Overview

This is an **entry level** course aimed that those with a reasonable biological background but **no significant experience with bioinformatics**. The course is broadly based around a series of exercises in which a combination of simple analytical tools and publicly available databases is applied to the investigation of a single human gene. The training manual for the course is comprised of detailed instruction for the tasks undertaken. Included are, questions (with answers) and discussion of and the interpretation of the results achieved. The manual is under constant development, a recent version is available online.

Outline

Participants are asked to imagine an interest in the disease **aniridia**. Course exercises then provide extremely detailed instruction leading participants to discover the gene primarily associated with this disease and all that is interesting about that gene and its protein products.

Objectives

This course is not intended to fully meet the requirements of many modern research projects. The intention is more to examine and practice the basic tools that underpin many current bioinformatic solutions and have changed little over the years. Specifically, we set out to show how many answers can be simply be looked up in an appropriate information source and/or how it is also usually a trivial matter to compute the similar/identical answers using readily available software.

We set out to explain the operation of the various programs used in the exercise, but only to the extent that allows a user to select parameters intelligently and to interpret results fully.

Timetable

The main topics covered, in order, are as follows:

Simple Information and Data (i.e. sequence) retrieval from public resources (primarily the **NCBI** and **EBI**).

Genome Databases (primarily **Ensembl** including access using **Biomart**)

Pairwise Sequence Alignment (**dotplots**, **global** and **local** alignments)

Database searching (using elements of the **blast** family)

Primer design (primarily using the **NCBI** services)

Secondary Structure Prediction (**GOR**, **JPred**)

Multiple Sequence Alignment (clustal, muscle, t-coffee …)

PSI-blast (an explicit example, but incorporated into a number of the previous topics)

3D Structure (a very superficial look specific to the exercise protein)

The timing can vary depending on circumstances. A very rushed **3 days** to a relatively leisurely **5 days**.

DPJ – 2016.03.03