Reference Guide – Designing CRISPR-Cas9 RNA Oligonucleotides for the No-Amp Targeted Sequencing Procedure

Introduction

The purpose of this document is to provide technical guidance on how to generate and select CRISPR RNA oligonucleotides (crRNA) that are compatible with the PacBio *Procedure and Checklist, No-Amp Targeted Sequencing Utilizing the CRISPR-Cas9 System*.

CRISPR-Cas9 RNA oligonucleotides are comprised of CRISPR RNA (crRNA) and trans-activating CRISPR RNA (tracrRNA). When annealed together or synthesized as one continual construct they makeup guide RNA (gRNA) or single-guide RNA (sgRNA), respectively. Regardless of the guide RNA type, the tracrRNA enables gRNA to be complexed with Cas9 nuclease. At the same time, the crRNA provides the targeting function of the gRNA-Cas9 complex to enable double-stranded digestion at a specific genomic location.

The PacBio procedure referenced above makes use of gRNA as opposed to sgRNA. The tracrRNA oligo is comprised of a universal sequence that does not require customization. It is the crRNA that requires customization for a specific target. The No-Amp procedure uses two gRNAs to excise the target region from genomic DNA (gDNA).

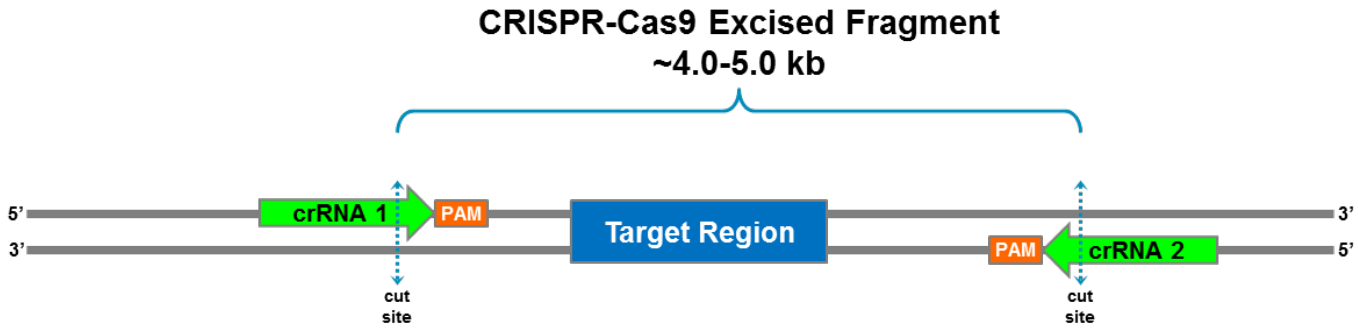


Figure 1. A depiction of a generic No-Amp targeted sequencing assay design. Various aspects of this illustration will be addressed in the design steps that follow. Note that the two flanking crRNA binding sites must be upstream or 5' to the target region with the 3-bp protospacer adjacent motif (PAM) sites closest to the target region necessitating a trans configuration for the crRNAs. The PAM site is a required targeting element located immediately 3' of the crRNA design in the target sequence, not in the crRNA design sequence.

This targeted sequencing method was developed for analysis of HTT, FMR1, ATXN10, and C9orf72 repeat loci but is readily applicable to other targets. The guidance below covers the design process for the ATXN10 repeat locus (as an example).

Part Number 101-839-600 Version 01 (August 2019)

Design Steps

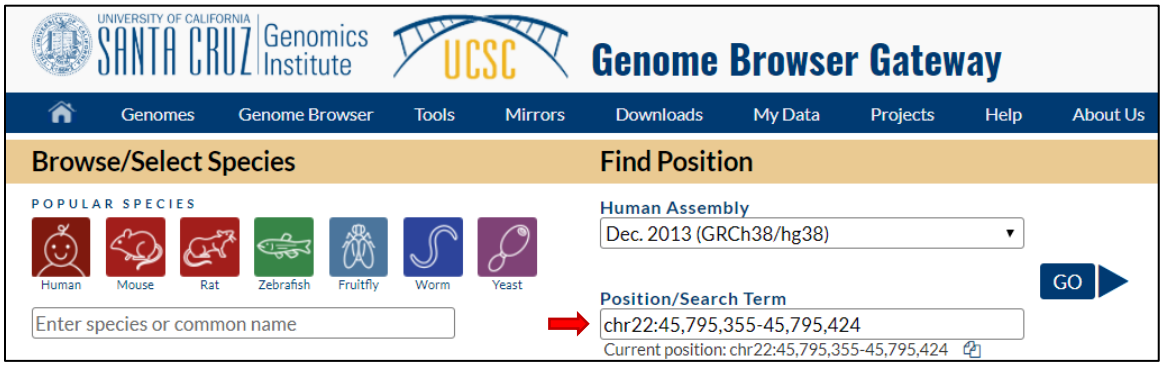
There are 4 steps to designing a pair of crRNAs for a particular target region:

1. Retrieve gDNA sequence with common SNPs identified
2. Generate crRNA designs
3. Select crRNA designs
4. Test crRNA designs

