Module 4.2: Learn - Coding

4.2 Coding

For this module, we are assigning to you a specific sex chromosome gene to analyze in the CCLE. You will use the same code you worked on for XIST in Module 3.2.

Assignment: Analyze expression of your assigned sex chromosome gene in the CCLE

In 'CCLE_gene_expression.Rmd' that you worked on in Module 3.2, there is a chunk called ChooseGene in a section called Gene of Interest. In this section, you can specify the name of the gene you would like to look up in the CCLE. This was set to XIST in Module 3, but we set it as a variable so that it can be easily changed to look at other genes in the data set. You can see in this chunk that the character string you set in this chunk will be searched in the Description column of that CCLE data table for an exact match. After you extract the rows that contain this gene name in the Description column into the 'chosen_gene_data' variable, take a look at the contents of this variable to make sure that your gene is in the data set and data has been successfully extracted. From there, you can go line by line, chunk by chunk, to prepare the same visualizations as we did for XIST in the previous module.

Specific steps and best practices for analyzing your assigned gene in the CCLE

Create a new directory for analysis of your assigned gene in your working directory

1. For working on ASU Sol supercomputer:

    1. Open an RStudio server session
    2. Go to the Terminal tab

3. Navigate to your /scratch directory that you are working in for this class ('cd /scratch/username/

4. To prevent us from accidentally overwriting output files from XIST analysis, let's make a new dir of your assigned gene and copy our code from last module into it

5. Make a new directory ('mkdir new_sex_chr_gene')

6. Copy code into that directory ('cp CCLE_gene_expression.Rmd new_sex_chr_gene')

7. Navigate into that directory ('cd new_sex_chr_gene')

8. Change the name of the code so we know which version we are working on ('mv CCLE_gene_ CCLE_gene_expression_mygene.Rmd')

9. To start working on that code, go to the Files tab on the right, find the directory and new copy of made, and double click 'CCLE_gene_expression_mygene.Rmd' to open

2. For working on RStudio Desktop on your own computer:

1. Open an explorer window and navigate to the working folder for this project where you downloa stored your code in Module 3.2

2. Make a new folder for analyzing your assigned gene

3. Copy and paste the code from last module 'CCLE_gene_expression.Rmd' into that new folder

4. Change the name so we know which version we are working on (CCLE_gene_expression_myg

5. Double click the Rmd so that it opens automatically in RStudio Desktop (or open RStudio Desk the File, Open File menu)

3. In the SetDirectory chunk, change the working directory path to be in the directory you just created (add 'new_ working_directory path)

1. We can keep the data_directory variable the same because the location of the data files have r

4. Go through section by section executing code

1. The chunks in the 'Exploring the CCLE data set' and 'Create lookup tables' sections will not be module

5. In the ChooseGene chunk in the 'Gene of Interest' section, change the 'chosen_gene' variable to be your ass

6. Go through and produce the various data visualizations as we did with XIST in Module 3.2

7. Add new visualizations to ask new questions as they arise

Possible questions to ask

There are many questions you can ask and try to answer with code, but here are a few to get you started:

1. Does my gene follow the expected expression patterns in cell lines with reported sex? (chromosome X ge reported female, chromosome Y genes high in cell lines reported male)
2. Are the distributions in cell lines reported female versus male distinct? Or do they overlap? Do they have l
3. Is there a type of cancer that shows especially high expression of my gene? In primary versus metastatic age group?
4. Can I look up the function of my gene or what other genes it interacts with? Does the expression of those patterns?
5. What does the expression of my gene look like in cell lines that did not have a reported sex? Does this ge those cell lines into two groups that could reflect the sex chromosome complement of those cell lines?

In this and the next module, you will be presenting these results as a manuscript. The manuscript will contain a questions you are asking and the results you got along with a technical summary of the code (packages and func answer those questions.

Optional follow-up activity: Looking up gene in human cancer tissues

In this research project we are looking up sex chromosome genes in cancer cell lines, but there are other public cancer tissues sequenced directly after removal from the body (without having been made into a cell line).

There is an awesome webtool to look at your gene's expression in large collections of cancer tissues: UCSC Xer
https://xenabrowser.net/heatmap/ ⤓ **(https://xenabrowser.net/heatmap/)**

There are a bunch of tutorials to help you to use all of the features for this tool, but here we will walk you through and age distribution in the context of your gene's expression. Here we are using XIST as an example:

1. The first prompt will ask you what study you would like to use, start to type "pan" in the search box with "I kno use" selected, and you will see an option to click on "GDC Pan-Cancer (PANCAN)", which is a giant collectior samples on which many types of genomics were performed.

