<u>**Dumont Lab Short Protocol: Rat Kangaroo Transcriptome**</u>
<u>**Updated July 2020**</u>

Rat Kangaroo Transcriptome: http://sluganalysis.ucsc.edu/#home
Username: sophie.dumont@ucsf.edu
Password : spindle2012
Transcriptome Explorer: http://sluganalysis.ucsc.edu/ratkangaroo/

Shortcut Method: Searching for Rat Kangaroo gene of interest:
1. Open the Transcriptome Explorer and search the official human gene symbol for your protein as follows: ProteinName*
   a. Eg, NDC80*, NuMA*, PRC1*, AURKB*, SMC1A*
2. Select the Trinity gene that is obvious to you which one is correct and skip to step 9.
   a. If it is not obvious to you which Trinity Gen to select, then proceed to step 3 below.

Full Method: Searching for Rat Kangaroo gene of interest:
3. Launch the Rat Kangaroo Transcriptome, and open the genome browser. Select "Tools" on the top left and select the "BLAT" tool.
4. Search for the human version of your gene of interest using the NCBI gene database.
5. Copy the FASTA protein sequence of your gene of interest (check for correct isoform).

mRNA and Protein(s)

1. NM_006101.2 → NP_006092.1 kinetochore protein NDC80 homolog

mRNA sequence   Protein sequence

6. Paste the protein sequence into the BLAT tool's search engine. This produces a series of hits with different scores and a comp number.
   a. The comp number corresponds to a unique transcript sequence on the Transcriptome Explorer.
7. Copy the comp number for the top hit(s).

ACTIONS	QUERY	SCORE	START	END	QSIZE	IDENTITY	CHRO	STRAND	START	END	SPAN
browser details	YourSeq	877	1	641	642	73.0%	comp61793_c0_seq19	+-	1332	3269	1938
browser details	YourSeq	874	12	641	642	73.3%	comp61793_c0_seq17	+-	1332	3236	1905

8. Open the Transcriptome Explorer and paste your comp number into the search bar.
9. Select the Trinity Gene and select the Trinity Isoform with the highest FPKM number.
   a. The higher the FPKM number, the more of those transcripts were sequenced and are presumably expressed in cells. (high FPKM ~ high copy number in cells)
   b. You may want to consider multiple isoforms with high FPKM when designing oligos for siRNA.
10. Scroll down to the transcript sequence to see the rat kangaroo sequence of your gene of interest.
11. Scroll further down and **record** the strand orientation.

Protein Detail

Name: m.205522   Length: 2057   Location: 1329-3386   Strand: (-)

12. To check if you have the correct gene, select the NCBI Blast link below your transcript sequence and use "mammals" for the organism parameter. Select BLAST when you've inputted your desired parameters.

13. Finding the protein coding region and correct Methionine start site in your gene of interest:

Method 1:

a. From the NCBI Blast search, select the gene that aligned best with your sequence.

b. If the query sequence (gene of interest) starts at number (X) other than 1, then this suggests that there are (X-1) amino acids in the rat kangaroo sequence that are not part of the protein coding sequence. The actual start site is at amino acid number X.

c. From the copied transcriptome nucleotide sequence, delete the corresponding nucleotide sequence preceding amino acid number (X).

d. If your strand orientation [step 9] is (-), then you must first reverse-complement the highlighted yellow transcriptome sequence before deleting the nucleotide sequence preceding amino acid number (X).

Method 2:

e. Copy the yellow highlighted nucleotide sequence from the transcriptome and paste it into this online translate tool, select translate. The highlighted sequence is the correct frame.

f. Similar to Method 1 above, delete the nucleotide sequence that precedes the first methionine site. Take into consideration strand orientation, as described in 11d.

   i. Note: The first Methionine start site might not necessarily be the correct one. You can check it by BLAST.