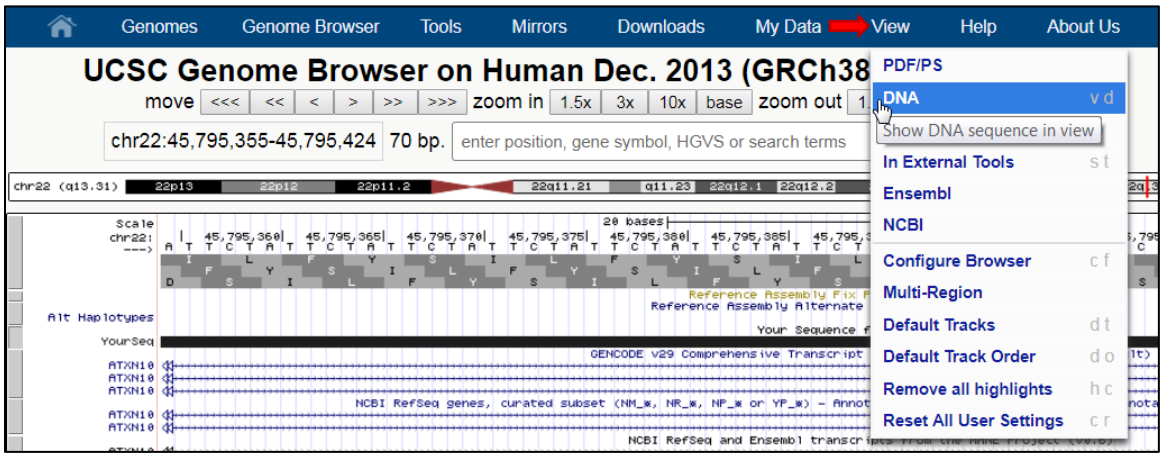
Retrieve gDNA Sequence with Common SNPs Identified

In this step, 5,000 bp of gDNA sequence containing the target region is retrieved in preparation for generating crRNA designs. Use a retrieval tool capable of marking common SNPs in the sequence. Common SNPs located within the crRNA design sequence or PAM site may alter the binding efficiency of the crRNA in the gRNA-Cas9 complex, ultimately reducing the number of on-target sequencing reads. The following instructions use the UCSC Genome Browser website (https://genome.ucsc.edu/).

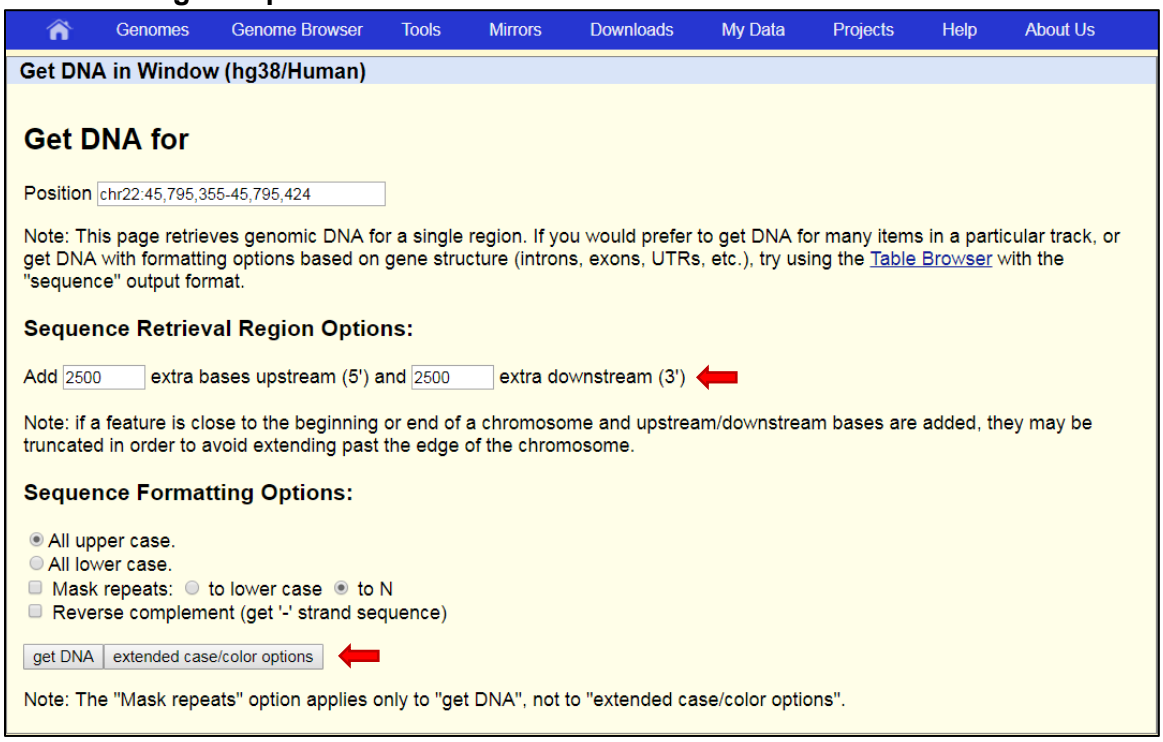
1. Identify the location of the target region. Here the target is the ATXN10 repeat locus in Chromosome 22 with ATTCT repeat motif (chr22:45,795,355-45,795,424; GRCh38/hg38). Use the latest reference assembly build available.
2. On the UCSC Genome Browser website, enter the chromosome and coordinates of the region of interest in the **Position/Search Term** field as shown below and click the **Go** button.



