**Hemoglobin bioinformatics teaching notes**

**Learning objectives:**

At the end of this worksheet, students should be able to:

* Find sequences and data by searching NCBI
* Identify basic information that is found within a GenBank file
* Perform a basic multiple sequence alignment (MSA) between two sequences

This case study was originally created to provide undergraduate students with an introduction to bioinformatics. One major purpose of this set of worksheets was to connect DNA mutations to the concept of protein structure and function.

Worksheets could be used independently, although they were generated to explore the connection between DNA mutations and proteins structure/function. This set of worksheets was used as an adaptation to the MoleCaseNet case Nicholas’ story (<https://molecular-casenet.rcsb.org/>) which further visualizes the protein structure and function using online 3D tools.

Through this set of worksheets (3 parts), students will be:

* Introduced to the structure of a gene
* Directed towards finding sequences and information from files using GenBank
* Creating an alignment of the a and ß-globin proteins
* Articulates the nucleic acid and amino acid mutation in ß-globin
* Reinforces the concepts of protein structure (primary, secondary, tertiary and quaternary)
* Directed to identify the chromosomal location of ß-globin
* Reinforce the structure of human chromosomes

**Pre-requisite knowledge**

The case provides much of the background information as well as instructions on how to access and view the bioinformatics tools. This is not a case study to teach the algorithms of bioinformatics (e.g., BLAST, MSA, etc.) but is consists of worksheets to lead students through the use of these resources. The definition of a gene is more of a molecular definition that includes the transcription product as well as the regulatory and promoter sequences within the definition (see definition used below). This is typically not the definition used in Genetics textbooks, but is similar to historical definitions and is inclusive of non-translated (non-coding) RNAs. This definition is used in the first worksheet.

**Definition of a gene as taught in class**:

A gene is a region of specific DNA bases, whose sequence is transcribed into coding or non-coding RNA and the non-transcribed DNA bases that regulate the temporal, spatial and developmental expression of that RNA

Students should be familiar with the concept of the Central Dogma of Molecular biology as well as the four levels of protein structure (primary, secondary, tertiary and quaternary). The chemical bonds that stabilize the different protein structures should have already been presented to students.

**Usage in class**

This set of worksheets could be used in a classroom setting, in a laboratory setting or as homework. The resources used are available on the Internet. The requisite data is provided in the worksheets. Each part of the case could be completed in a 50 min class period.

It is recommended that the students be provided with the worksheets in Word format so that they can copy and paste information directly into the document. This set of worksheets has been assigned as a document to be turned in completed or specific questions (highlighted in orange in the document) have been entered as questions to be filled out on an LMS (Connect, Sakai or Canvas have been used).

An answer key is available for each worksheet. The author has indicated in orange which questions are typically graded for worksheets 1 and 3 (all of the questions in worksheet 2 are graded). The student version does not have the shading for those questions.