J. Jedediah Smith

BIFX 504

11/17/2021

Watch the first 23 minutes of "A guide to proteomics data analysis using UniProt and InterPro" and answer these questions:

1. What resources are available at UniProt?

Detailed information on protein function, interactions, pathways, etc. Sequences, including isoforms, disease variants, and PTMs. Stable identifiers. All this is collected into the Swiss-Prot and TrEMBL, the two halves of the UniProt database.

1. What is the main difference between SwissProt and Trembl?

SwissProt is manually curated and reviewed, while TrEMBL is automatically annotated with expert curated rules.

1. How many reference proteomes are available on UniProt?

As of the webinar recording, UniProt has 20,125 reference proteomes.

1. How do you download all of the canonical protein sequences for an organism in FASTA format?

Navigate to the organism of interest, scroll down to components, and click the download button. Make sure the format setting is set to “FASTA (canonical).”

1. How can you identify what version of the proteome you are using?

Near the top of the page, there is a “last modified” category. This can be citied in papers so that readers know what version you used, as the entries are updated frequently.

1. How can you use UniProt to identify what protein a peptide came from?

Go to home page, scroll down to getting started, and click on “peptide search.” Or one can use the bar at the top of the page. Just input the peptide sequence and click search. This shows a list of what proteins link to the peptide, and at which residues.

1. What information is available on a protein entry on UniProt?

Protein name, gene name, organism, status, function, taxonomy, subcellular localization.

1. What can aspects of your protein of interest can you visualize in the Feature Viewer?

Domains & sites, sequence information, structural features, variants, antigenic sequences, PTM, proteomics, mutagenesis, and more.

1. What is "a variant" according to UniProt?

An amino acid change that is different from the canonical sequence and isn’t involved in a splicing event or isoforms. So possibly involved with disease or abnormal phenotypes.

1. How can you download the full amino acid sequence of a protein of interest?

From the protein page, scroll down and click sequences tab on the left hand bar. From there, click the FASTA button and it will open a page of the full amino acid sequence as raw text.

1. What information is in the FASTA header for a protein on UniProt?

Name, length, mass and checksum. If it is an isoform, it also shows a sample of how it is different from the canonical sequence.

1. How can you easily compare two isoforms of the same protein to each other?

You can use the align button to compare them, or add them to basket and align them that way.