3. In the next window, choose **DNA** under the **View** menu at the top of the webpage.

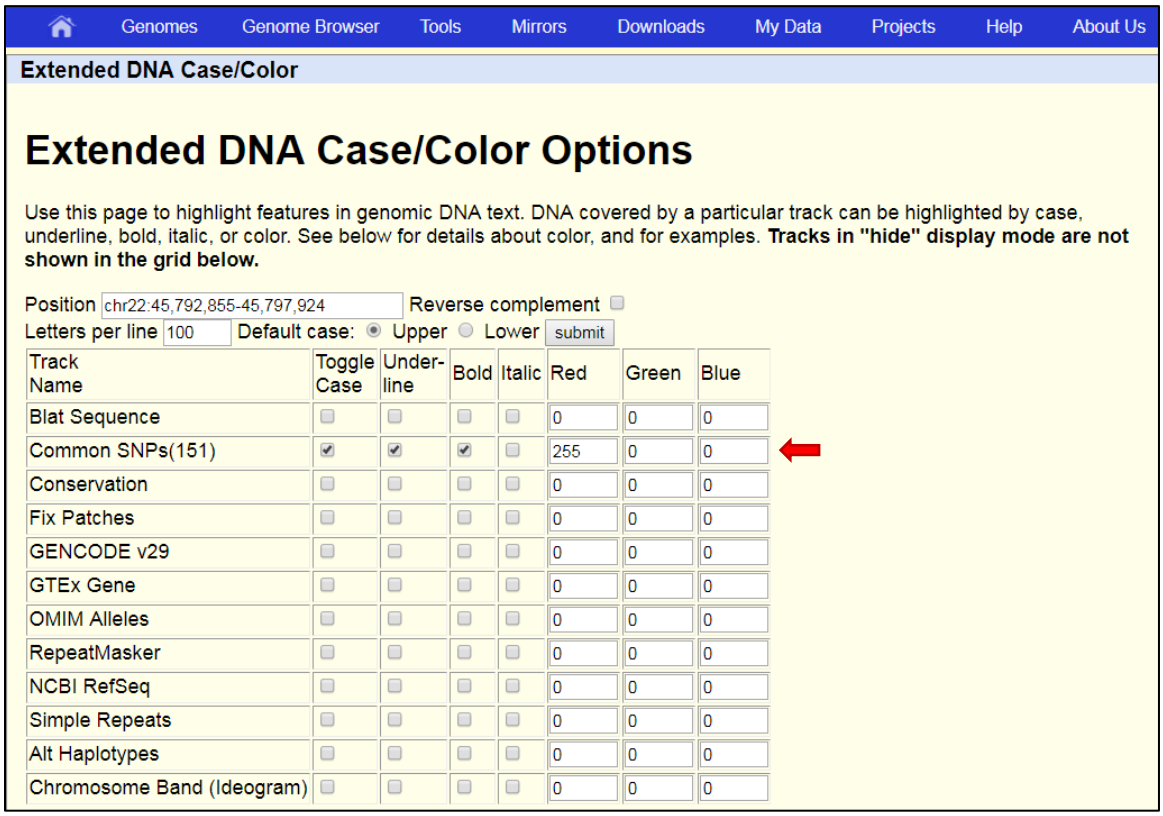


4. Retrieve 5,000 bases surrounding the target region (2,500 bp on each side). In the next window, enter 2,500 bp into the fields for extra upstream and downstream bases under the **Sequence Retrieval Region Options** section.



5. Click the **Extended Case/Color Options** button to advance to the next webpage and enable marking of SNPs in the sequence.

