CRISPR sgRNA selection

PART I: Know your gene!
In order to successfully design sgRNAs for CRISPR KO, you first need to make sure you have all the information necessary to make the right choices. This includes identifying the number of splice variants/transcripts, identifying potential start codons in the beginning of the CDS as well as knowing which parts of the ORF are exons. Finding this information can be done using Ensembl;

1. Go to the Ensembl website (www.ensembl.org).

2. In the box to the top right, type in your gene of interest (for example, type in ITGAV for evaluating the Integrin alpha V gene).

3. When the search results have been returned, select your desired organism in the “Restrict species to” list on the left hand side, to the left of the results list (for example, for mouse ITGAV click “Mouse” in the list).

4. Now, click on the gene link in the search results list (for example, for mouse ITGAV, click on “Itgav (mouse gene)” which is first in the results list).

5. Now you’re at the page for your selected gene. Have a look at the transcript table (if you can’t see it, click the “show transcript table” button). In the transcript table, you can see all the known transcripts of your gene of interest, what size of protein it produces (if it does produce protein) and more. The “main” protein isoform(s) of your gene of interest can usually be found at the top of the list, and this (or these) transcript(s) have additional information listed in the table such as RefSeq links.

6. Click the “main” protein isoform(s). This will take you to a page describing the transcript. Next to the “About this transcript” menu there will be information regarding the number of exons and variants within this transcript (for example, for the mouse itgav-001 transcript, the text will be “This transcript has 30 exons, is annotated with 41 domains and features, is associated with 32 variations and maps to 50 oligo probes”). Click the “xx exons” link to access exon and intron sequences for the transcript.

7. Scroll down to see the sequences for the transcript. Introns and exons are clearly marked, and the protein-coding regions are easily identifiable as these sequences are shown in blue capital letters. If your gene has more than one transcript to take into account, have a look to determine if they are rather similar (i.e. have approximately the same start site and comparable protein size) or if they are very different. If the transcripts are different, you may not be able to knock them out simultaneously using only one sgRNA, instead you may have to design different sgRNAs to efficiently target them both (this applies to situations where for example the sgRNA cuts very early in the gene in transcript #1 but in transcript #2 the cut occurs in the end or middle of the gene if using the sgRNA).
PART II: Identify sgRNA candidates!

1. Go to the CRISPRscan website (www.crisprscan.org).

2. In the upper left corner, click “By Gene”. Once the search page has loaded, select the organism you are working in in the first box. In the second box select “Gene”. Finally, type in your gene of interest in the text box (for example, for mouse ITGAV, select “Mouse – mus musculus” in the first box, “Gene” in the second box, and type in “ITGAV” in the text box).

3. Hit the “Get sgRNAs” button. After a short moment a list of sgRNA candidates will appear. They are listed according to scoring, i.e. high-scoring sgRNAs that are deemed to have high efficiency are at the top of the list. On the far right, all of the protein-coding transcripts of your gene of interest are listed in columns. For each sgRNA there is a numeric value listed in the transcript columns showing where the sgRNA will cut – if there is no number this means that the sgRNA does not cut this specific transcript. The columns titled “Cong” and “Hsu” contain information on predicted off-target hits in the genome of your selected organism.

4. Now it’s time to sort the sgRNA candidates. Copy the first 10 to 20 hits in the results list, and paste into an Excel file.

5. If you have very many hits with a score of 55 or higher, you can sort out all the hits that have numbers in the “Cong” and “Hsu” columns, but you want to have at least 10 different sgRNAs left in your list for the next step. The higher the score, the more stable the sgRNA and thus also the activity of the sgRNA. However, the score does not reflect other suitability factors such as the amount of off-target hits, or where in the gene the sgRNA cuts.

6. Now, have a look at the transcript information to the far right. Identify the “main” transcript(s) and have a look at the numbers in that column. You’re looking for sgRNAs with low numbers – ideally ones that will cut approximately 90-150nt (cutting “later” in the gene can also be ok but it’s better to be as close to the start as possible), after the ATG start codon (or where the blue capitalized text starts in the exon view on Ensembl).

7. Select 2 different sgRNAs that fulfill the criteria (i.e., sgRNAs that have a score of at least 55, few predicted off-target hits, and that cut early in the CDS). The actual sgRNA sequence is marked in capital letters (If it isn’t possible to find sgRNAs that match the criteria using CRISPRscan Gene search, you will have to search for suitable sgRNAs using sequence input on CRISPRscan, see Part III for further instructions).
PART III: Finding sgRNA candidates using sequence input

If you’re having a hard time finding good sgRNA candidates using CRISPRscan’s gene search, you could try using the sequence input search function in CRISPRscan.

1. Look at your gene of interest in Ensembl. Have a look at the transcript table again and select the “main” protein isoform(s) of your gene of interest. If there is more than one, have a look at the start sites (blue capital text in the exon view page) of each of the transcripts to see if it would be possible to find a common sequence for these in the beginning of the CDS, that allows for simultaneous targeting of these transcripts.

2. Copy the sequence of the starting exon (this is not always exon #1, look for the exon where the blue capital letters begin) and paste it into the sequence input box on CRISPRscan (press the “By sequence” button at the top of the CRISPRscan page). If this first exon is very short (i.e., less than 100bp), copy and paste the sequence of the next exon instead. Make sure you have selected the correct species, then press “Get sgRNAs”. It may also be a good idea to test the next exon as well, especially if the starting exon is rather short, or if your selected sequence doesn’t return any suitable sgRNAs (while you have more control of where cutting occurs by choosing a specific exon to search for sgRNAs in, you may find that not all exons have characteristics that allow for efficient sgRNA targeting).

3. Select 2 different sgRNAs that fulfill the criteria (i.e., sgRNAs that have a score of at least 55 and few predicted off-target hits). The actual sgRNA sequence is marked in capital letters.

PART IV: Primer design!

Now that you know the sequence of your desired sgRNAs, you must convert them into primers that can be ordered, annealed and ligated into the vector. Below is an example to show how to this:

To find the REV seq, identify the complimentary sequence to your sgRNA.
FWD seq: GGC TGG TGA TGA TGC TCC CA
The complimentary seq: CCG ACC ACT ACT ACG AGG GT

Next, reverse the complimentary sequence to get correct orientation.
Complimentary seq before reversing: CCG ACC ACT ACT ACG AGG GT
Complimentary seq after reversing: TG GGA GCA TCA TCA CCA GCC

Finally, add the flanking overhangs to construct primers that allow for ligation into the vector (marked in red below).
FWD primer: CACCCGGCTGGTGATGATGCTCCCA
REV primer: AAACTGGGAGCATCATCACCAGCCC