sgRNA design for CRISPR-Cas9 mutagenesis

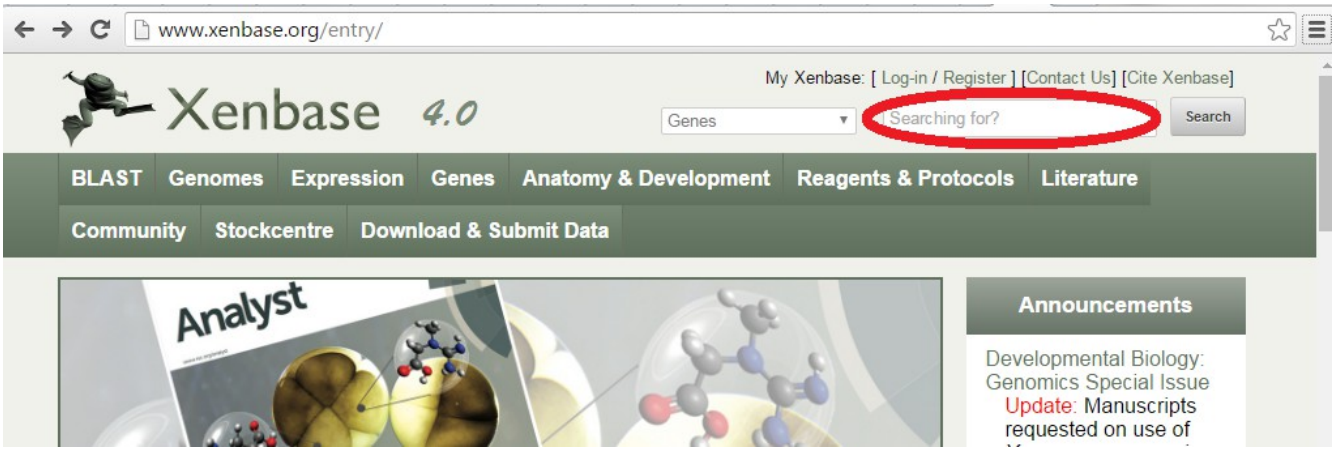
Once you have identified your gene of interest, you need to start working on designing an sgRNA construct that you will use to mutate your gene of interest.
Before you start, make sure you install ApE plasmid editor on to your computer.
http://biologylabs.utah.edu/jorgensen/wayned/ape/

Here is an abbreviated step-by-step design process followed by more in depth explanations

1) Install ApE on your computer.
2) Type in your gene name in Xenbase and search.
3) Click on GBrowse 9.1
4) Left-click on JGI primary transcript and choose Genome Details.
5) Download FASTA file of your gene.
6) Open FASTA file in ApE.
7) Highlight all the sections of the sequence in the ApE file.
8) Either
   a)first search for CCNNNNNNNNNNNNNNNNCC (CCN18CC) sequence within exon sequences in ApE file (check "also find rev-com of string")
   OR
   b) first Go to CRISPRdirect, put in mRNA sequence (without introns) for your gene of interest and choose all sgRNA sequences that are CCN18CC or GGN18GG, and ideally choose sgRNA sequences that target both genes on long and short chromosomes but have the fewest off-target sites.
9) Among the potential candidates, eliminate according to inclusion and exclusion criteria.

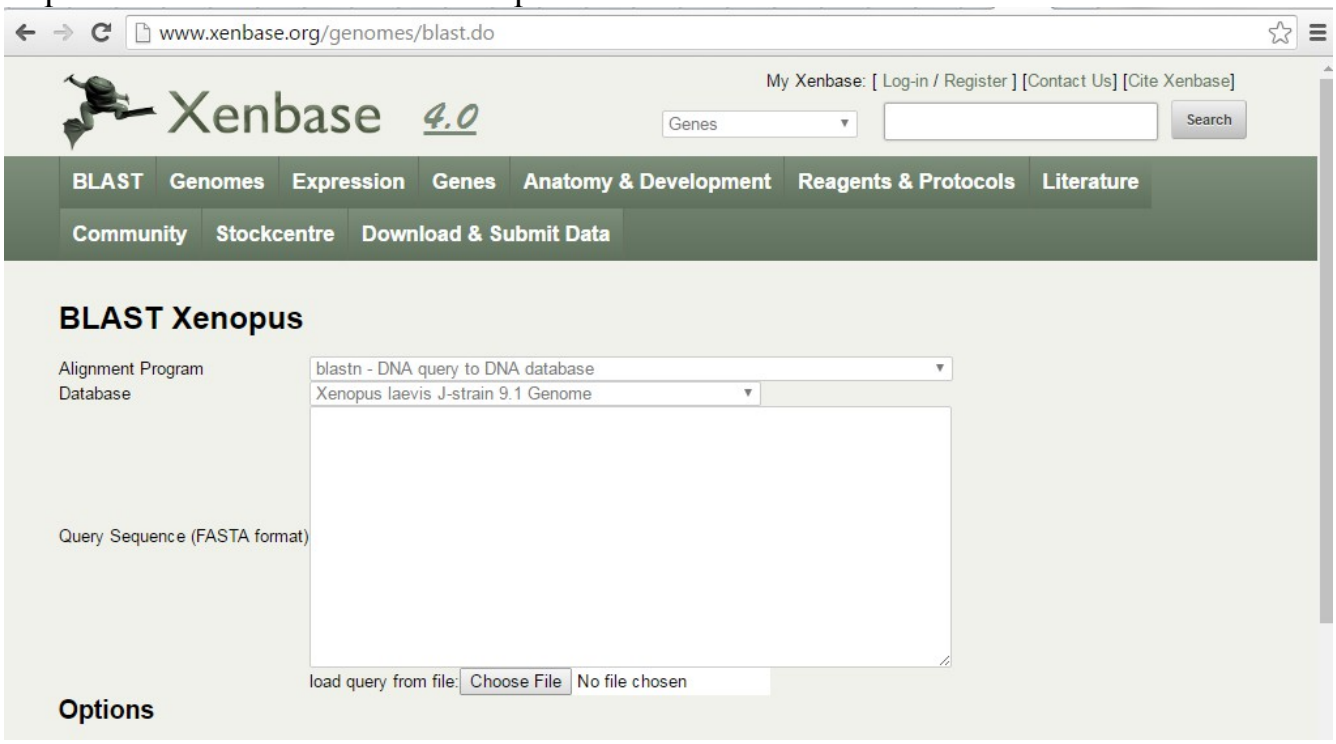
   Once you have identified a good candidate, you will start designing your oligos for cloning your sgRNA construct. This will be discussed in the next section of the manual.

