**Practical Assignment**

**Module name: Introduction to Databases and Resources**

**Session name: Protein Classification and Resources**

**Trainer: Shaun Aron**

**Participant:** <*write your name here>*

**Date:** <*write today’s date here*>

**Protein Classification and Resources**

**Introduction**

Please go through the entire practical exercise. No formal answers are required for **Task 1**. This section of the practical is focused on you exploring and becoming familiar with the resources. You are welcome to make your own notes and comments for this section. Please complete the answers for **Task2: Annotating a Protein Sequence of Interest** section **ONLY** in this document and upload this to the Vula website before the deadline for submissions.

**Tools used in this session**

Uniprot <http://www.uniprot.org/>

ExPASY <https://www.expasy.org/proteomics>

**Please note**

* **Hand-in information** If you are formally enrolled in the IBT course, please upload your completed assignment to the Vula ‘Practical Assignments’ tab. Take note of the final hand-in date for each assignment, which will be indicated on Vula.

**Task 1: Protein Resources**

**Task 1: Instructions**

1. Go to the Uniprot web site: http://www.uniprot.org/

(a) What is the resource used for?

(b) What databases can be searched using the resource?

(c) What are the different ways in which the databases can be searched?

1. Go to the main ExPASy web site https://www.expasy.org/proteomics

(a) What type of resource is this?

(b) What databases are available via this resource?

(c) What types of data analyses can you carry out using the tools on this resource?

(e) What information is stored in the ProSite database and what can it be used for?

**Task 2: Annotating a Protein Sequence of Interest**

**Task 2: Instructions**

You have been provided with the protein sequence for the longest isoform encoded from an unknown human gene. You can find and download the sequence from the Vula site: Resources -> Week 2 Introduction to Databases and Resources -> practical session 4.fasta

1. Navigate to the UniProt homepage <http://www.uniprot.org/>
   1. Select the BLAST option from the top menu
   2. Copy and paste the full sequence into the search box and click on the Run BLAST button.
   3. The BLAST tool allows you to search a query sequence against a database of known annotated sequences. In this case we are searching our sequence of interest against the UniProtKB database. Be sure to change the database to UniProtKB. Do not worry about understanding the details of how the BLAST algorithm works, as this will be covered in an upcoming module.
2. State the protein names of the top two hits that the query sequence matched to in the UniProtKB database.
3. Which other organism/s has a protein sequence with the highest identity to our query sequence of interest? Provide the scientific and common name for this organism/s.
4. Click on the accession ID for the top hit in the results table.
   1. What gene is this protein sequence an isoform of?
   2. Has entry for the protein sequence for this gene been reviewed? What is the source for this protein sequence? What experimental evidence is there for this protein sequence?
   3. What is the key function of the protein sequence?
   4. In which location in the cell is this protein usually expressed?
   5. What is the name of the main protein domain identified in this sequence? Hint : Look at the Feature Viewer page for the Uniprot entry for the protein.
   6. What is the source of the protein domain annotation for this sequence and what approach has been used to identify the domain?
5. Find the linkout from UniProt entry page for the protein to the InterPro database. *Hint: look for “view protein in InterPro”*  under the Cross References section.
   1. What information can you find on this page about the protein sequence?
   2. What are the 2 main domains found in this protein sequence?
   3. What biological process is the protein involved in?

**Task 2: participant’s answer**

<*start typing your answer here*>