**STEP 2. Creating Samples of Interest**

1. Run protein prediction pipeline to convert RNA-seq data to predicted enzyme abundances (REQUIRED)
   * Requirements: RNA-seq data for samples of interest, preferably with most genes and not significant amount of missing values
2. Create a new folder within the ‘FBA-pipeline\Code + Models\data\protein\input’ folder, named with whatever you want the dataset name to be.
3. Place (1) RNA-seq dataset, and (2) copy of “\_OPTIONS\_.xlsx” file in this folder. Do not change the name of the options file.
4. Fill out the \_OPTIONS\_ file with dataset information.
5. Run ‘FBA-pipeline\Code + Models\data\protein\protein.ipynb’
6. Calculate Vmax values for new samples (REQUIRED)
7. Run ‘FBA-pipeline\Code + Models\data\vmax\vmax.ipynb’
8. Create clinical files with sample information (Optional)

* Clinical information is not needed for creation of FBA models
* Clinical information must be processed manually in the ‘FBA-pipeline\Code + Models\data\clinical\clinical.ipynb’ notebook. Useful information includes cell line/tumor attributes, radiation response, and drug response.
* Name the folder within ‘FBA-pipeline\Code + Models\data\clinical\’ with the same dataset name from Steps 1-2.
* Name each file with the same names as samples from Steps 1-2.

1. Process sample mutation data (Optional)
   * Mutation data is not needed for creation of FBA models
   * Mutation data must be processed manually in the ‘FBA-pipeline\Code + Models\data\mutation\mutation.ipynb’ notebook. See notebook sections for TCGA and CCLE samples for example processing.
   * Output files should be formatted the same as TCGA and CCLE sample files, with gene symbols and Envision scores
   * Name the folder within ‘FBA-pipeline\Code + Models\data\mutation\’ with the same dataset name from Steps 1-2.
   * Name each file with the same names as samples from Steps 1-2.