1) To install A plasmid Editor (ApE), go to
   http://biologylabs.utah.edu/jorgensen/wayned/ape/
   and install the version that best fits your operating system.
   Then "Run" ApE software.
2) To make sure you can find your gene of interest in the *Xenopus* genome using Xenbase www.xenbase.org
   If possible do not use Internet Explorer as your browser when using Xenbase.
   You may simply type in your gene name in the red oval below OR

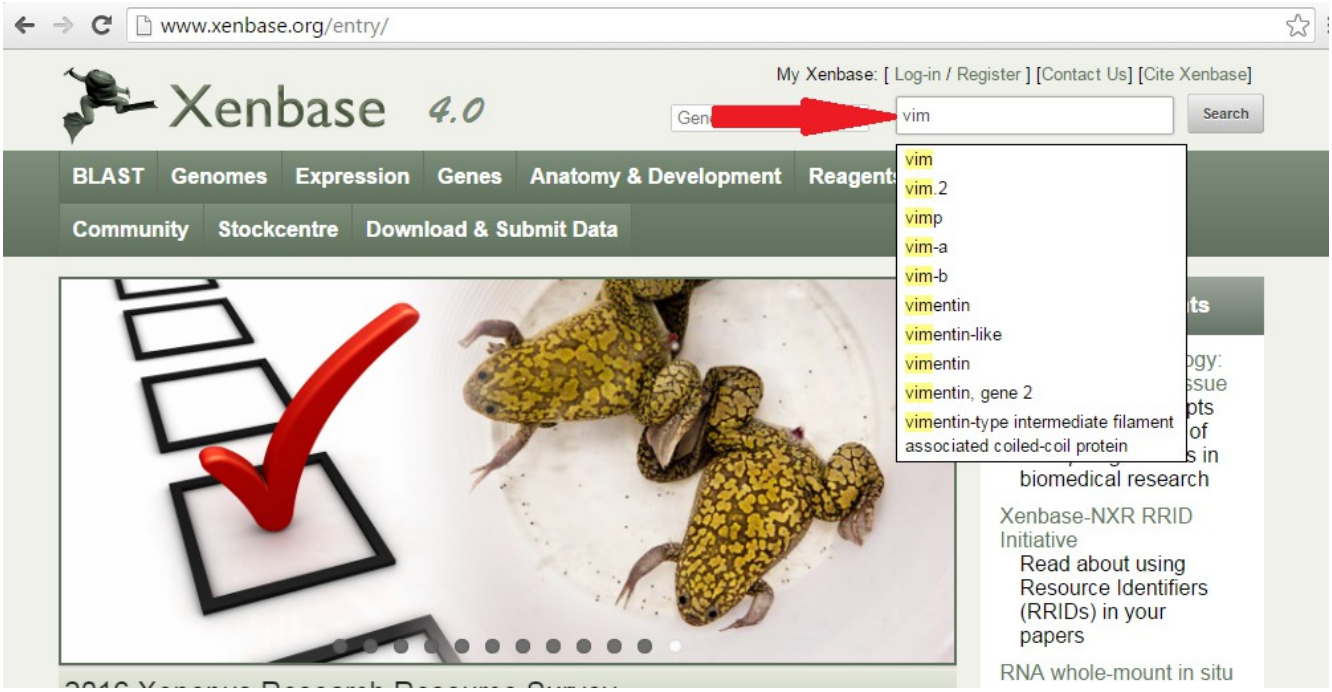


If you were unable to find a gene name but have a DNA sequence, you may BLAST the sequence to find identical or similar sequences in the *Xenopus* genome using the BLAST option (yellow oval below).

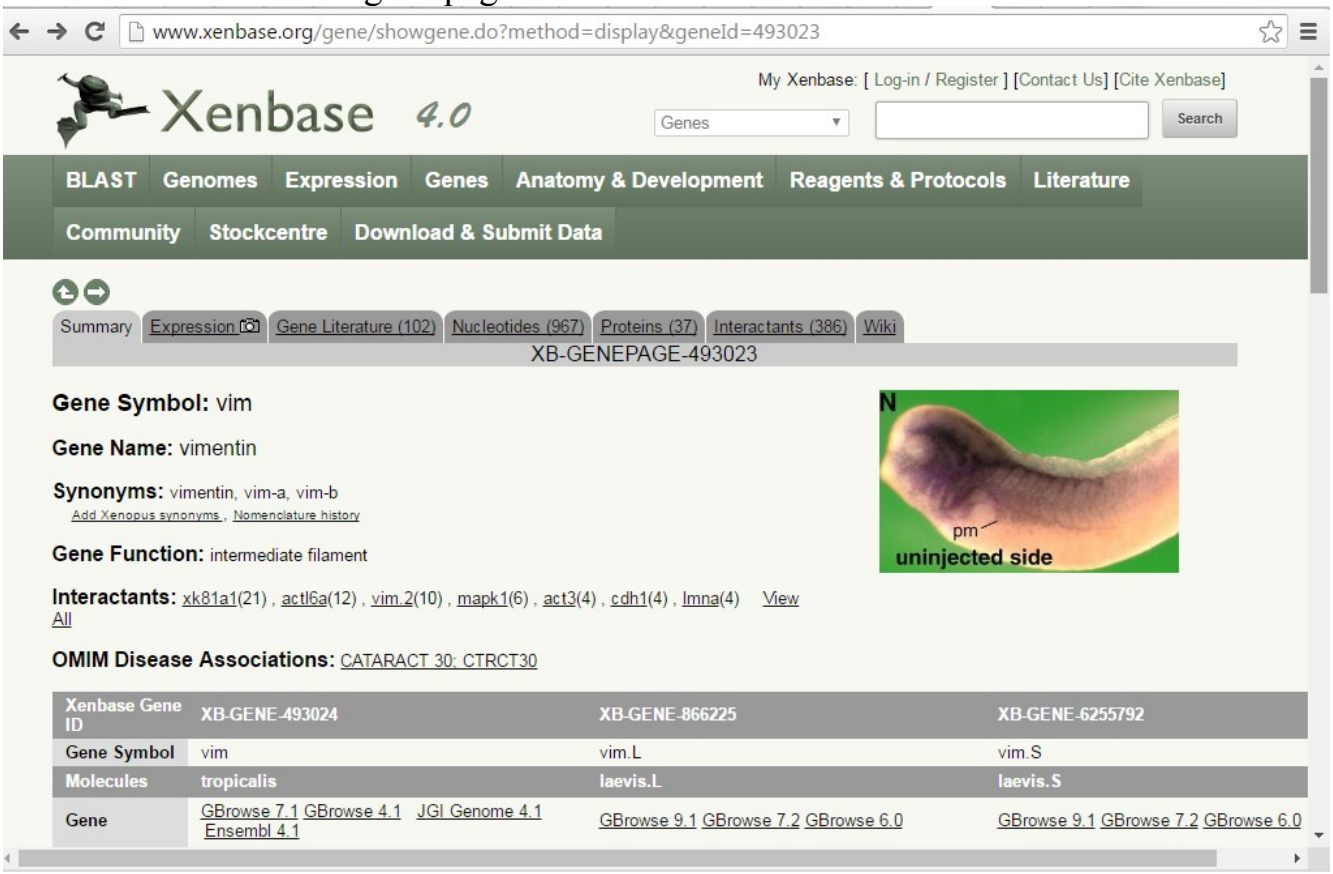
Here in BLAST, you may use different types of searches such as DNA or protein sequences to search for similar sequences.