2. After clicking the "TO FIRST VARIABLE" green button below, it will ask you to select variables we want to visu
   select our gene of interest (type 'XIST' in the 'Add Gene or Position' box and select it when it comes up) and
   inside the 'Dataset' search box and select "Gene Expression" from the "Basic" menu).
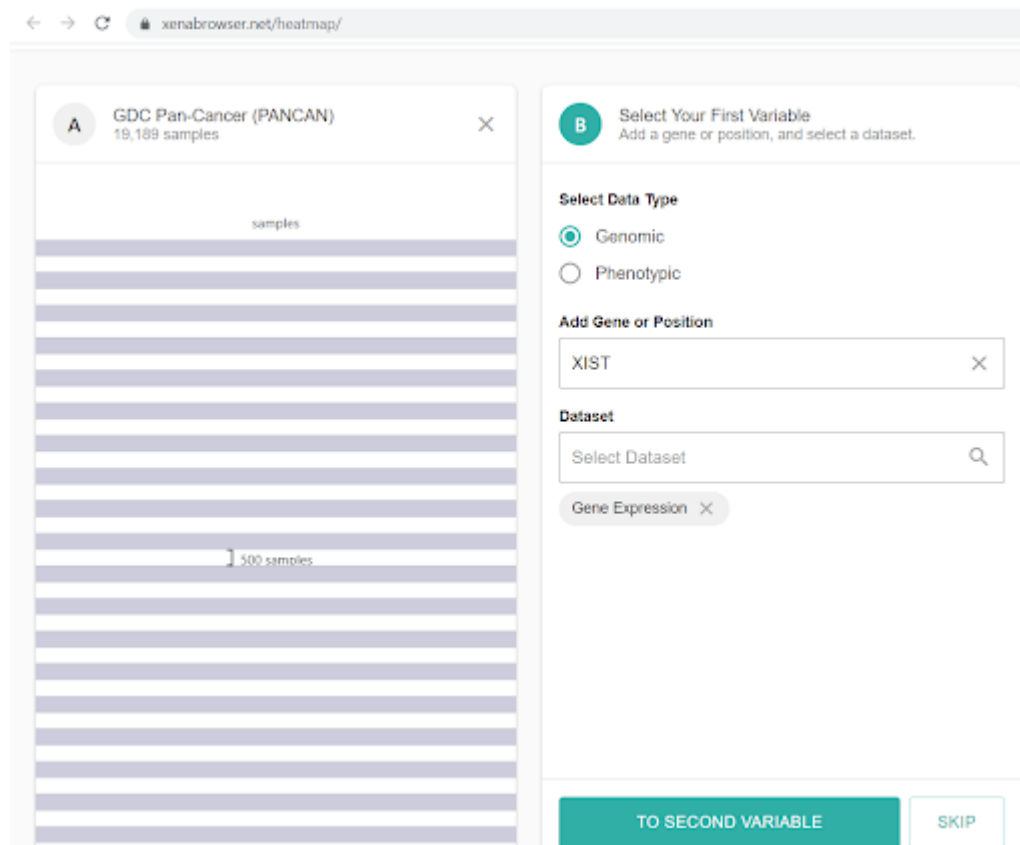
3. After clicking the "TO SECOND VARIABLE" green button, you will see that the data has been sorted for XIST second box. You can select the other variables you want to view, which we will select to be "Phenotypic" and "Age at Diagnosis in Years".
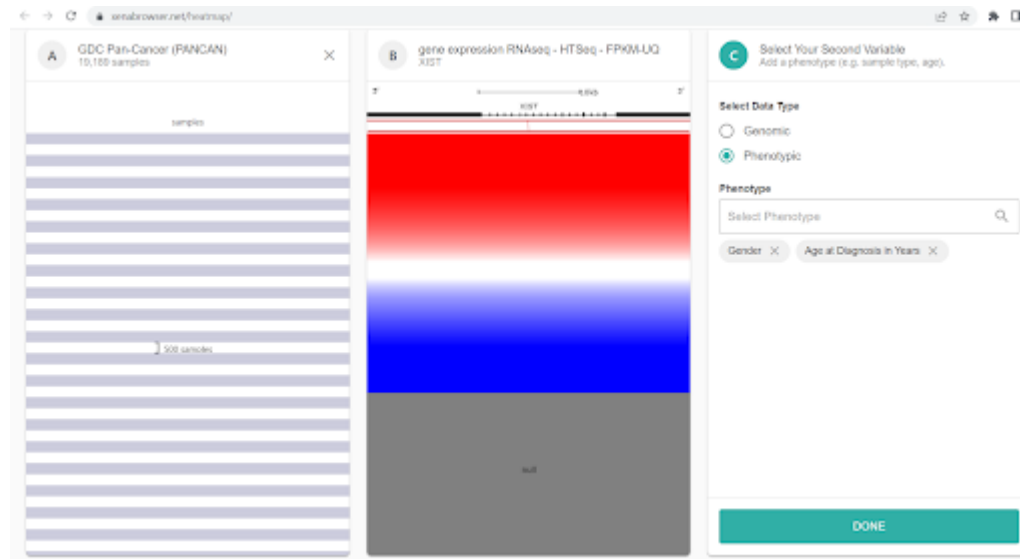
4. When you click the green "DONE" button, you will see two new columns on the side that show the "Gender" a
