors with respect to the homologous recombination process resulting from generation of a DSB in cells. They serve as the donor template to induce HR in cells that have been targeted at a specific locus using CRISPR/Cas9 or TALE nucleases.   A general HR vector will contain the following features:

1) Homologous sequences (to the template with DSBs) at 5' and 3' ends of the DSB (cutting) site
2) An expression cassette bearing a promoter, insert (cDNA, microRNA, non-coding RNA, etc.), fluorescent marker, or selection agent to select cells that have undergone HR.
3) In the case of a tagging vector, a fluorescent marker (e.g. GFP) is included as part of the 5' homologous region that allows for seamless fusion into the last coding exon of the gene to be tagged.

The 5' and 3' homologous sequences (termed "homology arms") should be exactly homologous to the template (genomic sequence) with the DSB, and preferably directly adjacent to the actual DSB site. An example of homology arm design is shown in Figure 4:
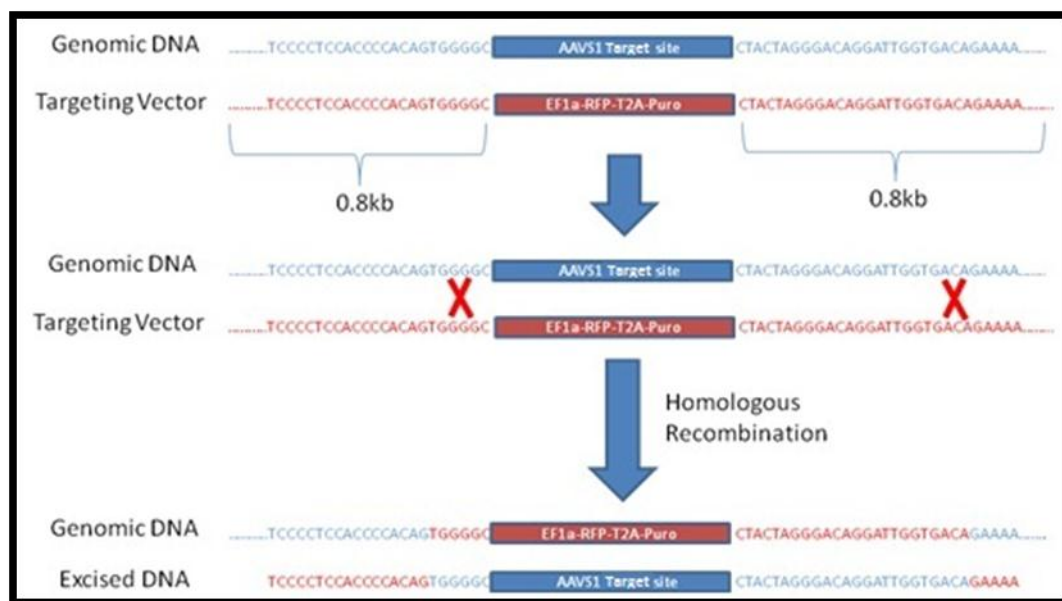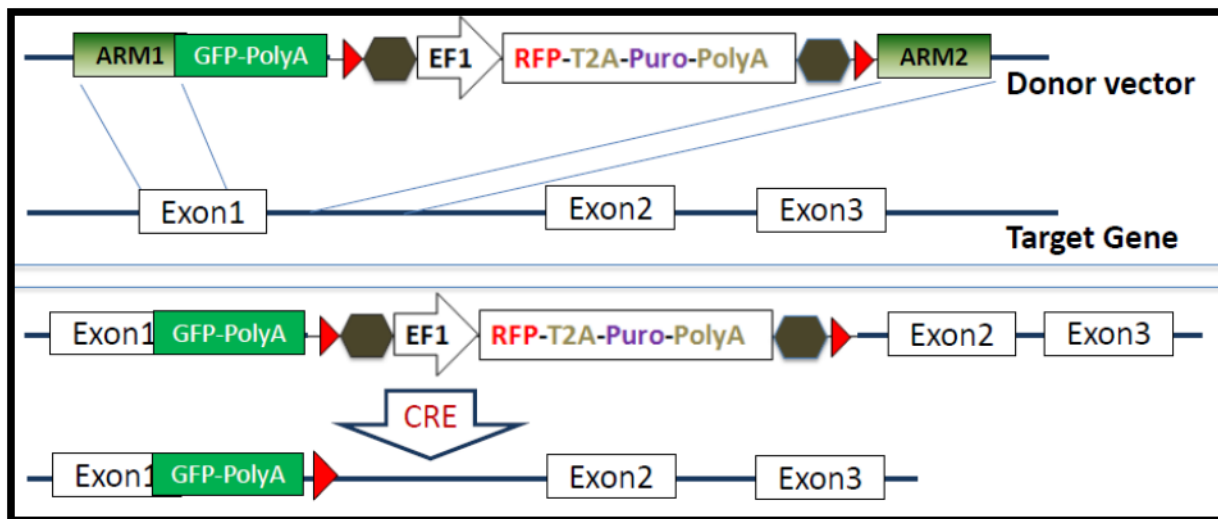