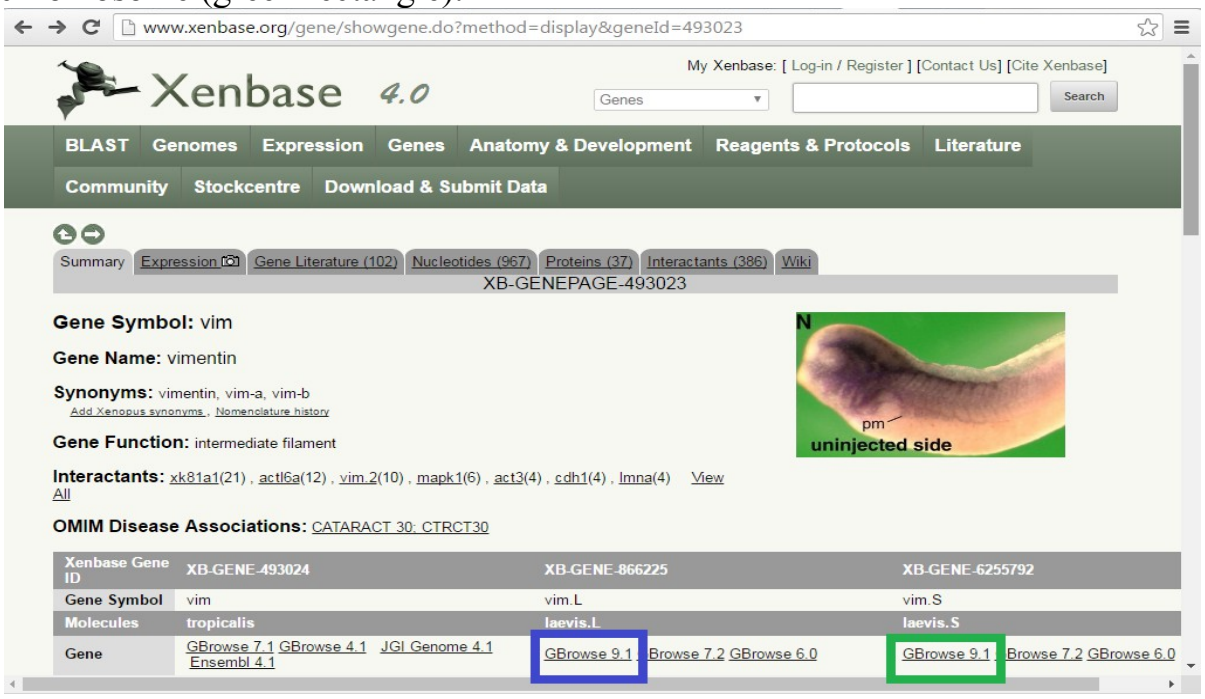
Let's say you have already identified a gene name and you decide to reach Xenbase information using your gene name.
I will give you an example with my gene of interest, *vimentin*.
As I begin to type the gene name *vimentin*, Xenbase starts giving me some suggestions already (see red arrow). *vimentin* is abbreviated as *vim*.



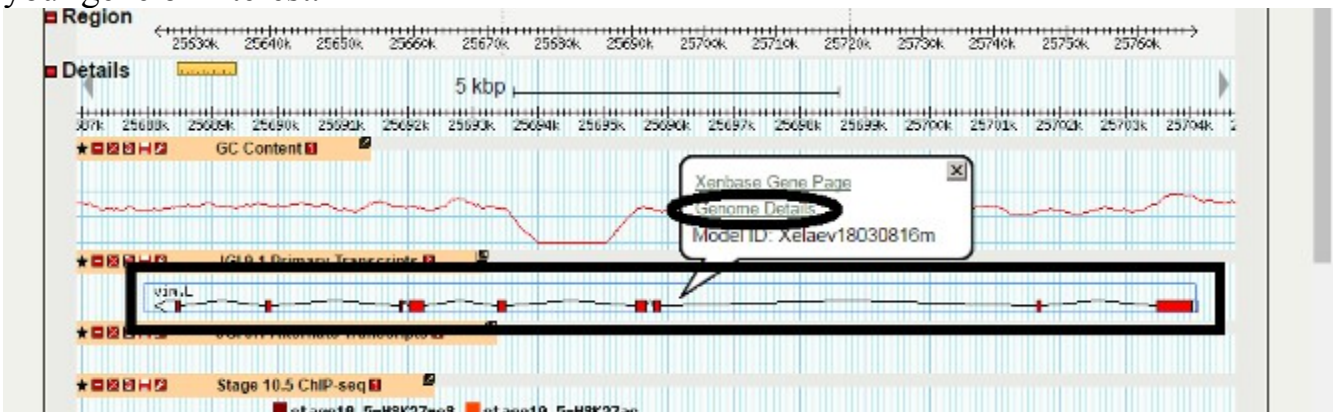
Below is what *vimentin* gene page looks like as of 8/30/2016.



