Lab 19*

Cancer Mutation Mini-Project

i Instructions

Save this document to your computer and open it in a PDF viewer such as Preview (available on every mac) or Adobe Acrobat Reader (free for PC and Linux). Be sure to add your name and UC San Diego personal identification number (PID) and email below before answering all questions in the space provided.

Student Name UCSD PID UCSD Email

Background:

To identify somatic mutations in a tumor, DNA from the tumor is sequenced and compared to DNA from normal tissue in the same individual using variant calling algorithms.

Comparison of tumor sequences to those from normal tissue (rather than 'the human genome') is important to ensure that the detected differences are not germline mutations.

To identify which of the somatic mutations leads to the production of aberrant proteins, the location of the mutation in the genome is inspected to identify non-synonymous mutations (i.e. those that fall into protein coding regions and change the encoded amino acid).

As you go through this mini-project please remember to:

- Download the PDF version of this lab sheet (as noted above).
- Type all your answers directly in the space provided below each question.
- Save and upload your completed PDF to gradescope.

Good luck!

^{*}http://thegrantlab.org/teaching/

Questions:

Visit the following webpage and download your student specific sequences. These sequences resulted from an NGS analysis of patient healthy and tumor tissue.

N.B. Note that these sequence are unique for you and you must download your sequences and use them to answer the following questions in the space provided.

Q1. [1pt] What protein do these sequences correspond to?

Q2. [6pts] What are the tumor specific mutations in this particular case (e.g. A130V)?

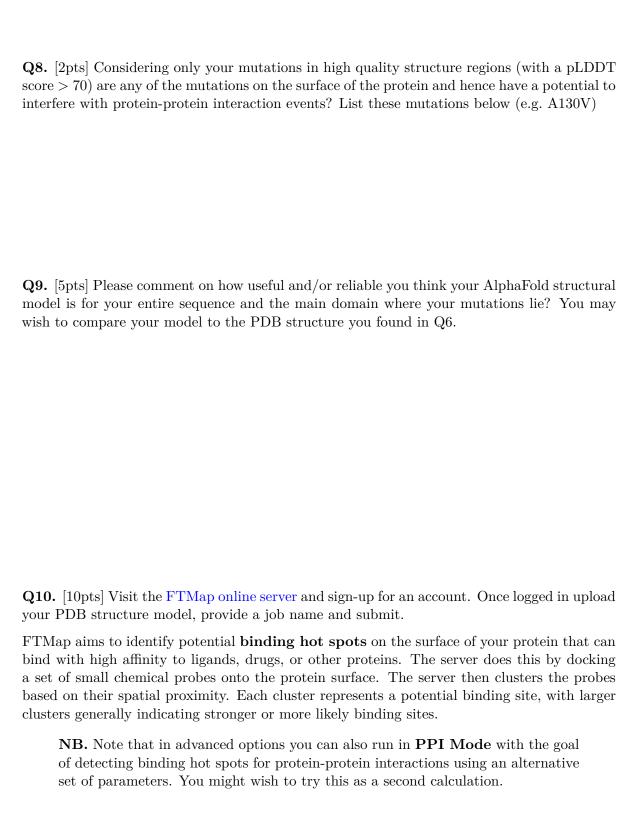
Q3. [1pts] Do your mutations cluster to any particular domain and if so give the name and PFAM id of this domain? Alternately note whether your protein is single domain and provide it's PFAM id (e.g. PF02196).

Q4. [2pts] Using the NCI-GDC list the observed top 2 missense mutations in this protein (amino acid substitutions)?

Q5. [2pts]	What two TCGA projects have the most cases affected by mutations of this gene?
Q6. [3pts]] List one RCSB PDB identifier with 100% identity to the wt_healthy sequence
	the percent coverage of your query sequence for this known structure? Alternately a most similar in sequence PDB structure along with it's percent identity, coverage e.
Optional	Extension:
	[S] Using AlphaFold notebook generate a structural model using the default parameter mutant sequence.
mode amin	that this can take some time depending upon your sequence length. If your el is taking many hours to generate or your input sequence yields a "too many o acids" (i.e. length) error you can focus on the main PFAM domain of interest canswer to O3 above)

Once complete save the resulting PDB format file for your records and use Mol-star (or your favorite molecular viewer) to render a molecular figure. In this figure please clearly show your mutant amino acid side chains as spacefill and the protein as cartoon colored by local alpha fold pLDDT quality score. This score is contained in the B-factor column of your

PDB downloaded file. Upload this image to GradeScope.



Are any of the identified "hot spots" near your cancer specific mutation sites? If so which mutation site(s)?

- End of Lab -