6. Select the desired SNP marking attributes and retrieve the sequence by clicking the **Submit** button.



7. Copy the resulting target sequence into a text editing program (e.g. Microsoft Word) and mark the region of interest (highlighted yellow below). Note that common SNPs are marked as instructed above.

```
>chr22:45792855-45797924
TCGACATTTTTACTGACTTTGTTAAATTCTTCTAAACTTGCATTGGCTGTTATACCTACATAgGGAGCCACATGAAGATCATAGATAAGGTGAAT
CAAATGATTTACATGTTTACTGACAAGAGCTTTTTTAGTTCCACCCAAATGCTTTCTCCTAATAGCCCAGCCAGGCCTTTTGCAGGCTCTGCAGT
TGCTTCTCAGGACCGTGGTCACCAGGTTTGTCTACAGTTCTCCCACAAAACCTGACATATCTTAAAATTCAAAAACAATTGTATTGTACTCTGTT
GATGTCCAACAGACTTTTCTTATTTTAAAAAATCTCCATAACCATTAAATTAGATAAAGTACCATCAAAACCCAAAGACATAGAATTGTTTGCTA
GAGAGCTTTGTGTTCAGCATCTCCTAGTCAGTTGTGAATTTGTCCTCAGCATGTCTTCCATCAGTGACAACATCAgTCCAGAGTGTAGTAATTTA
AAAATTGTCTTCACGAAGCCTGGCAGTAATTGTTCTGTGCtTTTCTGCTGTCCATCATCCAAATACCAGCTGACGGAAACATCTACTGCATTTCA
ATAAGGGTAACAATGGTAGCAGAGTCCACATCTTGAGGTCTTATTCCATTTAATAAATCTCTCACTGCTCCTTCTGCTATTCTGAATTCACGATT
CTCTTTAACAAATAATTCTGTCAAGCTCTTCAATCTTTCTGTGACAAATGACTAGAATTCTGGAGCCTGCACCTTCATTTTAGGCAGCTATTGCT
TCAGAAGTCACAGTCTGTTCCACCAGCCTTTGCCAAGGATGCAGGTGCCACAGCATCTCAAGAGAAGTCACCAATCACACAGCTCCATGCTGAGG
CCCTCTTGCCTGCTACTGAGGAGCTCTTGCCACCCCCGATTCCTGCCTTTCCCCAGTGATGCAGGACAGGTAAGCCCCCAAATTGGGACTTAGCC
TGGGAAGGTTCTTGGCTTTGCCCGGGAAATAATTCAAGGGTGAGCTGGTGGTGGTAAACAGCAACTTTGATTGAAGCAGCAGTGTATAGCAGCAG
CAGAGGTGCTGCTCCTTGCAGAGCAGGGCTGCCCTATAGGCAGTGTGCCCAGAATAGCAGCTCAGAGACAGGTCTGCACTCATATTTATACCCAC
TTTTAACTATATGCGAATTAAGGGGCAGTTTATGCAGAAATTTCTAGGAAAAGGGTGGCAACTCTTAAGTCATCAGGTTGTTGCTgTGGaAAGGG
GCGGTAACTTCTGGGTGTTGCCATGGCAATGGTAAACTGACATGACATATTGGTGGGTATGTCTTATaGAAAATGGCTTCCACCCCTGACCTACC
TGTTTTAGTTAGTCCTCAATTTGGTCAGGTGTCTGAGCCCTGCCTCCTACCTCTCCAGGACATGGGCAGTAGCCTGGCTGGGCACTTTAAAAAGG
ACAGTTGTATCCAAGCCTTTGAttTTTTTCTGTCAATTTTAACATTTTCTGTCTCAGAAATATCTATTTCATCTAGATTTTCACATTTTTTTCAT
AAAATTGTTCATATTTTAATATTTTATCTAGGTTGTCTAAAACCTGTTTCTCCCTTTTAATTCTATTCTTTCTTAAAAAAAAAAAAAGGATGTCT
CTTTTTTCCTTCCTCAGTCTTTCTGGAATTTTCCATTCTATTAATGTTTACAAATAACCAGCTTTTGGTTTTGGCTAATCTTCTCTGTAGTGTTG
TTGTTCTAATTCACTGACTTTCTTGCCTGTCAGTTTATAAGAACAAAGATAAAAATGATAACAGGCTTTCCTTCAGAAATGATGCAAGACAGGAA
TCAATGGAATGACATCTTAAAAATACTAAAGGAAAACTGTTAATCTAGAATTATGTATCCAGCCAAAATATCCTTCAAAAATGAAGGTcAAAAAA
AAGAAGGTAAAATGTCCTGTAAGAAATGTTAAATGAAGTTCAACAAGGTGAAAGTAAATGATAGCACATGGAAACCTGACCTATACAAATATATA
CAAAGGATCAGAATCCCTGGAAAAGTTAAATATATGGGTAAAAAGAAAAGATTTCTATTCTCACTTTTAATTTTCTTTGAAATGTAATTGGCTCT
TTAAAGTAATCATAATGGCAGTATATTAGGGGATTTTTAATGTGTATATAGAAGAAAAATATATGAGAGTAATTGCCCAAAGATTAAACAGAGTA
TACAGTTACAACATTCTTATATTGTATGTGGGGTAGTATGGTATTATTTGTAGATAGACTGCAACAAGGTAAAGATGTATACTCTAAACTCCAGT
GCAACCACTTTTAGAAAAAGAGAGAGACAGTTAACAAGCCAGTATTGGAGAAATAAGTGGAATATTTAAAAATGCTCAGTCTCAAAAGAGTCTGG
AAAATAAGAACAAAGAATAGTGAAGACAAATAGAAAACAGATGGCAGAATGATAAACTCAATCATGTTGATAAATATATTAAATGTAAATGGCTT
AAATATCCAACTAAAAGACTACTAGAATGGATTCTATTCTATTCTATTCTATTCTATTCTATTCTATTCTATTCTATTCTATTCTaTTCT
ATTcTtTTGAGATGAAGTCTCTCTATGTTGCCCAGGCTGGAGTGCAGTGGCgCGATCTCAGCTCACTGCAACCTCTGCCTACCAGGTTCAAGCA
ATTCTCCTGCCTCAGCCTCCCAAGTAGCTGAGATTACAGGTTCACCACCACACCCAGCTAATTTTTGTATTTTTAGTAGAGACGGTGTTTCGGGA
AGTCAGGGACCCTGAACGGAGGGACTGGCTGAAGCCATGGCAGAAGAACATAAATTGTTAAGATTTCATGGACATTTATTAGTTCCCCAAATTAA
TACTTTTATAATGTCTTACACCTgTCTTTACTGCAGTCTCTGAACATAAATTGTGAAGATTTCATGGACATTTATCACTTCCCTAATCAAcACTC
TTATAATTTCCTATGCCTGTCTTGTCTTTAATATCTTAATCTCGTCATCTTCATAAGCTGAGGATGTATGTCACCTCAGGATCCCGTGATGATCA
CGTTATCTGCACAAATTGTTTGTAAAGCATGTGTGTTtGAACAATATGAAATTTGGGCACCTTGAAAAAGAACTGGGTAACAGCGATTTTCAAGG
AACAAGGGAGATAACCATAAGGCCTGACTGCCTGCAGGGCCGGGCAGAACAGAGTCATATTTCTCTTCTTTCAGAAAGTGACTAGGAGAAATATC
GCTGAATTCTTTTCTCAGCAAGGAATAACCCTGGGAAACGAATGCATTCCCAGGGGGAGgTCTCTAAAATGGCCGCTCAGGGAGTGTCTGTCTTA
TGCAGTTGTAGATAAGGGATGAAATACTCCCTGGTCtCCTGCAGCGCCCTCAGGCTTGCTAGGATTAGGAAATTCCAGCCTGGCGAATTCTAGTC
AGACCGGTTGaCTGCTCTCGAACCCTATTTCCTGTTAAGATGTTTATCAATGACAGTGTGTGCCCAGCAGGACATGGACCTTCATCAGTAATTCT
AGTTTCGCCCTGGCCTTGTGATCTCACTCTGCCTCTCTGCCCTTGTGATATTTCATTGCCTTTGAAGCATGTGATCTCTGTGACCCACACCCTGT
TTGTACACCCCTCCCCTTTTGAAACCCCTAATAAAAAGTTACTGCTTTTGtGGCTCAAGGGGCATCACGGAACCTGCCAACATGTGATATCGCTC
CCAGAGACCCAGCTGTAAAATTTCTCTCCTTTGTACTCTTTCTCTTTATTTCTCAGACCGGCCGACACTTAAGGAAAATAGAAAAGAACCTACGT
TGAAATATTGGGGGCTGGTTCCCCCGATAGATGGGGTTTCACCGTGTTGGTCAGACTGGTCTCGAaCTCCTGACCTCAAATTATCCTCCCACCTT
CCGCCTCCCAAAgCATTGAGATTACAGGCGTGAGCCACTGTGCCTGGCCTATACTGGATTTTGAAAGGCAAGACTCAATTATATGCTATGATAAG
AAATATACTTTGAATATAAAGGTGAAGATACCTTAAGAGTGAAAGAATGATATACCATGTAAACACTTATAAGAAAGCTGGAGTGGCTATATTAA
TATTAGACAAAATAGACCTAAGGACAAAGAATTTCCCTAGATGTAAAAGAGTGATATTTCaTAATGATTAAAGGGTCAGTTCATCAAGAAGACAT
AATAGCCTAAACATACATTGCTAATAGcAGAATTTCAAATAACTGAAACCAAAACTGACACCTGAAAGAAAAAATCGATCTGCAGTCATAATCGG
AGATATCAGcGTTCTTATCTTTATAAATTACAGAACAAGCTGAAAGAAAAGCAGGAGGAATATGGAAAACTGGAATACCATCAGCCACCTTGATC
AgGTGAAATTTAGTGAACACTGCACCAATAACAGCATGTATGTTCTTTCCAGTACACATAGAACATTCACCAAGACAGACCATATGCTGGGGCCA
TAACAGTGCTCAATAAATGTAGAAGAACTGAAAACATAGACTATATTCTGTCTCTACAGCAGAATTATAAGTTAAAACTATAAAATATCTCTAAT
ATCCCAAGATATTTGAAAATTAAACAGTACACTTCTAAATAACCCACATGTCATAGACACAACCATATCGAAATTAGAAAATATTTTGAACTGAA
TAAAAATGAAAAATGGCATACCTAAAtTGTAAGATGGGGATAAAGTAGTGTTTAGGAAATTTAAAATACATGAGAAAACAAGACAGATTTAAGAT
CAGTCATCTTAGTTTCCAACTTAAGAAGGTAATAAAACCAAGAGTAGATTAAACCAATGTAAGTAGAAGAAAAGAATTAACAAAGATAAGAATGG
AAATCTTTGAAGTAGAAAaTAAAATAGACATAGAGTGGAGAAAGTCCATGTAATCAGAAGCTGGATGCATATTTAAAAAgAtTGATCAATCTTTC
CAGATAAATCTgTCTACCTGTCAAGAAAATGACACAAAATATAAATGACCTCTATCAGCAGTGAAAGAGGAGACATCATTaTAGATATTACAGAC
ATTAAAAAGATAATAAGAGAATAGCATAAACAACT
```