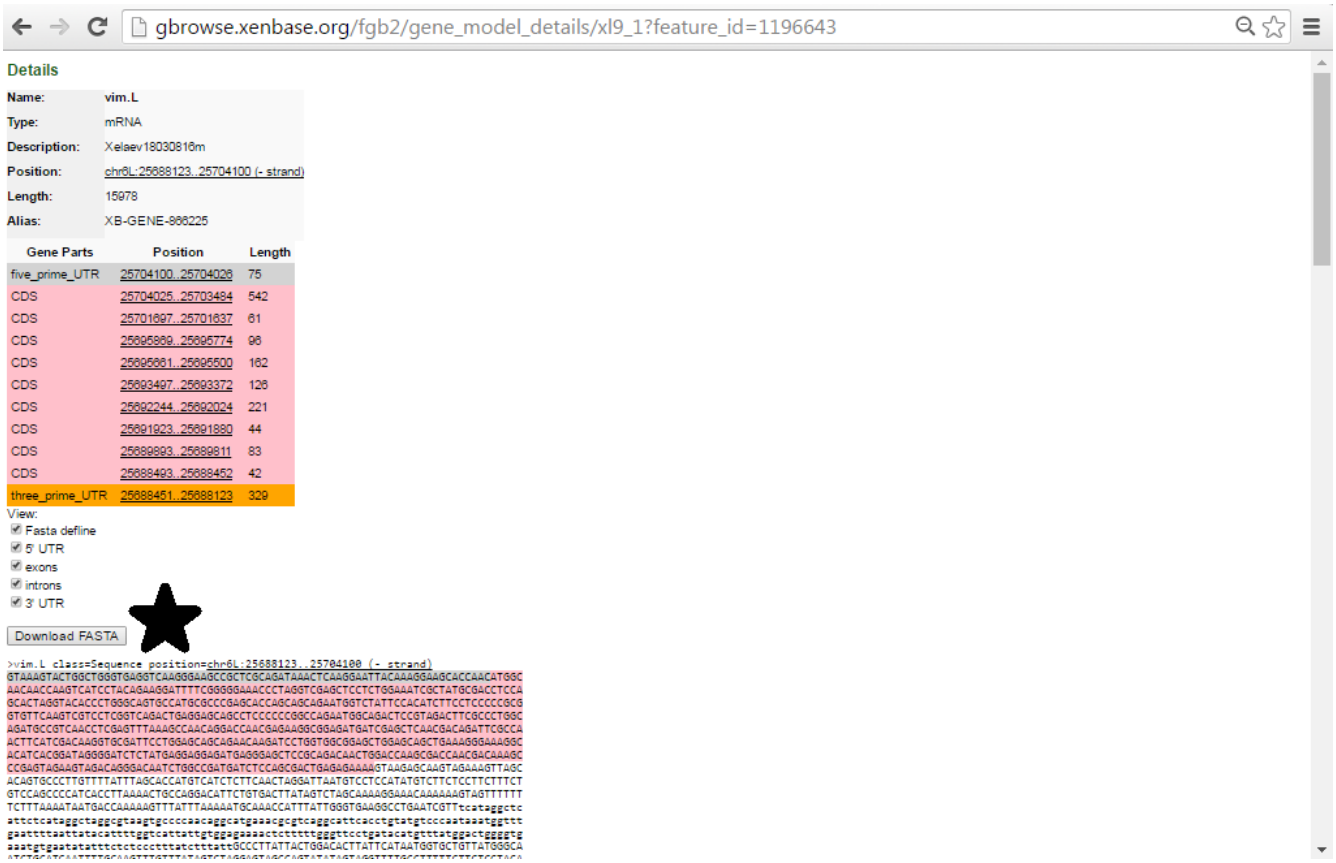
3) Click on GBrowse 9.1, blue rectangle first for *Xenopus* genes on the Long (L) chromosomes. Later we will also compare this sequence to the one on the Short (S) chromosome (green rectangle).



4) Once on the GBrowse page, scroll down to find the JGI primary transcript section (black rectangle and left-click on that RNA structure that displays introns and exons from the primary (most abundant RNA transcript) NOTE: Some browsers such as Internet Explorer may not let you click on this. Use Chrome or Firefox).
Now choose Genome Details (black oval) to reach the gene structure information for your gene of interest.



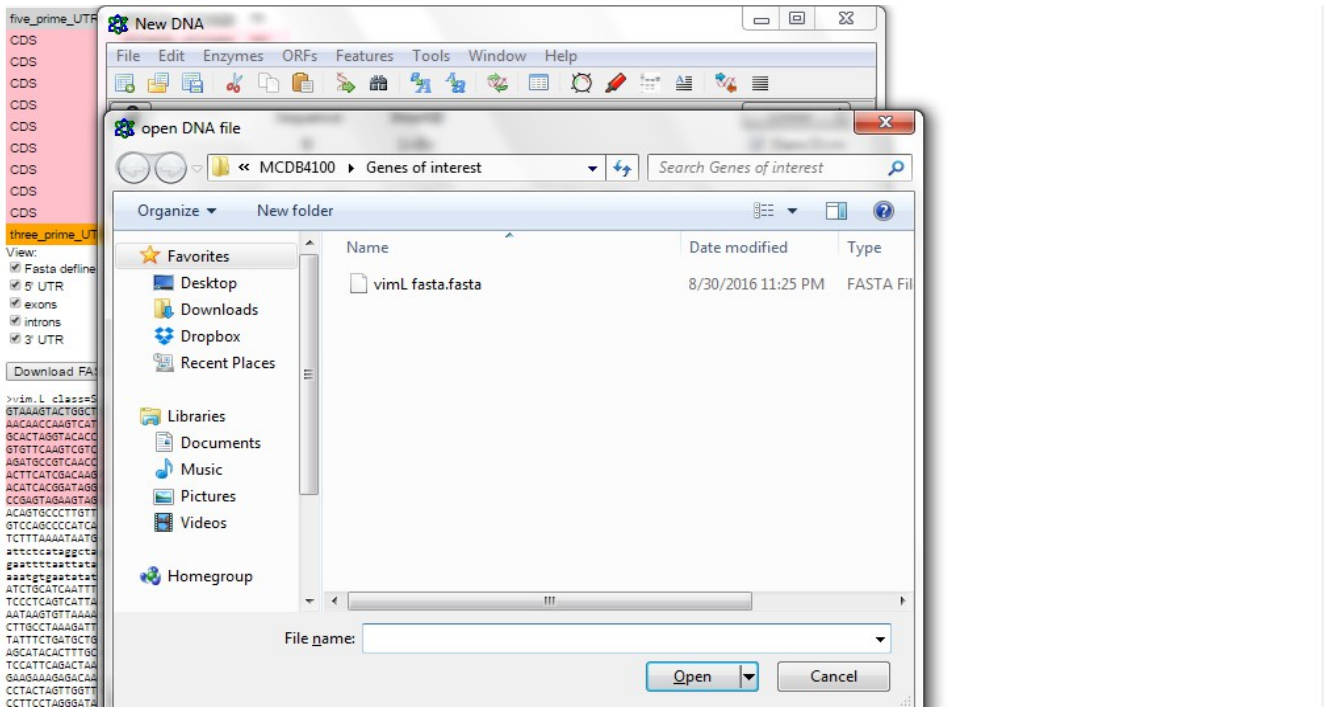
Genome Details will display 5' UTR highlighted in grey, Exons highlighted in pink,