in Years" of the rows in the same order as the second box where your gene expression is plotted.
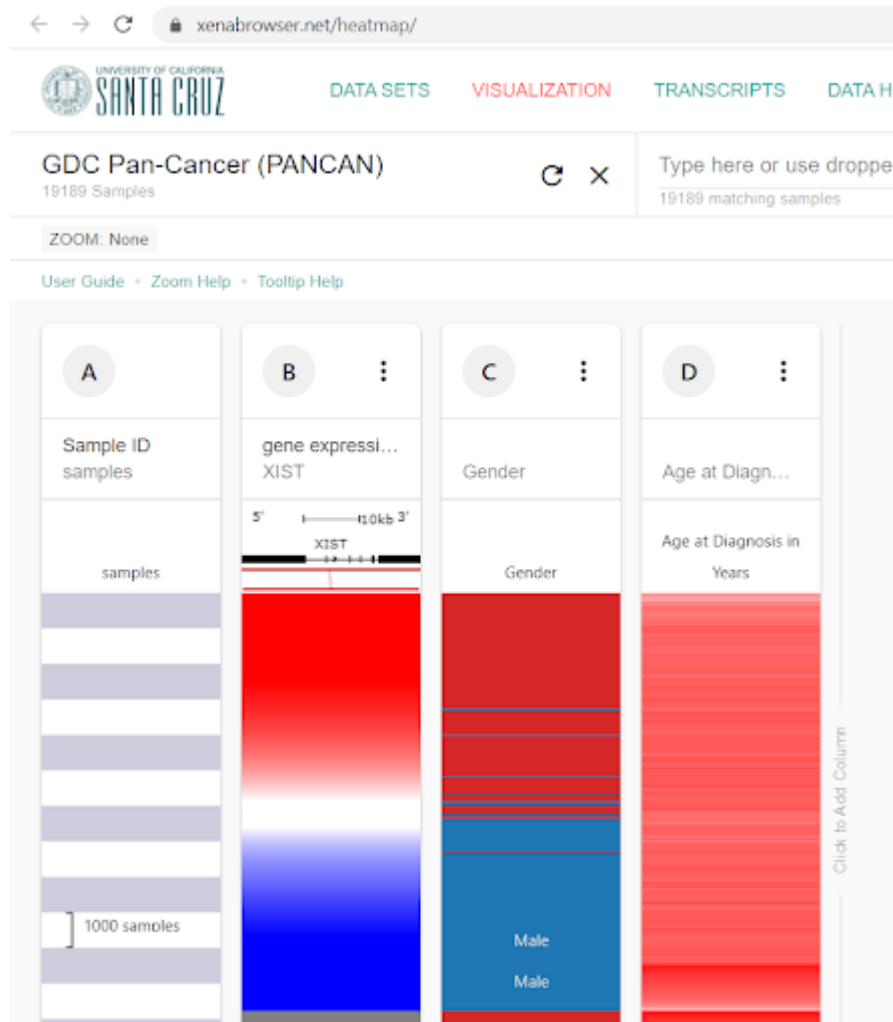
So here we can see that there are in fact some XIST high samples that have a "Gender" (reported sex of the can "Male". We can tell this because there are blue lines in the "Gender" in a region where the XIST expression in th (high). Also, the opposite is also true, there is one red line in the area that goes with low XIST expression indicati tumor sample that has low XIST expression. There doesn't appear to be a trend for age with XIST.

You can follow this same process to get a look at your assigned gene's expression.