FOSS Project

Bioinformatics Laboratory Tutorial for an Introductory Undergraduate Bioinformatics Course:
Lab 1. Using Google Dataset Search to Retrieve Public Datasets and Lab 2. Utilizing Creative Commons to License a product (Graded: 50 points).

Student learning objectives (What you will learn)
Learn about Google Dataset search
Learn how to read a scientific abstract
Learn about scientific data sets and DAVID (**D**atabase for **A**nnotation, **V**isualization and **I**ntegrated **D**iscovery) tools to analyze this data
Learn about licensing data and the terms and conditions of different licenses
Learn how to create reports
Learn to think critically and write scientifically

Student core competencies (What you will learn to do)
Retrieve and download a dataset linked to a scientific article
Analyze a data set (gene expression data) using the DAVID bioinformatics tools
Read a scientific abstract and extract relevant information
Create and license a product using creative commons
Develop critical thinking and team skills.
Create and write reports

Lab 1. Using Google Dataset Search to Retrieve Public Datasets

The task:

There exist thousands of data repositories on the web, that provide access to millions of datasets. To enable ease of access to this data, Google has launched Dataset Search, so that scientists or anyone else can find the data required for their work. Similar to how Google Scholar works, Dataset Search lets you find datasets wherever they're hosted, such as a publisher's site, a digital library, or an author's personal web page.

Imagine that you are a scientist working in the bioinformatics field. Your PI has requested that you reanalyze the data in a specific public data set from 2011 (because the wealth of knowledge in the field has grown since 2011), he needs your help to determine the difference in the gene expression profile in etoposide sensitive versus etoposide resistant breast cancer cells. However, your lead PI has gone on vacation giving you 'extremely vague' instructions on how to accomplish this task. Work in groups using the instructions below that have been improved upon by a kind colleague in your field to accomplish this task.

- To use Google Dataset Search type "Google Dataset' into the Google window.

We are going to search and download a specific set of data and the linked paper from this portal.

- Retrieve information about the dataset GSE28414 by copying and pasting the paper name: 'Differential gene and microRNA expression between etoposide resistant and etoposide sensitive MCF7 breast cancer cell lines' into the Dataset search box
- Find the original paper by copying and pasting its unique identifier (in Google datasets) in a web browser (https://doi.org/10.1371/journal.pone.0045268)
- Download the paper to your desktop by clicking on the unique identifier link and once on the landing page to PLOS ONE downloading the file as a .pdf.
- Read the paper abstract and write a summary of what this data indicates in the biological sense so that you can understand the work.
- Go back to your dataset search and download the dataset file by clicking on the link/ blue button 'explore at figshare.com' and double clicking DATA SET
- Group activity: In your groups find 2 other ways to access and download this dataset and report out. (Ans: pubmed, PLOS One website, Google scholar etc.)
- Name the different types of experimental data contained in this dataset ie; Western blot data etc.
- Use the paper that you downloaded as a reference to re-analyze the microarray data comparing only the top 20 overexpressed genes in MCF7/MCF7VP dataset using the DAVID tool (https://david.ncifcrf.gov/) to find functionally related gene groups, (https://david.ncifcrf.gov/gene2gene.jsp).
- Work together in your groups to follow the instructions in the DAVID web pages since your absent minded PI has forgotten to give you instructions on how to do this.
- Identify the gene groups and put forward a hypothesis on why you think these gene groups are important in this study.
- Create a report for your PI with this information (rubric and instructions on Moodle). This report is worth 20 points.
- Create a concept map of the steps that were required to complete this activity. The concept map is worth 10 points.
- Upload your report and concept map to Moodle by 12 midnight Feb 25th, 2020.

Lab 2. Utilizing Creative Commons to License a product

Your PI emails you from Paradise Beach on the Greek Island of Mykonos to get an update on how the project is going and asks you to look into the licensing of above data set and send him the related information. He further asks you to create an image from the data set after you have reanalyzed the data and license it under a creative commons license. Once again giving you very vague input on how you are to accomplish this task.

- Go back to your dataset search results and scroll down to 'License'
- What kind of license does the data hold? (Attribution 4.0 (CC BY 4.0))
- Click on the link to the license for the dataset to learn about the license and discuss the license terms in your group and what they mean. Report out.
- What is a license and why are licenses required? Discuss in groups and report out.
- In your groups discuss what does Creative Commons does and draw a figure to depict what CC does. Share and explain this figure to the class. https://creativecommons.org/about/
- Explore the different kinds of licensing at https://creativecommons.org/use-remix/cc-licenses/ and write down the different types of licenses that are available.
- Group activity: In your groups create an image from your newly reanalyzed dataset and then license said image.
- Go to https://creativecommons.org/choose/ and choose a licensing option for your image and license your image.
- Upload your image, license information and your rationale for using this particular license in a brief report to Moodle by 1st March, 2020, midnight. The report is worth 10 points.

Minute paper activity (in class)
In one minute write down the main concepts that you learnt today. I will potentially grade this activity for extra credit.

Graded Homework
Make a video explaining licensing to your Grandma/elderly relative using examples that they can relate to. Feel free to be extremely creative! Upload the video to YouTube and submit the link on Moodle by midnight 4th March, 2020. We will watch your videos in lab/class and they will be graded. The video is worth 10 points of your grade.