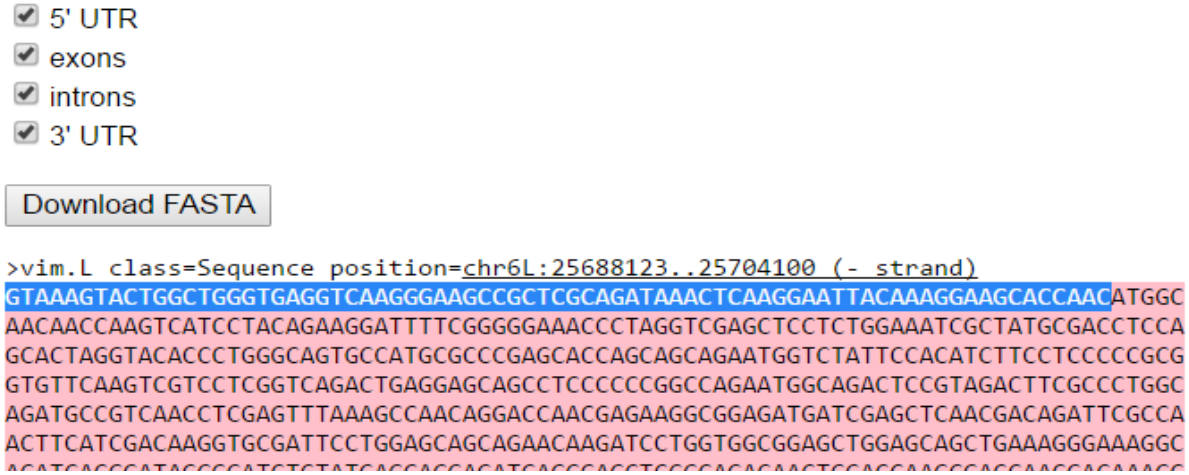
Introns in white (not highlighted) and 3' UTR highlighted in orange.

5) Click on “Download FASTA” button (next to black star) and save the file with your

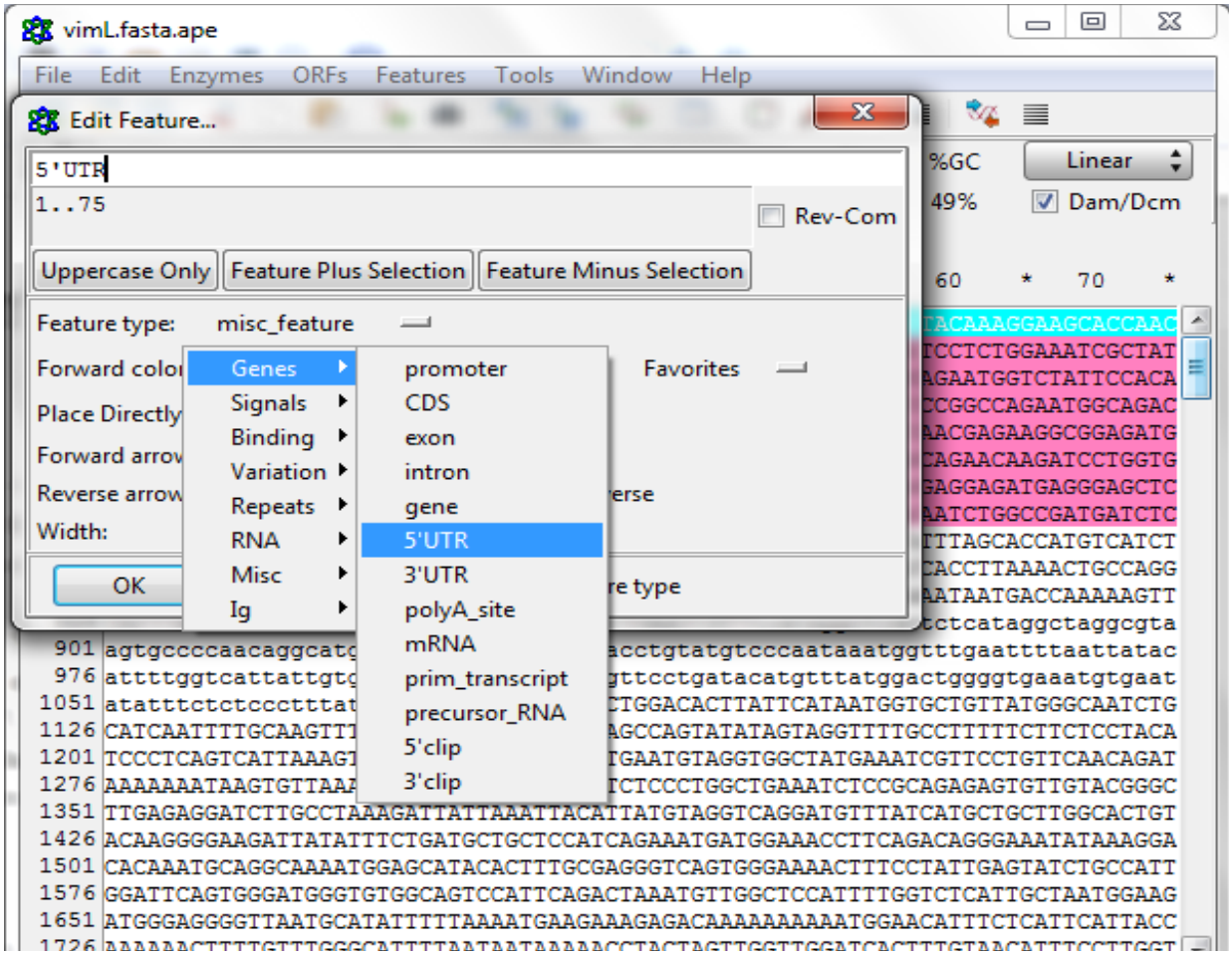
gene sequence name (e.g. vimL).
You can now open this FASTA file in ApE.
6) Run ApE software and under “File”, choose “Open”, select your FASTA file and open it.