Generate crRNA Designs

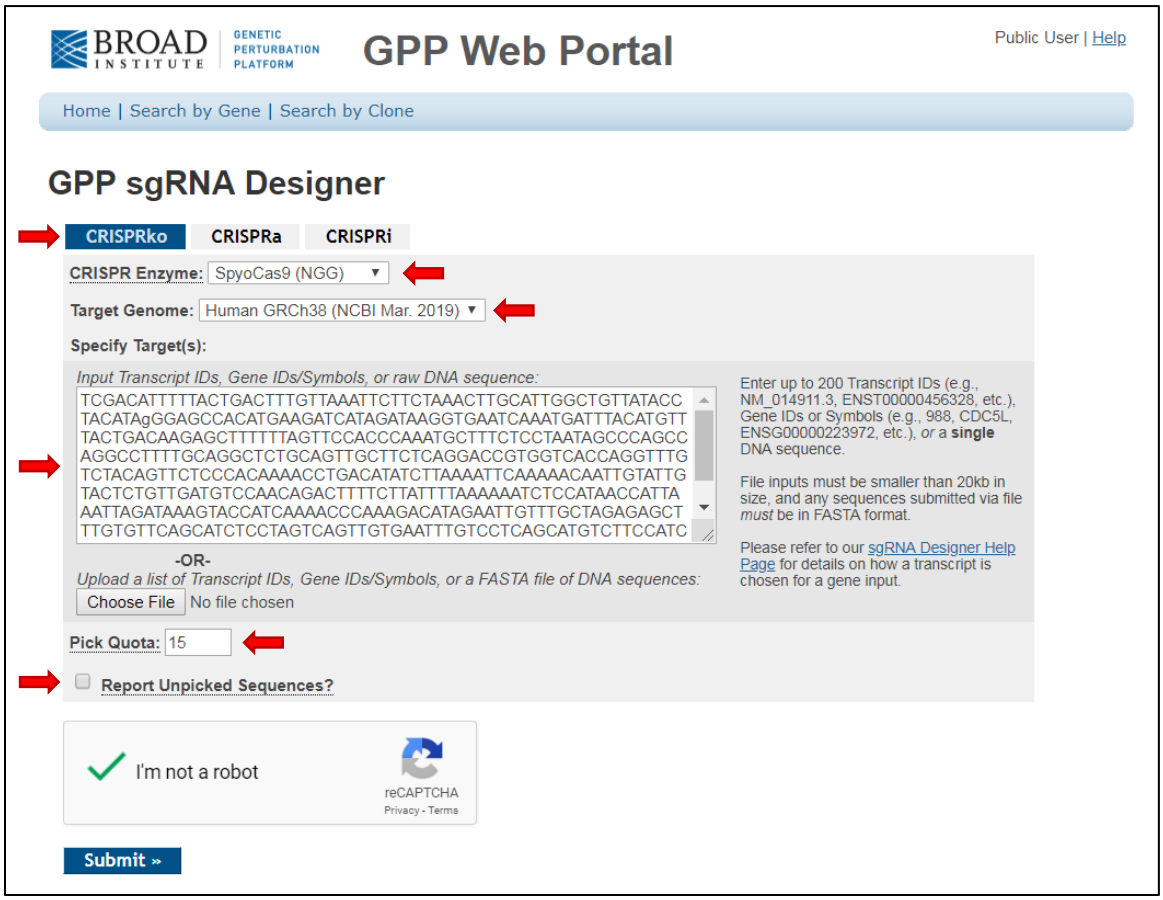
In this step, the retrieved sequence is submitted to an online CRISPR-Cas9 design tool to generate several crRNA designs. Note that many websites use the terms crRNA and gRNA interchangeably, but, as explained in the introduction, crRNA is the targeting component of gRNA. Also, at times a distinction isn't made between gRNA and sgRNA. As pointed out in the introduction, gRNA is constructed by annealing separate crRNA and tracrRNA oligos together and sgRNA is formed by synthesizing the crRNA and tracrRNA as one continuous oligonucleotide.

