



## Please reference

Flies AS, Lyons AB, Corcoran LM, Papenfuss AT, Murphy JM, Knowles GW, Woods GM, Hayball JD. [\*PD-L1 Is Not Constitutively Expressed on Tasmanian Devil Facial Tumor Cells but Is Strongly Upregulated in Response to IFN- \$\gamma\$  and Can Be Expressed in the Tumor Microenvironment\*](#). Front Immunol (2016) 7:581. doi:10.3389/fimmu.2016.00581

Flies AS, Blackburn NB, Lyons AB, Hayball JD, Woods GM. [\*Comparative analysis of immune checkpoint molecules and their potential role in the transmissible tasmanian devil facial tumor disease\*](#). Front Immunol (2017) 8:513. doi:10.3389/fimmu.2017.00513

## Materials

Computer

## Software

SnapGene <http://www.snapgene.com/>

CLC sequence viewer (free version) <https://www.qiagenbioinformatics.com/products/clc-sequence-viewer/>

Webcutter <http://rna.lundberg.gu.se/cutter2/>

Tool for finding potential restriction sites in a DNA sequence by introducing silent mutations that do not change the protein sequence. This is used to introduce restriction sites in the plasmid during construction that can then be used for downstream modifications of the plasmid.

Russel Lab <http://www.russelllab.org/aas/>

If you need to change an amino acid to introduce a restriction site, then check the amino acid properties and consequences of substitutions to see which amino acid is the best substitution (Gray & Barnes, 2003).

## Methods

### Identify and characterize the gene

1. Identify gene of interest
2. Find approved gene name from the HUGO Gene Nomenclature Committee for your gene-of-interest (GOI) <http://www.genenames.org/>
  - a. e.g. FCGR2B (Fc gamma receptor IIb)
3. Use Ensembl or NCBI to search for the gene by name
  - a. <http://asia.ensembl.org/index.html>
  - b. <https://www.ncbi.nlm.nih.gov/gene>
4. Locate the transcript sequence for the gene
  - a. Multiple versions may exist for each gene
    - i. e.g. FCGR2B\_X1
5. Save the GenBank file cDNA transcripts to your computer using the appropriate naming convention
  - a. species code \_ approved gene name \_ reference code
    - i. e.g. sha\_FCGR2B\_ENSSHAT00000004489 (Ensembl)
    - ii. e.g. sha\_FCGR2B\_XM\_012547618.1 (GenBank)



- b. See species table below for appropriate species code
    - i. e.g. Tasmanian devil: *Sarcophilus harrisii* (sha)
    - ii. Ensembl has full list of three letter species codes
  - c. Ensembl codes start with ENS, then the species code, the either G for gene or T for transcript.
    - i. e.g. *ENSSHAG00000003908*
    - ii. e.g. *ENSSHAT00000004489*
  - d. You can also create a free NCBI account to save your transcripts online
  - e. The FASTA (.fa or .fasta) sequences are also ok, but the GenBank (.gb) have additional useful information
6. Open the files using SnapGene
  7. Copy the coding sequence (CDS) for the GOI and search the coding sequence in the cDNA file
  8. Create a new feature in the cDNA file that indicates where the coding sequence is located
    - a. Highlight the DNA sequence
    - b. Select Add Feature and set the feature type to ‘CDS’ for coding sequence
  9. Copy the amino acid translation (i.e. protein sequence) of the coding sequence and use this to characterize the protein using freely available analysis software
    - a. Phobius: predicts transmembrane topology and signal peptide predictor <http://phobius.sbc.su.se/>
    - b. ProtParam tool on the ExPASy bioinformatics server: predicts molecule weight and extinction coefficients <http://web.expasy.org/protparam/>
      - i. Remove the SigP from the protein sequence for the molecular weight and extinction coef analysis, because the SigP should not be present in the final protein product
    - c. Eukaryotic Linear Motif (ELM) resource: annotation and detection of eukaryotic linear motifs <http://elm.eu.org//index.html>
    - d. Simple Modular Architecture Resource Tool (SMART): identification and annotation of signaling domain sequences <http://smart.embl-heidelberg.de/>
  10. Save the results from each of the above programs
    - a. Save as either a webpage (html) or PDF
    - b. Save with the correct file names that matches your SnapGene file name
      - i. sha\_FCGR2B\_ENSSHAT00000004489\_Phobius
      - ii. sha\_FCGR2B\_ENSSHAT00000004489\_ProtParam
      - iii. sha\_FCGR2B\_ENSSHAT00000004489\_ELM
      - iv. sha\_FCGR2B\_ENSSHAT00000004489\_SMART
  11. If you don’t already have a gene table, create one using a spreadsheet and record the key information for the protein in the gene table so that you have a single reference to lookup key information about your protein. You can also save most of the information in the SnapGene (.dna) files
  12. Update the SnapGene file with key information from each prediction tool
    - a. Phobius: signal peptide (SigP), extracellular domain (ECD), transmembrane domain (TMD), intracellular domain (ICD, aka cytoplasmic domain)
    - b. ProtParam: molecular weight in kilodaltons (kDA),
    - c. ELM: key motifs (e.g. ITIM, ITAM)
    - d. SMART: Ig or TNF domains
  13. Make sure to save your SnapGene file. You can also export it as a GenBank (.gb) file as a backup
  14. In SnapGene copy the amino acid sequence and create a new protein sequence in