Figure 4. Schematic Diagram of Homologous Recombination (HR) Process

In Figure 4, a region of the human AAVS1 safe harbor locus (in blue) is targeted by a gRNA + Cas9 in order to insert (knock-in) an EF1α-RFP-T2A-Puro expression cassette (in red) present in the HR Targeting Vector through homologous recombination (HR). The HR targeting vector contains homology arms at the 5' and 3' end of the expression cassette which each include ~0.8kb of sequence homologous to the genomic DNA surrounding the targeted AAVS1 locus. This region of homology is crucial for the success of the homologous recombination reaction, as it serves as the guide template for specifically targeting the exogenous cassette into this genomic locus. The typical size range for homology arms varies by the application, but it should be anywhere from 0.5kb to 1kb for each arm for efficient recombination to occur. Please note that the actual regions of recombination (Red "Xs") at the 5' and 3' of the target site can vary widely, thus it is difficult to predict the actual sites as this is determined by the cell.

For gene tagging applications (see Fig. 5 below), the 5' homology arm must consist of 0.5-1kb of genomic sequence ("Arm1") upstream of the stop codon which is cloned into the HR tagging vector in such a way as to provide a seamless, **in-frame** junction between the end of the target coding sequence and the GFP marker (or T2A-GFP marker) in the tagging vector. The 3' homology arm ("Arm2") is cloned into the HR tagging vector based on sequence downstream of the target cutting site, to complete the tagging vector for use in HR as mediated by a suitable Cas9/gRNA or TALE-Nuclease.   The end result is an endogenous C-terminal fusion or T2A-linked protein (depending on the tagging vector) that can be tracked within the cells of interest for dynamic spatial studies without resorting to traditional overexpression approaches, resulting in the study of the tagged protein in a biologically-relevant context.

Figure 5. Schematic Representation of Gene Tagging using Homologous Recombination (HR) Tagging Vector

With the CRISPR/Cas9 system, the cleavage site is 2-3bp upstream of the protospacer adaptor motif (PAM) immediately following the guide RNA sequence; therefore, homology arms should be designed to be as close as possible to this cut site (<10bp) at both ends for efficient HR reaction. (For TALEN-

mediated cleavage, homology arms are designed to be adjacent to the spacer region between TALEN binding sites, which spans 15-30bp and is the site of the DSB.)

Important Caution when Designing Homology Arms for Use with CRISPR/Cas9: It is best practice to avoid including the full target sequence (gRNA sequence + PAM) in the HR Tagging Vector, to ensure that the donor vector is not targeted for cleavage by the Cas9/gRNA complex. For tagging vectors, it may be difficult to avoid adding the full sequence (particularly when using a gRNA targeting an exon) since the homology arm sequence must seamlessly fuse to the tag in the HR tagging vector. (It is easier to delete or modify the gRNA sequence from the homology arms when targeting a 3'UTR or intron). We would highly recommend gene synthesis approaches (as opposed to direct PCR from genomic DNA) to clone in HR arms into our tagging vector, which allows the user to control the exact sequence cloned into the arms. This will allow the user to mutate the obligate PAM site proximal to the gRNA target site or use silent mutations to prevent cutting of the donor vector by Cas9/gRNA complex.

For gene tagging, the most straightforward approach to construct homology arms is to use a combination of gene synthesis of the 5' and 3' homology arms and SBI's Cold Fusion Cloning system **(http://www.systembio.com/molecular-tools/cold-fusion-cloning/overview).** By synthesizing sequences that have partial homology to the cloning vector and the genomic DNA, both 5' and 3' homology arms can be "fused" into the cloning vector without the hassles (and extraneous nucleotides) required of traditional ligase/restriction enzyme cloning. This process is done sequentially for each homology arm to rapidly and accurately assemble the complete HR tagging vector for delivery into target cells in conjunction with CRISPR/Cas9 or TALE-Nuclease systems.

Summary of gRNA Design for Tagging Applications:

1. gRNA cutting site should be in the 3' UTR region

This avoids the possibility of introducing indels in the exonic sequence which may affect the reading frame of the endogenous gene and tagging vector.

