Name: Brian Tinsley

Date: 1/23/22

TCGA Website Scavenger Hunt

QBIO Public Data Analysis

Nicole Black, David Wen

TCGA (Home Page):

The Cancer Genome Atlas (TCGA), founded in 2006, is a cancer genomics program hosted by the _____ National Cancer Institute and the National Human Genome Research Institute. The publicly available data from this project includes _____genomic_____, epigenomic, _transcriptomic_____, and proteomic data. This data was collected from 20,000 different samples that span 33 different cancer types, including colorectal cancer, which we will be focusing on this semester.

Program History:

Describe one outcome or impact of TCGA: _____
Changed the way cancer patients are treated in the clinic

Briefly skim the "Timeline & Milestones" page. When did TCGA publish their paper on colorectal carcinoma? _____july 2012

Because TCGA is a public dataset, and one of the first of its kind, they faced some initial concerns regarding the ethics of releasing health data to the public. Choose one of the papers in the "Ethics & Policies" section to skim. What is one way that your paper addresses these privacy concerns? _____
Qualified investigators may gain access to controlled data for research purposes following agreement to terms and conditions for responsible use of the data, such as secure storage in computers not linked directly to the internet

TCGA Cancers Selected for Study:

List three criteria used to select which cancers to study: _____Poor prognosis
Overall public health impact
Availability of samples meeting standards for patient consent

Open the colorectal adenocarcinoma page and read TCGA's provided background. List one interesting fact you found: _____rectal and colon cancer are nearly the same_____

Publications by TCGA:

TCGA published (at least) one paper on each of their studied cancer types. These papers, called marker papers, include an early analysis of the data, including any molecular characterizations that were performed. Read the abstract of the colon and rectal cancer paper. List any genes you come across (these may be good starting points for your future analyses of this cancer): _____ARID1A, SOX9 and FAM123B_____

Using TCGA:

Go to the Genomic Data Commons (GDC) Data Portal via the link on this page then launching the Data Portal. This portal lets you view TCGA's data in a visual way. Let's explore this website. According to the Data Portal Summary, there are _____70_____ projects in the GDC data portal. Now click on the "Projects" tab. Notice that not all projects in this data portal are TCGA-affiliated, though TCGA does make up _____33_____ of the projects included.

Name: _____

Date: _____

TCGA Website Scavenger Hunt

QBIO Public Data Analysis

Nicole Black, David Wen

Using TCGA (Continued)

Under the "Program" tab, select just TCGA studies. According to the graph at the top of the page, tp53 is the most mutated gene in TCGA projects, affecting approximately 35 % of cases.

Return to the GDC Portal home page. Now click the colorectal image in the diagram to the right of the page. This directs you to the "Exploration" tab and automatically selects all primary sites associated with colorectal cancers. These sites include the colon, the rectum, and the rectosigmoid junction. Now select TCGA as the program, and TCGA-COAD as the as the project. This is the data we will be focusing on this semester.

The table on this page shows each patient along with their data. Explore the data provided for the first patient (Case ID: TCGA-CA-6717).

Now explore the Cases, Genes, Mutations, and OncoGrid tabs. What is one takeaway from the plots provided here: APC and TTN genes mutated more frequently than tp53

Discussion:

Discuss the following questions with your group and take notes on your conversation:

1. What is the goal of TCGA?

the goal of TCGA is to allow researchers everywhere to have access to public data in hopes of speeding up the process of new findings to help cure, treat, and prevent cancer

2. What are some ways that we use TCGA's data for our own cancer research? (Think about the types of data available and brainstorm some research questions that can be proposed given that data.)

we can look at the most mutated genes and nucleotide sites, and then compare that to clinical data to see if some clinical categories have a high correlation with these mutations or if maybe different treatments prevent the cancer

3. What are the benefits and drawbacks of TCGA or other large publicly available datasets?

benefits of TCGA are plentiful, allowing students like us to make a real impact in research, the drawbacks are that we lose sensitivity to the severity of cancer and how tragic it is… we are treating it like numbers

Please email a pdf of your worksheet to blackn@usc.edu before the start of the next class with the Subject Line: QBIO 490 Homework 3 – LastNameFirstInitial. For example: QBIO 490 Homework 3 – GuionK.