The following instructions use the GPP sgRNA Designer webtool on the Broad Institute GPP Web Portal website (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design). Designs generated by this webtool are compatible with the No-Amp targeted sequencing procedure.

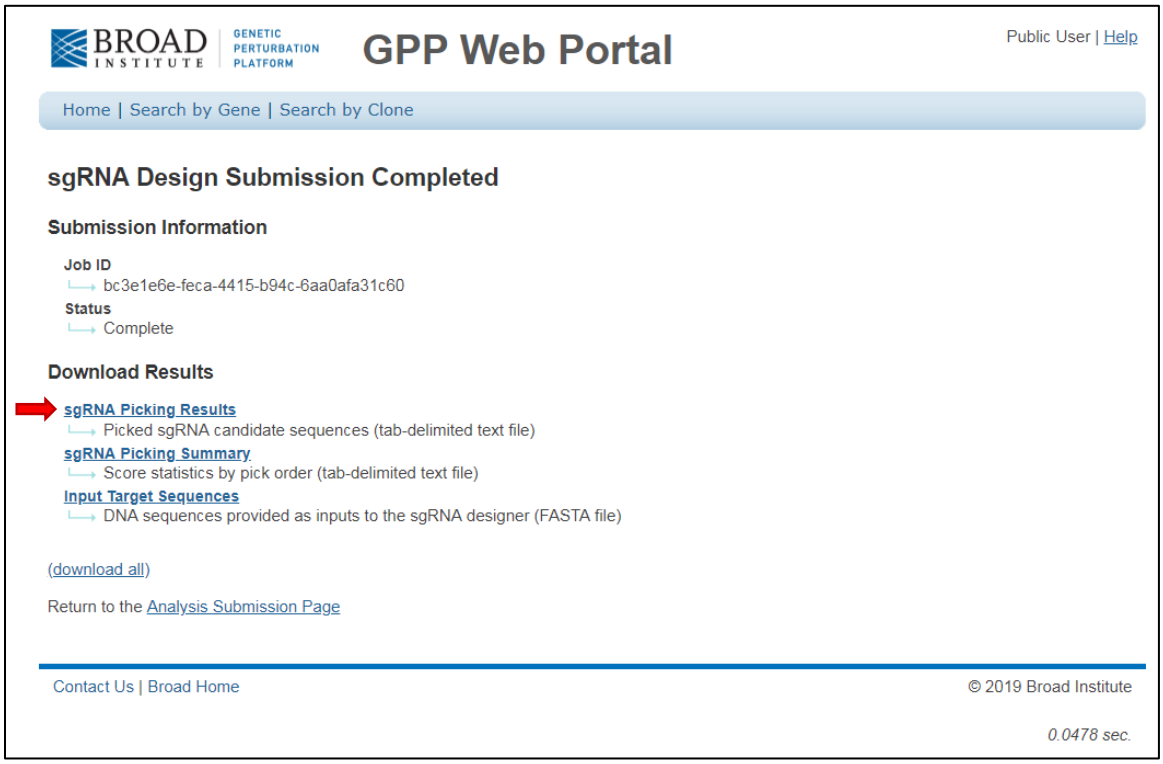
1. Select the first 500 bp from the beginning of the retrieved target sequence and submit to the GPP sgRNA Designer webtool. Despite the "sgRNA" reference in the website screen capture below, it is crRNA designs that are generated.

   Make sure the **CRISPRko** tab is active and that **SpyoCas9 (NGG)** is selected in the **CRISPR Enzyme** pull-down field. The No-Amp procedure uses Cas9 nuclease from *S. pyogenes* and is associated with the specific PAM sequence, NGG.

   Under the **Target Genome** pull-down field choose the latest build. Enter a value of "15" in the **Pick Quota** field and make sure the **Report Unpicked Sequences?** box is unchecked. Click the **Submit** button to generate results after verifying you are not a bot.



2. Download the “sgRNA Picking Results” file in the next window by clicking the link.



3. Repeat Steps 1 through 2 in this section for the last 500 bp at the end of the retrieved target sequence. In the next section, crRNA designs flanking the target region will be selected from the first 500 bp of sequence and last 500 bp of sequence. This will generate a 4-5 kb fragment containing the target region after CRISPR-Cas9 digestion.

Select crRNA Designs

In this step, crRNA designs are selected from the output of a CRISPR-Cas9 design tool based on orientation and uniqueness. They are then mapped to the gDNA target sequence obtained earlier in the design process to determine the presence of SNPs in the designs. Final design selections are then made for testing. Following the previous section, the "sgRNA Picking Results" files generated separately for the first and last 500 bp of the 5 kb target sequence are used below.