2. In-frame insertion of 5' homology arm into tagging vector

The control of insertion of the 5' homology arm into the tagging vector is crucial for correct expression of the fluorescent tag such that the tag protein (e.g. GFP) is in-frame with the cloning site (e.g. EcoRI site of the cloning vector, GAATTC, is in frame with ATG of copGFP – please see the figure for HR120PA-1 vector below). Gene synthesis of the arms will allow for precise control of the homology arm sequences to avoid frame issues.

3. Seamless cloning of homology arms is best done using a combination of gene synthesis and Cold Fusion Cloning

By using synthesized homology arms with appropriate end sequences appropriate for Cold Fusion Cloning (see 5' and 3' arm design schematics below), linearization of the tagging vector is the only step involved for seamless ligation of the synthesized homology arms into the tagging vector without introduction of additional nucleotides.

4. gRNA PAM motif site should be mutated or deleted in the completed tagging vector

In order to prevent targeting of the completed tagging vector by CRISPR/Cas9, the gRNA PAM site should be mutated or deleted, which can be done during the synthesis of the 3' homology arm which will contain the PAM site.

Final Homology Arm Design for GFP-tagged CD81 Protein (Compatible with Cold Fusion Cloning System) Using HR120PA-1 vector

 5' Arm Design

5' AAAACGACGGCCAGTGAATTC -- [1.0kb of Homology Arm Sequence Upstream of Stop Codon – **DO NOT INCLUDE STOP CODON**] – GAATTCATGGAGAGCGACGAG 3'

 Vector Homologous Sequence
EcoRI Restriction Site

3' Arm Design

5' GAAATAACCTAGATCGGATCC GGCCCCGCAGCTCTGGTCACAGGGACCTCTGCAGTGCCC -- [1.0kb of Homology Arm Sequence Downstream] – GGATCCCCGTCGACTGCATGC 3'

Vector Homologous Sequence
BamHI Restriction Site
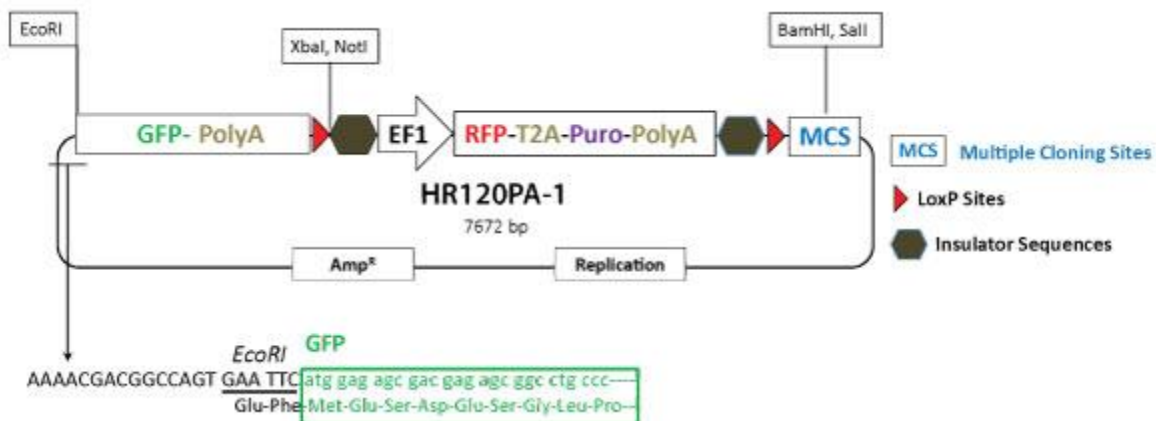PAM site
PAM Mutation C->T
gRNA target site

Cloning of the 5' Homology Arm into HR Tagging Vector (HR120PA-1)

The completed 5' homology arm can be readily cloned into our HR Tagging Vector (e.g. HR120PA-1) using SBI's Cold Fusion Cloning Kit:

1. Linearize 1-2ug of the HR120PA-1 vector using EcoRI enzyme, preferably a "high-fidelity" version such as EcoRI-HF to ensure complete digestion of the vector.

2. Clean-up the digestion reaction and run out the reaction on a 2% agarose gel. Gel-purify the insert, and quantitate DNA using NanoDrop or other suitable UV-Vis spectrophotometric method. Dilute the purified DNA to a final concentration of 10-100ng/ul.

3. Resuspend the 5' homology arm in suitable buffer to a final concentration of 20-200ng/ul.

4. Please refer to pgs 6 and 7 of the Cold Fusion user manual (http://www.systembio.com/downloads/Manual_Cold_Fusion_083010.pdf) for detail on setting up the cloning reaction.