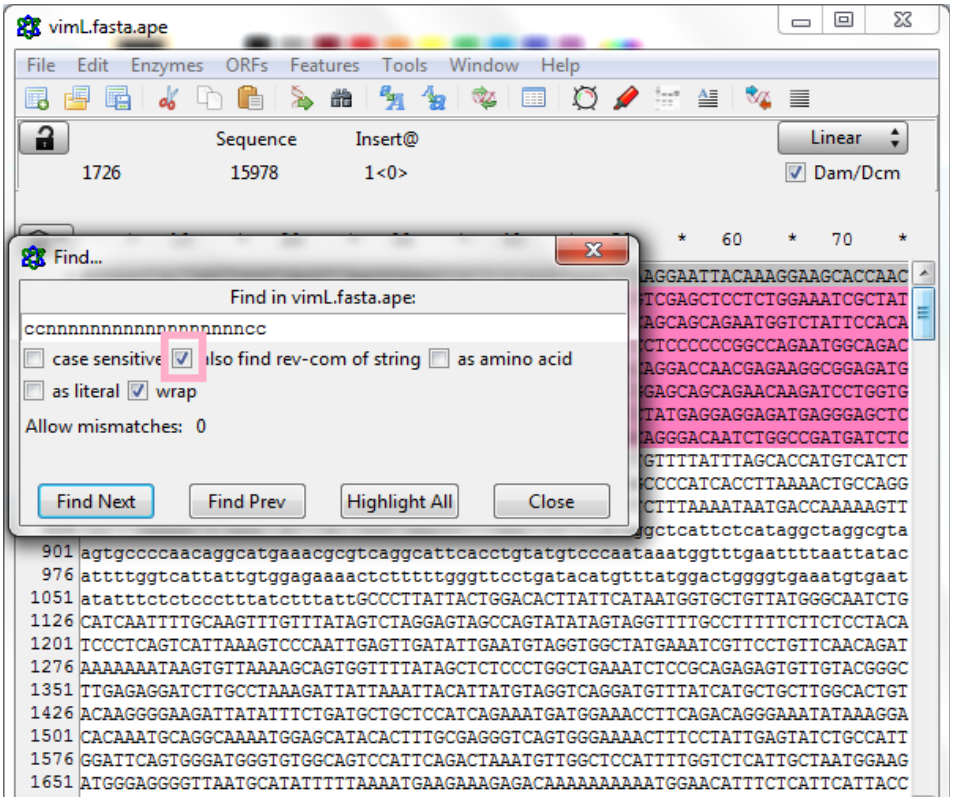
7) Unfortunately, this highlighted information does not transfer to ApE; therefore, we will have to do it manually.
One by one, highlight the 5'UTR, exons and the 3'UTR in the ApE file (leave the introns un-highlighted).
In order to highlight each section, begin with the 5'UTR.
Select and copy (CTRL-S) the 5'UTR sequence on Xenbase Genome Details page.



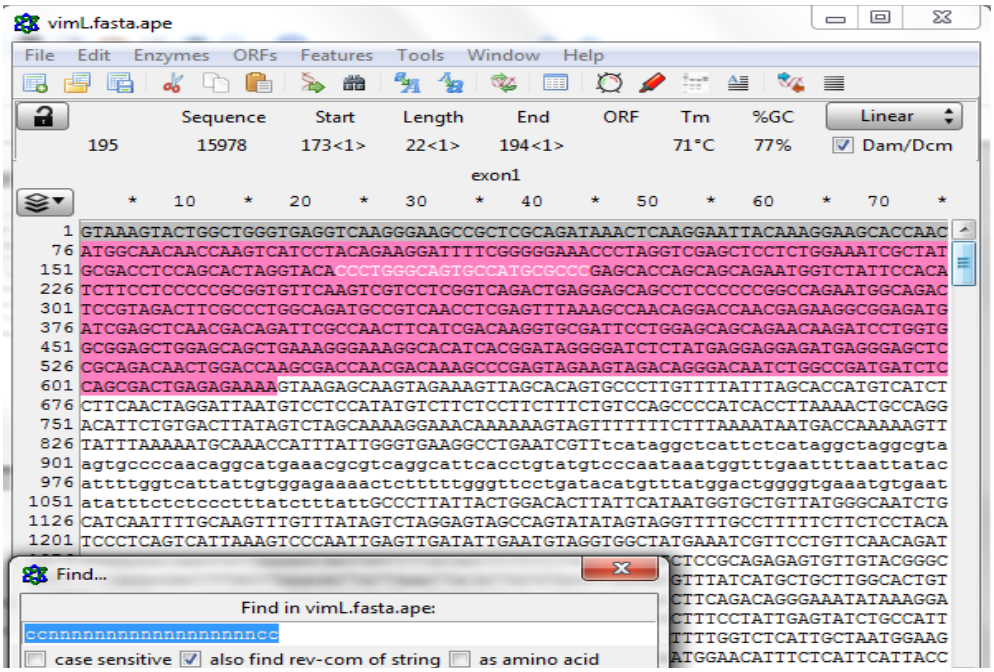
Go to ApE and click CTRL-F (Edit>Find).
Paste (CTRL-V) this sequence into the Find tab.
Click ”Find next” button and see that the ApE file now has highlighted the 5'UTR sequence.
Now select the sequence, and go to “Features” tab and click on “New Feature”.
Here, name the Feature as “5'UTR” and select Feature Type>Genes>exon.
Choose a “forward color” for this 5'UTR (you may choose gray to stay consistent with Xenbase or another color; but you should have a unique color for sections of a gene and these colors should be consistent across different gene files that you work on)



8a) Once all sections are highlighted you may search for potential sgRNA sites by searching for CCN18CC sequences in your file.
Using Find option (CTRL-F or Edit>Find), search for
CCNNNNNNNNNNNNNNNNCC
Remember to check the “also find the rev-com of string” (pink square)