#### CLC Sequence Viewer

- a. Include the full sequence name to match your SnapGene file
    - i. e.g. sha\_FCGR2B\_ENSSHAT00000004489
  - b. Include the common name
    - i. e.g. devil
  - c. Include the Latin species name
    - i. e.g. Sarcophilus harrisii
15. Create a new folder for each gene
    - a. e.g. FCGR2B
  16. Save the protein file into the correct folder
  17. You can then find the orthologs and homologs for the gene by searching the NCBI or Ensembl databases for sequence similar to the DNA and/or protein sequence for your GOI.
  18. Use NCBI blast to find homologs for the species listed in the gene table below
  19. Download multiple sequences simultaneously (tick boxes for matched genes that are relevant)
  20. Save the files to your computer or NCBI account
  21. Copy the protein sequences into CLC Sequence Viewer to align the protein sequences
  22. Determine if key structural features of the protein are conserved among species
    - a. e.g. ITIM or ligand binding sequence

#### **Insert the gene into a plasmid expression vector (Gibson assembly)**

1. Choose an expression vector (aka plasmid) for the gene.
2. In SnapGene, set the origin of the plasmid to the first base at the 5' end of the ORI. This creates a standard reference point for all of the plasmids in our library
  - a. Open the SnapGene file and place the cursor between the first base of the 5' ORI and the base directly preceding the 5' base
  - b. View -> set origin
3. Linearize the plasmid either by PCR or restriction enzymes
  - a. If doing PCR linearization it is usually better to do a restriction digest first to avoid the PCR going all the way around the entire plasmid
4. Use the Gibson assembly or NEBuilder tool in SnapGene to construct the new plasmid *in silico*
  - a. 'Actions' -> 'Gibson assembly' ->
    - i. NEBuilder and Gibson assembly are very similar
5. \*\*\* Note: People with experience constructing plasmids might find it more effect to just cut and paste genes, but all insertions/deletions must be checked carefully to make sure restriction sites are correct and ORFs are in-frame.
6. Select the number of fragments to be inserted
7. In the 'Vector' tab, select how you will linearize the plasmid
  - a. e.g. SfiI restriction digest
8. Click the 'Fragment' tab and select the DNA sequence to be inserted
  - a. e.g. the coding sequence for the GOI – FCGR2B
9. \*\*\* Make sure the fragment to insert is in the correct orientation \*\*\*
10. Click the 'Product' tab
11. Select the bacterial transformation strain
  - a. e.g. DH5a
12. Choose the overlapping primers with at least 20 base pairs overlapping onto the vector
13. Click the regenerate the upstream and downstream restriction sites from the vector



- a. Only do this if you actually want to regenerate the restriction sites. There are many cases in which you will not want to regenerate the restriction sites.
14. Click 'Assemble'
15. Save the linearized vector, the fragment, and the assemble plasmid in the appropriate folder.
  - a. e.g. pAF1
16. Check everything in the vector to make sure all coding sequences are in-frame and all sequences are in correct orientation
17. Adjust primers as needed
18. Order primers

Table Species. Species names, abbreviations, and common names for species used in comparative sequence analysis. All species abbreviations match Ensembl species prefixes except humans, which do not have a prefix in Ensembl.

#	Short name	Common name	Abbreviation	Species name	Taxid
1	devil	tasmanian devil	sha	<i>Sarcophilus harrisii</i>	9305
2	bat	little brown bat	mlu	<i>Myotis lucifugus</i>	59463
3	cat	domestic cat	fca	<i>Felis catus</i>	9685
4	cattle	cattle	bta	<i>Bos taurus</i>	9913
5	dog	domestic dog	caf	<i>Canis familiaris</i>	9615
6	hamster	golden hamster	mau	<i>Mesocricetus auratus</i>	10036
7	human	human	has	<i>Homo sapiens</i>	9606
8	mouse	house mouse	mus	<i>Mus musculus</i>	10090
9	opossum	gray short-tailed opossum	mod	<i>Monodelphis domestica</i>	13616
10	panda	giant panda	ame	<i>Ailuropoda melanoleuca</i>	9646
11	platypus	platypus	oan	<i>Ornithorhynchus anatinus</i>	9258
12	wallaby	tammar wallaby	meu	<i>Macropus eugenii</i>	9315