1. Open the "sgRNA Picking Results" files (tab-delimited text file) in Microsoft Excel and save each as an Excel Workbook (.xlsx file extension).
2. Sort the tables by the "Pick Order" column to order the crRNA design picks as shown in Tables 1 and 2 below in the next step. Only the columns to aid in mapping the designs to the 5 kb target sequence are shown, and other columns have been hidden. Again, "sgRNA Sequence" as appears in the table is synonymous with "crRNA Sequence".
3. Identify crRNA designs compatible with the No-Amp procedure.

   Orientation of the crRNA design relative to the target sequence is **critically** important to the No-Amp targeted sequencing procedure. The gRNA-Cas9 complex cuts 3-bp upstream of the associated PAM site. After cleavage, the gRNA-Cas9 complex remains associated with the DNA on the 5' side of the cut site, resulting in low or no read coverage of the associated DNA.

   Compatible crRNA designs must have 3' ends directed toward the target region as shown in the figure below. Fragments generated with the correct crRNA design orientation will not be encumbered by the presence of the gRNA-Cas9 complex in the ligation step following CRISPR-Cas9 digestion.

Figure 2. Illustration of four possible locations for crRNA designs for a generic target region. The two flanking crRNA designs for any target region must be located on opposite strands with the 3' end of each design closest to the target region. This will also place the PAM site in-between the crRNA design and the target region. Designs crRNA1 and crRNA2 are compatible with the No-Amp targeted sequencing procedure whereas designs crRNA 3 and crRNA 4 are not (shown crossed-out) as the 5' ends of these designs are closest to the target region with the PAM sites not adjacent to the target region.

In both tables below, the rows containing designs with compatible orientations are labeled in italicized, bolded text. Compatible “forward” designs generated for the first 500 bp of sequence before the target region are located on the sense strand (Table 1). Compatible “reverse” designs generated for the last 500 bp after the target region are located on the antisense strand (Table 2).

Table 1. *Forward crRNA designs from the first 500 bp of the 5 kb ATXN10 sequence.*

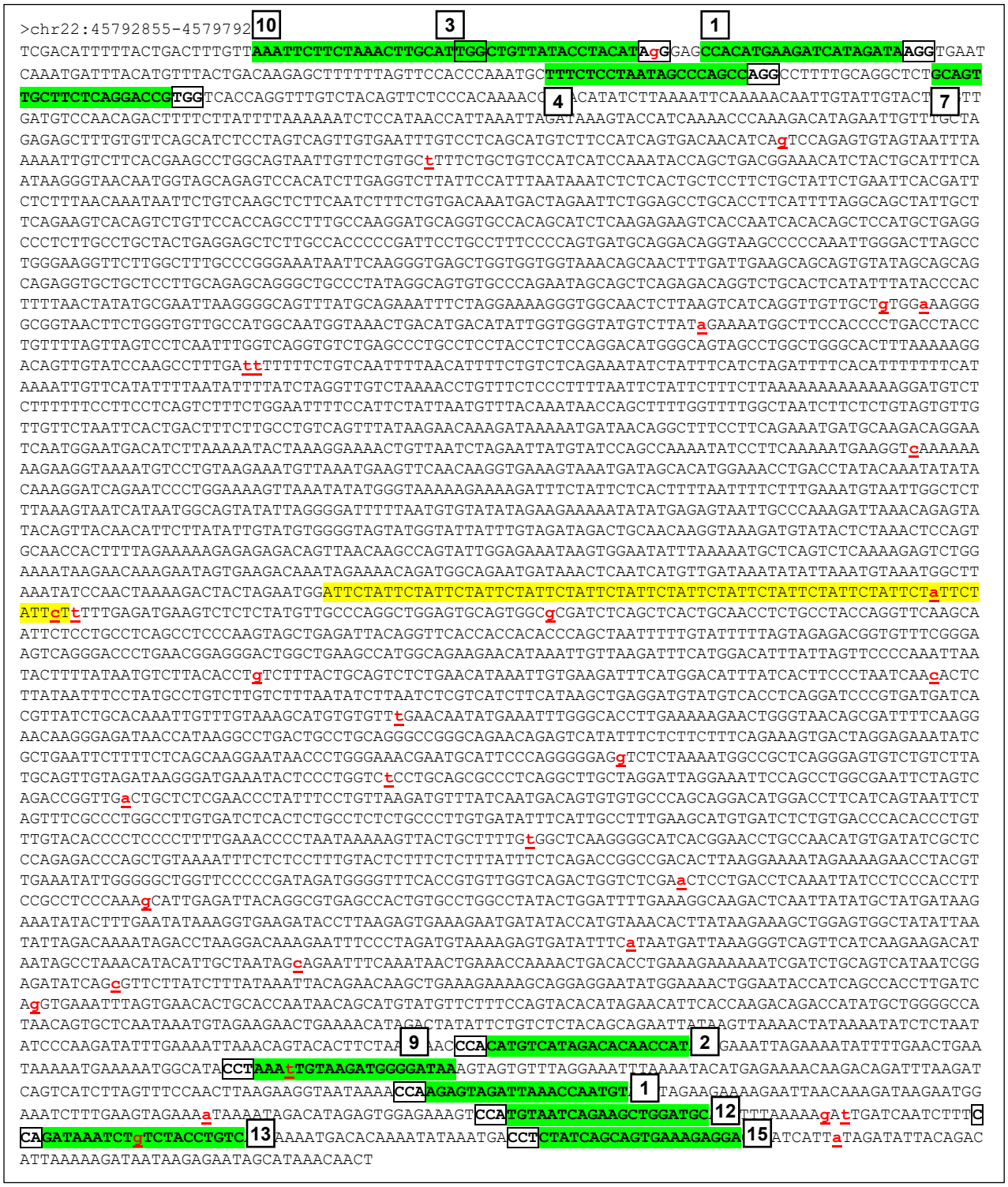
Orientation	sgRNA Cut Position (1-based)	sgRNA Sequence	PAM Sequence	Pick Order
sense	***85***	***CCACATGAAGATCATAGATA***	***AGG***	***1***
antisense	218	GGGAGAACTGTAGACAAACC	TGG	2
sense	***59***	***ATTGGCTGTTATACCTACAT***	***AGG***	***3***
sense	***165***	***TTTCTCCTAATAGCCCAGCC***	***AGG***	***4***
antisense	297	TTAAAATAAGAAAAGTCTGT	TGG	5
antisense	209	GTAGACAAACCTGGTGACCA	CGG	6
sense	***203***	***GCAGTTGCTTCTCAGGACCG***	***TGG***	***7***
antisense	74	CCTTATCTATGATCTTCATG	TGG	8
sense	***60***	***TTGGCTGTTATACCTACATA***	***GGG***	***9***
sense	***41***	***AAATTCTTCTAAACTTGCAT***	***TGG***	***10***
antisense	61	CTTCATGTGGCTCCCTATGT	AGG	11
antisense	167	AGCCTGCAAAAGGCCTGGCT	GGG	12
antisense	142	ATTAGGAGAAAGCATTTGGG	TGG	13
antisense	177	GCAACTGCAGAGCCTGCAAA	AGG	14
antisense	159	AAAGGCCTGGCTGGGCTATT	AGG	15