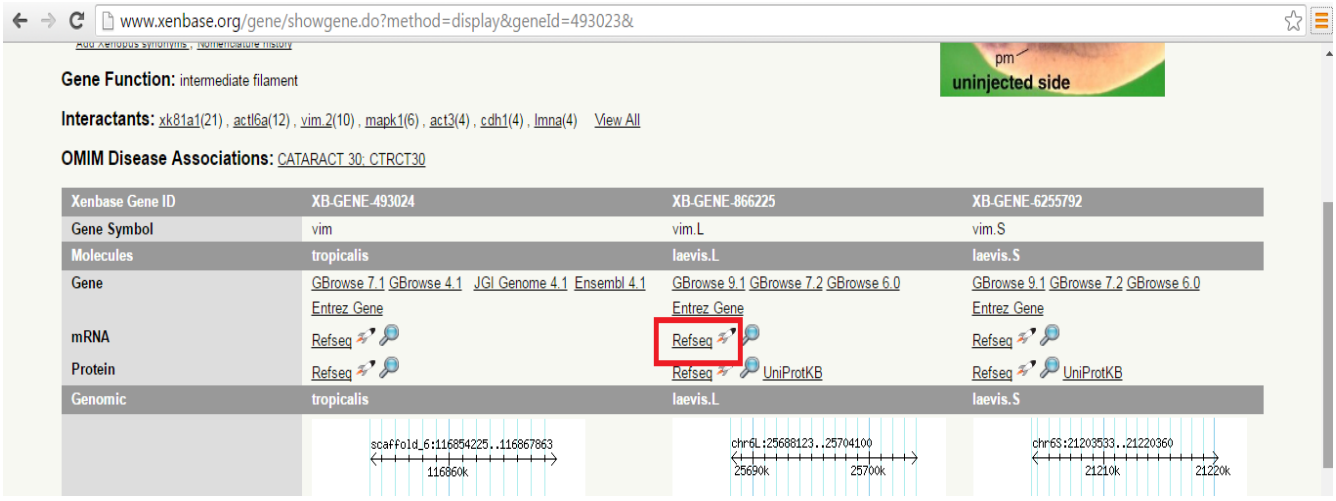


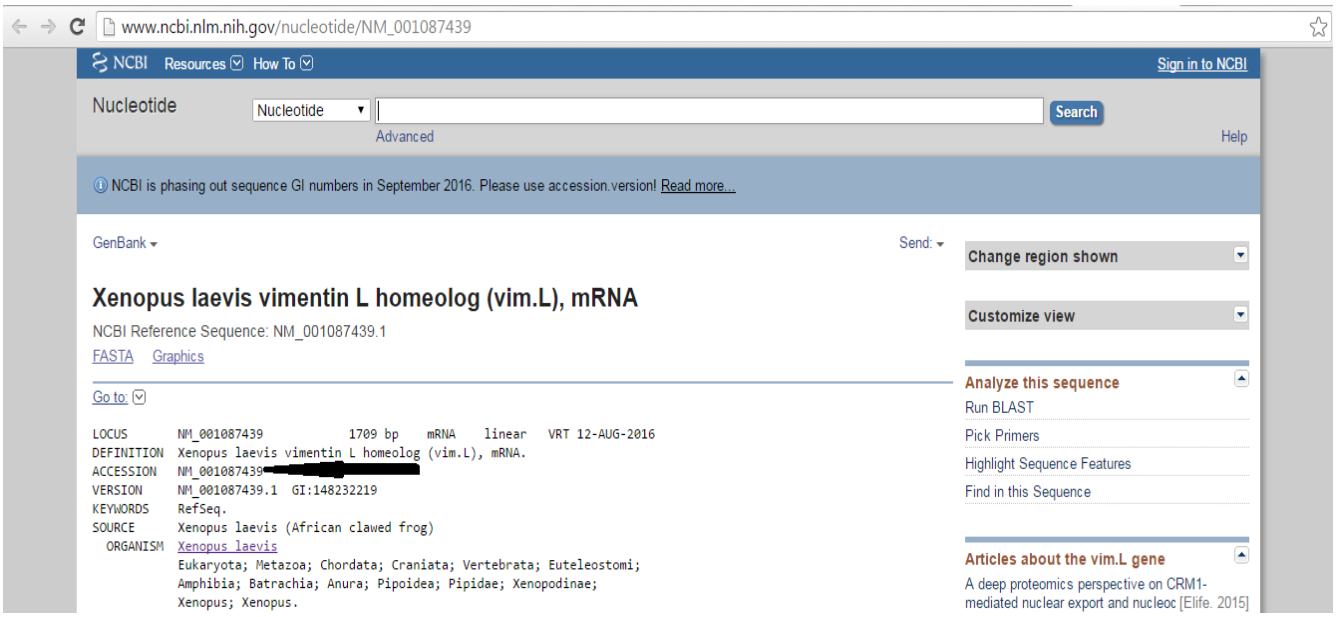
Every time to find CCN18CC sequence within an exon sequence, highlight it with a color (distinct from the rest of the gene sections but keep the color consistent for all the sgRNAs). Don't add these in as new features because some of these sequences may not meet all the inclusion and exclusion criteria and may be eliminated.



8b) Before or after you have searched for CCN18CC sequences in your ApE file, you must go to CRISPRdirect webpage to determine specificity of sgRNA sequences that can be designed within your gene sequence.
http://crispr.dbcls.jp/

You need to put in mature (spliced) RNA sequence into CRISPRdirect since most pre-RNA with intronic sequences are too long for CRISPRdirect to analyze. Plus, putting in mature RNA allows you to pick sgRNAs within exons and exclude intronic target sites. You can find mRNA sequence in Xenbase listed as mRNA reference sequence (RefSeq, red rectangle) that links you to NCBI. You may click on the magnifying glass symbol to view the accession number and gene sequence or you may click on RefSeq to connect to NCBI.



