PROJECT FILE ORGANIZATION

- Supplemental Data tables folder
- All other files are in ~/pipeline_output/

Within pipeline_output/there is one directory for each dataset analyzed:

1. CCRCC
2. BRCA
3. UCEC
4. PDAC
5. LSCC
6. LUAD
7. HealthyTissues
8. Savitski
9. Mouse

Each of these directories has several subdirectories:

1. AA_subs_pipeline: contains output dictionaries from running decode_pipeline and other files needed for reproducing figures
2. RNAseq_data: contains subset of files generated in custom_database_generation pipeline
    a. Note: Mouse and Savitski analysis were not run with custom databases
3. MQ_output: contains MaxQuant output files in 2 directories:
    a. Dependent_peptide_search
    b. Validation search
    c. Note: Savitski directory has MQ_output with dependent peptide search results and MSFragger_validation with validation search results
4. Publication: contains publication source of raw data, publication supplemental files, including clinical metadata
5. Databases: contains sample-specific protein databases generated from custom_database_generation_pipeline, with appended substituted peptides

Pipeline_output/ also contains a directory with the results from Prosit rescoring.

Pipeline_output/analysis_dependencies/ contains all files generated through data analysis that are required for replicating the publication figures. This directory should be downloaded to replicate the analysis and figures in the manuscript.