**Practical Assignment**

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

**Session name: Bioinformatics Databases and Resources**

**Trainer: Shaun Aron**

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

**Date:** <*write today’s date here*>

**Bioinformatics Databases**

**Introduction**

Please go through the entire practical exercise. No formal answers are required for **Task 1** and **Task 2**. These two sections of the practical are focused on you exploring and becoming familiar with the two resources. You are welcome to make your own notes and comments for these two sections. Please complete the answers for **Task3: Finding and extracting information** section **ONLY** in this document and upload this to the Vula website before the deadline for submissions.

**Tools used in this session**

NCBI [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)

EBI [www.ebi.ac.uk](http://www.ebi.ac.uk)

**Please note**

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

**Task 1: Exploring the NCBI resources**

**Task 1: Instructions**

Open a browser and navigate to the main NCBI webpage ([www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov)):

1. Describe briefly what resources are found in each section of the webpage:
   1. The left hand panel
   2. The right hand panel
   3. The centre panel
   4. The lower panel
2. From the right hand panel or the drop down menu (select the database and leave the search box empty and click search to access a database or resource homepage) navigate to the homepage for the following resources and describe what each resource is used for:
   1. PubMed Central
   2. Genome
   3. PubChem
   4. SRA
   5. dbGap
3. Navigate to the **Gene** resource and have a look at the types of queries that can be used to search for information. (No answer necessary)
4. Select the DNA and RNA tab from the menu on the left on the homepage.
   1. What information does the **RefSeqGene** database contain?
   2. What are the download and submission tabs used for?

**Task 2: Exploring the EBI resources**

**Task 2: Instructions**

Open another browser window and navigate to the EMBL-EBI webpage ([www.ebi.ac.uk](http://www.ebi.ac.uk)):

1. Navigate to the services page.
   1. How are the resources/databases categorised in comparison to NCBI?
   2. Select the DNA and RNA category. Compare the resources and databases available here in comparison to NCBI. Briefly describe what the main differences are.
2. Navigate to the Uniprot website from the homepage.
   1. What is the resource used for?

**Task 3: Finding and extracting information**

**Task 3: Instructions**

1. You are about to conduct a study examining the genetic variation present in the *LDLR* gene in a local population group in your country. As a starting point you would like to find all relevant information on the gene. Using either the NCBI or EBI resources, find the following information about the gene.
   1. What is the full name of the gene?
   2. What is the chromosome number and genomic location of the gene?
   3. How many protein-coding transcripts have been annotated for the gene?
   4. Provide a brief description of the function of the protein encoded by the *LDLR* gene.
   5. What is the accession number for the genomic sequence for the GenBank entry for the gene?
   6. Is there a RefSeq entry for the *LDLR* gene? If so, provide the accession number of the sequence from which the RefSeq was derived.
   7. What disease is associated with mutations in the *LDLR* gene? Provide the associated OMIM entry number for the disease.
   8. Provide a citation (in any format) for a journal article that used whole exome sequencing to identify new variants in the *LDLR* gene
   9. What is the accession number for the RefSeq mRNA and protein sequence encoded by the *LDLR* gene?
   10. What is the length of the protein encoded by the longest *LDLR* transcript?
   11. Extract and paste the amino acid sequence for the longest protein encoded by the *LDLR* gene.

**Task 3: participant’s answer**

<*start typing your answer here*>