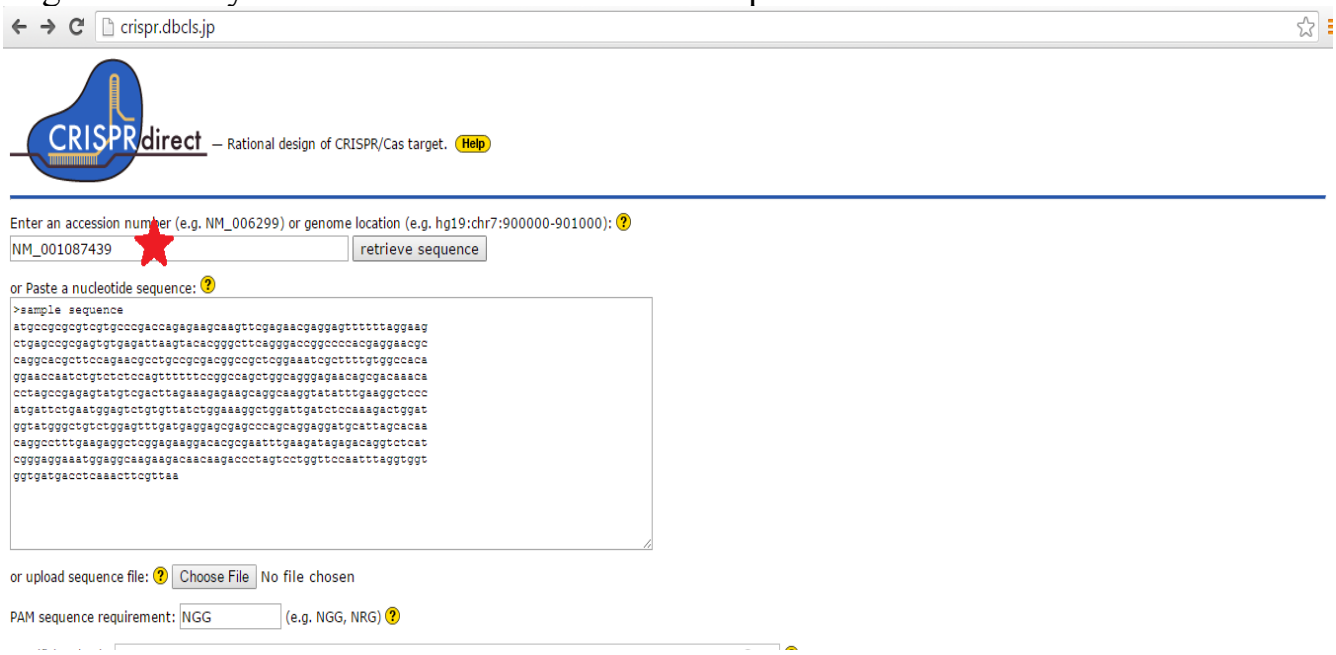


Find the gene accession number (black arrow) and copy it.

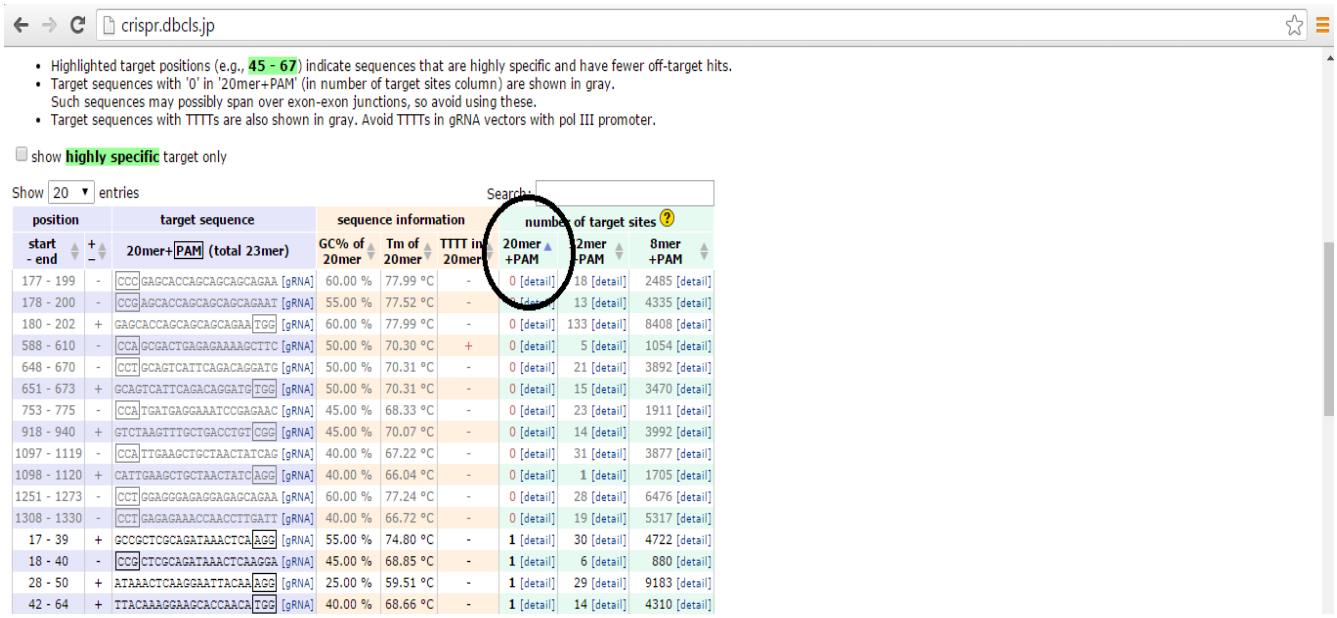
On CRISPRdirect webpage, paste the accession number into the “Enter an accession number” line (red star) and click “retrieve sequence”
Then in the specificity check tab, select “Frog (*Xenopus laevis*)”, and click the “design” button. design

DO NOT choose Frog (*Xenopus tropicalis*) by mistake.

NOTE: Remember the RefSeq sequence links to the most abundant splice isoform; therefore, you may need to check for alternative splice forms to find shared exons to target. This may be done in Xenbase under the Expression section.



You will need to scroll down to see the list of the candidate sgRNA sequences.
Click on “20mer+PAM” (black circle) to order the sgRNA candidates according to target site count.
We don't want zero counts (no target within sequence you provided).
We prefer to choose two count candidates since most if not all of the two count candidates will be targeting the two genes on long and short chromosomes that we want to target.



Once you have the candidates in order, look at the ones that target two genes and see if you can find candidates that have CC on both ends of the sequence (GG on both ends of the sequence would work, too).

Copy each of these sequences either into an excel sheet or directly into ApE (see description in 8a where you use “Find” and highlight options.
At this point go to Step 3 and Click on GBrowse (green rectangle) to get the gene sequence from the short chromosome and follow the same procedure upto Step 9. If possible pick sgRNA candidates that are identical for long and short chromosome genes.

Now, make sure your choice meets inclusion and exclusion criteria.

9) Inclusion and exclusion criteria for target sequence
   Sequence must be CCN18CC or GGN18GG
   Sequence ideally is identical for both long and short chromosome genes. If identical target sequence cannot be found, we will need to pick two separate sgRNAs from the “1” target site group (black circle above) to target both genes individually and inject them together into frog embryos.
   Sequence ideally has few off-target sites (as 12+ or 8+mer sequences).
   Sequence ideally is within an early exon.
   Sequence ideally has high GC content (>45% to 85%) and high Tm.
   Exclusion criteria:
   Sequence does not target across exon-intron junctions or within introns or within UTRs.
   Sequence must not have 100% identity to any part of the human genome. You will need to blast your candidates into the human genome on NCBI BLAST link and

eliminate all candidates that show 100% identity to a sequence in the human genome.
Sequence should not have an internal DraI restriction enzyme site. To check for DraI sites, in ApE, go to Enzymes>Enzyme Selector, then click on DraI to highlight it (blue arrow) and then click “Graphic map” (green oval). Do not choose a sequence that can be digested by DraI enzyme (TTT^AAA).