Table 2. *Reverse crRNA designs from the last 500 bp of the 5 kb ATXN10 sequence.*

Orientation	sgRNA Cut Position (1-based)	sgRNA Sequence	PAM Sequence	Pick Order
antisense	***224***	***ACATTGGTTTAATCTACTCT***	***TGG***	***1***
antisense	***40***	***ATGGTTGTGTCTATGACATG***	***TGG***	***2***
sense	119	CATACCTAAATTGTAAGATG	GGG	3
sense	270	AGAATTAACAAAGATAAGAA	TGG	4
sense	308	AAAATAAAATAGACATAGAG	TGG	5
sense	204	TCTTAGTTTCCAACTTAAGA	AGG	6
antisense	***39***	***TGGTTGTGTCTATGACATGT***	***GGG***	***7***
sense	118	GCATACCTAAATTGTAAGAT	GGG	8
antisense	***112***	***TTATCCCCATCTTACAATTT***	***AGG***	***9***
sense	117	GGCATACCTAAATTGTAAGA	TGG	10
sense	136	ATGGGGATAAAGTAGTGTTT	AGG	11
antisense	***327***	***GCATCCAGCTTCTGATTACA***	***TGG***	***12***
antisense	***376***	***GACAGGTAGACAGATTTATC***	***TGG***	***13***
sense	334	AAGTCCATGTAATCAGAAGC	TGG	14
antisense	***425***	***TCCTCTTTCACTGCTGATAG***	***AGG***	***15***

In addition to being located on the appropriate strand, designs are further selected by their uniqueness following the pick order (green shaded cells in Tables 1 and 2 above). For example, in Table 1, designs with pick orders 1, 3, 4 and 7 (Designs 1, 3, 4 and 7) all occupy unique 20-bp positions on the sense strand and are all highlighted as potential designs. However, Design 9 almost entirely overlaps with Design 3, therefore, preference is given to Design 3. Lower ranked designs with minimal overlap with higher ranked designs are considered as a potential design as is the case with Design 10 which overlaps with Design 3 by 2 bp.

4. Map the selected crRNA designs that are both properly oriented and unique to the 5 kb target sequence in order to determine the presence of common SNPs in crRNA and downstream PAM sequences.
   The sequence below shows the five forward designs highlighted in green in Table 1 and the 6 reverse designs highlighted in green in Table 2. PAM sites are enclosed within a box.



Part Number 101-839-600 Version 01 (August 2019)

5. Select a final set of designs for testing which do not contain common SNPs. Carefully inspect all designs for the presence of common SNPs in the crRNA sequence and downstream PAM sequence. Of the forward designs for ATXN10, Design 3 has a SNP in the PAM site. Of the reverse designs for ATXN10, Designs 9 and 13 have a single SNP in the crRNA sequence. All other designs present valid selections. Because not all crRNAs are equivalent in practice, choose several crRNA designs to be tested.
   We recommend testing the top 3 compatible designs as ranked by pick order from the forward (sense) and reverse (antisense) flanking orientations. For the ATXN10 target above, this would be Designs 1, 4 and 7 from the forward flanking design set and Designs 1, 2 and 12 from the reverse flanking design set.

Test crRNA Designs

It is advised to order and test several designs before choosing a final crRNA design set for a particular target. The crRNA designs can be tested with the No-Amp procedure or with a plasmid construct containing the flanking crRNA design sequences. This is a more efficient and economical testing model for a small number of targets as CRISPR-Cas9 digestion products are analyzed directly by agarose gel electrophoresis instead of sequencing.

We recommend ordering crRNA oligos from Integrated DNA Technologies (IDT) under the product offering “Alt-R® CRISPR-Cas9 crRNA”. The universal 67mer tracrRNA used in conjunction with crRNA to form gRNA is also ordered from IDT under the product offering “Alt-R® CRISPR-Cas9 tracrRNA”. These RNA oligos contain chemical modifications to protect from degradation by cellular RNases.

Revision History (Description)	Version	Date
Initial Release.	01	August 2019