5. Screen colonies for clones with the correct inserts by restriction digestion analysis and sequencing

Cloning of the 3' Homology Arm into HR Tagging Vector (HR120PA-1)

1. Using a miniprepped clone (from Step 5 above) containing the correct insert, linearize 1-2ug of the vector using BamHI enzyme, preferably a "high-fidelity" version such as BamHI-HF to ensure complete digestion of the vector.

2. Clean-up the digestion reaction and run out the reaction on a 2% agarose gel. Gel-purify the insert, and quantitate DNA using NanoDrop or other suitable UV-Vis spectrophotometric method. Dilute the purified DNA to a final concentration of 10-100ng/ul.

3. Resuspend the 3' homology arm in suitable buffer to a final concentration of 20-200ng/ul.

4. Please refer to pgs 6 and 7 of the Cold Fusion user manual (http://www.systembio.com/downloads/Manual_Cold_Fusion_083010.pdf) for details on setting up the cloning reaction.

5. Screen colonies for clones with the correct inserts by restriction digestion analysis, sequencing, or PCR. You should verify correct insertion of the homology arms at both the 5' and 3' ends before transfection of target cells (See Section IV)

The final HR tagging vector after cloning of the HR arms would look similar to the following:

AAAACGACGGCCAGT<mark style="background:cyan"></mark><mark style="background:red">GAATTC</mark> **[1kb of 5' Homology Arm Sequence]** <mark style="background:red">GAATTC</mark>**[copGFP-LoxP-Ins1-EF1a-RFP-T2A-Puro-Ins2-LoxP]**GAAATAACCTAGATC<mark style="background:magenta">GGATCC</mark>GGCCCCGCAGCTCTGG<mark style="background:lime">T</mark><mark style="background:grey">CA</mark>CAGGGACCTCTGCAGTGCCC**[1kb of 3' Homology Arm Sequence]** <mark style="background:magenta">GGATCC</mark>CCGTCGACTGCATGC

<mark style="background:cyan">**Vector Homologous Sequence**</mark>
<mark style="background:magenta">**BamHI Restriction Site**</mark>
<mark style="background:red">**EcoRI Restriction Site**</mark>
<mark style="background:grey">**PAM site**</mark>
<mark style="background:lime">**PAM Mutation C->T**</mark>
<mark style="background:teal">**gRNA target site**</mark>

Cells which are successfully targeted will have GFP fused to the last coding exon of human CD81 as well as the rest of the donor cassette [**LoxP-Ins1-EF1a-RFP-T2A-Puro-Ins2-LoxP**] downstream. The elements between the loxP sites can be readily excised out by transfecting targeted cells with Cre recombinase (Cat #CRE100A-1)

V. Protocol for Co-Transfection of HR120PA-1 Tagging Vector with Cas9/gRNA plasmids and Characterization of Cells

1) Plate 200,000 to 300,000 cells (e.g. 293T cells) into a single well of a 12-well plate in 1 ml of appropriate growth medium. Include a single well of cells as negative control (which can be non-relevant plasmid DNA)

2) Next day, or when cells are 50-60% confluent, co-transfect target cells with Cas9 plasmid(s) and the HR targeting vector using a suitable transfection reagent following the manufacturer's recommended protocol for 12-well plates. The use of reduced or serum-free media containing no antibiotics to dilute the vector/transfection complex is highly recommended.

Note: For 293T cells, we suggest 0.5 µg of SBI's Cas9 SmartNuclease vector in conjunction with 0.5 µg of the HR tagging vector into cells for efficient cleavage and HR reaction. For other cell types, we suggest optimizing the amounts and ratios of Cas9 plasmid to targeting vector for optimal results.

3) Allow at least 12 hours before changing transfection media to complete growth media

4) Assay for positive HR events 96 hours after co-transfection. Select cells with insertion of tagging vector using fluorescent or antibiotic selection. If using selection by Puromycin, select cells for a minimum of 5-7 days prior to further characterization. Cutting efficiency of Cas9 can be measured by Surveyor Nuclease Assay and HR efficiency by % of fluorescence signal via FACS sorting.

5) After selection of cells in Puro, remaining colonies can be propagated for further characterization by PCR genotyping and Sanger sequencing to confirm tagging of one or both alleles. SBI offers the EZ-Genotyping kit for fast characterization of engineered cells:
http://www.systembio.com/genome-engineering-ez-genotyping-kit

6) Colonies possessing the desired tagged allele(s) can be subsequently passaged and clonally propagated.

7) If desired, the remaining cells can be transfected with the Cre recombinase to excise the cassette inserted into the genome.   This will preserve the GFP-tag, but will remove the Puro and RFP cassettes present in the HR120PA-1 backbone.