ACTIVITY 1:

1. What is the biological question?

Paper A: Impact of genetic variation on three dimensional structure and function of proteins
—> Systematic and qualitative analysis of cases of proteins where their atomic-level structural differences caused by variation have been determined
—> **Structural comparison** of wild type and mutated proteins (specifically for Single Nucleotide Variations (SNV), accounting for 90% of sequence difference) and their **functional effect** of the single AA change.

Paper B: Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10000 patients
→ Find patterns of alterations across "tumor types" that could possibly be targeted therapeutically
→ large-scale clinical sequencing of 10'000 patients (tumor vs control sequence data) to

2. What is the method?

Paper A:

- data is collection of proteins from from PDB (one AA mutated!), which was first "filtered" (must be matched by UniProt and 3D structure), then duplicate entries were removed. Resulted in dataset of 374 unique proteins for which 3D atomic level coordinates are available.
- To analyze **functional consequence** of given SNV on gene product, literature was reviewed and the mutation was classified by whether it had effect on:
    - a) activity
    - b) stability
    - c) binding
    - d) assembly
    - e) rearrangement.
    - Problem: variation of detail with which each SNVs related changes have been experimentally analyzed
- **Structural** location of the SNV was manually categorized into 2 main groups
    - 1) surface
    - 2) buried
    - and for each, whether it falls within
        - a) loop
        - b) alpha H
        - c) beta S.

Paper B:

- large-scale, prospective clinical sequencing initiative using comprehensive assay (MSK-IMPACT= hybridization capture-based next-generation sequencing panel capable of detecting protein coding mutations, copy number alterations….) to compile tumor and "normal"/healthy (control) sequence data
- With this data, identification of:
    - clinically relevant somatic mutations

- novel noncoding alterations
- mutational signatures that were shared by common and rare tumor types.
- Determine prevalence of "actionable" mutations
- Matching patients to molecularly targeted therapy

3. What significant scientific contribution does the paper make?

Paper A:

- Main target of paper: *Identify patterns concerning sites at which point mutations occur*
- EXPECTED distribution (considering secondary structures) → 46% alpha, 24% beta, 30% loop
- OBSERVED: 79% of SNVs lie on protein surface (!)
  - 52% in loop
    - Discrepancy explainable, as AA changes in loop can often be compensated for without affecting structure/function)
  - 34% alpha
  - 14% beta
- 21% were buried in core of protein
  - 42% alpha
  - 31% beta
  - 27% loop
- **"STRUCTURAL" PATTERN thus:**
  - Proteins with SNV's on surface are more frequently found in loops
  - Proteins with SNV's in core are more frequently found in alpha H / beta S

- **"FUNCTIONAL PATTERNS": Range of possible SNV effects at protein level are significantly greater than currently assumed** (prediction of (functional) consequences remains challenge…).
  - SNV can have multiple effects on protein structure and function thereby also affecting multiple of the six categories (e.g. activity and aggregation…)

Paper B:

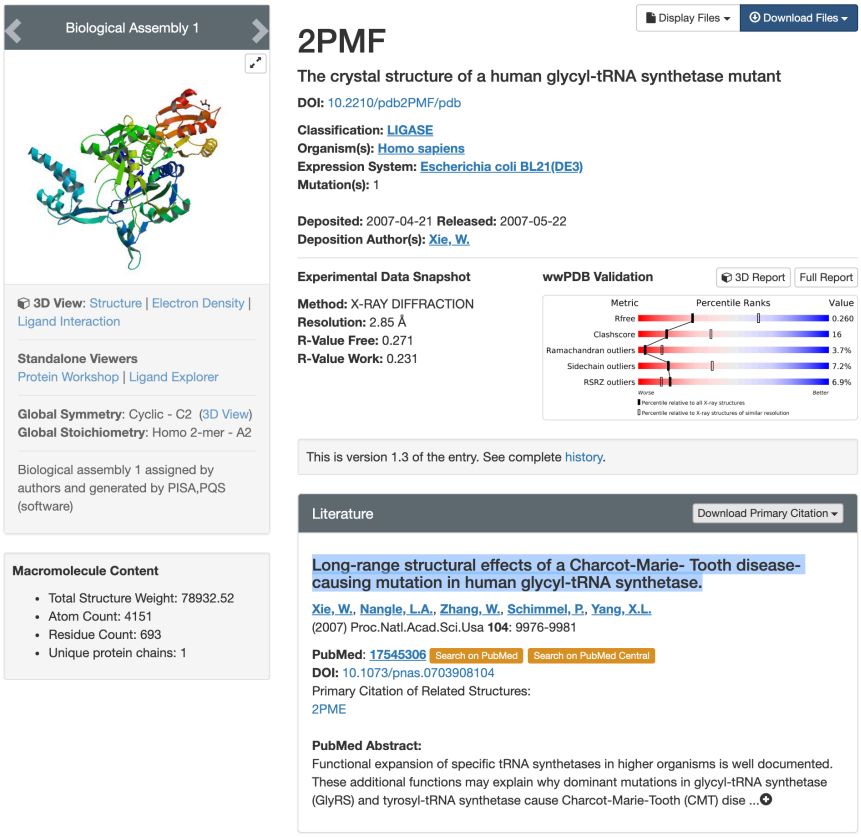
- Use of data! Guide selection of genomically matched therapies for patients
  - Data set of manually reviewed mutations, CNAs and genomic rearrangements
  - 300 detailed tumor types (providing prevalence of driver alterations across cancers)
  - Data set as potential resource for identification of novel biomarkers to inform prognosis and predict response/resistance to therapy

ACTIVITY 2:

1. Go through the papers and identify one protein whose mutation is associated with a disease state.

human glycyl-tRNA synthetase mutant (2PMF) is associated with Charcot-Marie-Tooth disease

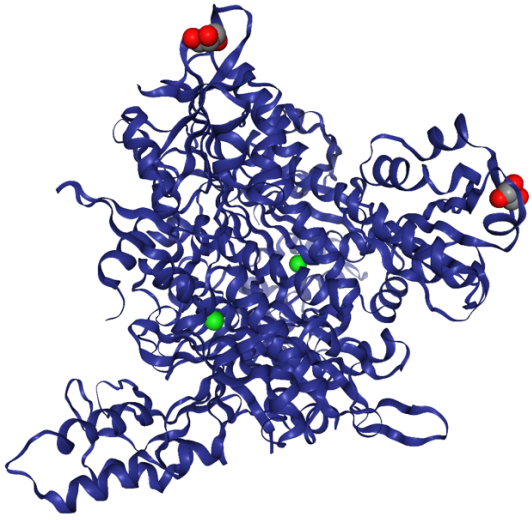
2. Go to the PDB and search for the protein using its name. Give a brief summary of the available information in the PDB:



1. How many entries are there for this protein?
   15
2. Which method(s) were used in characterization of the structure?
   X-RAY DIFFRACTION

4. What is the Uniprot ID?
   P41250
5. How many chains does it have?
   1
6. What is the sequence length?
   693
7. Which ligands (if any) are present?
   1. Chloride Ion
   2. Glycerol
8. Examine the 3D structure using 3D view.

9. Color by chain. (+ ligands as spacefill)



10. Play with all of the options and prepare a structure figure.
Options: color by molecule type