Welcome to the Bioinformatics Research Team!

In our collaborative study on cancer genomics, we will investigate treatments for melanoma cancer using human tumors and mouse models. On this team, you are stepping into the role of a bioinformatics researcher and will be using specific skills and training to uncover the genetics of patient tumors. Before we dive into our research, let's first explore careers in bioinformatics and learn about what bioinformaticians do to contribute to a project like this one.

Careers in Bioinformatics

You are now a member of a collaborative group of professionals who work primarily with biological data.

Let's take a look at an example of a real bioinformatics research team: here we have a data analyst, a postdoctoral research associate, and an information technologist.

1. The **data analyst** works with research data. In their work, they typically write computer code or use specialized computer software to sort large data sets and identify patterns. They have a bachelor's degree but could also have advanced degrees or certificates to gain expertise and responsibilities in their career.
2. The **postdoctoral research associate** leads a specific research project. They can work solely on computational research using computers or could also work with samples in the laboratory. This team member has a bachelor's degree and a graduate degree (PhD).
3. The **information technologist** supports this computational team with their knowledge of computers and networks. An IT specialist makes sure the technology runs smoothly by maintaining the network, databases, and all hardware associated with computing. An IT specialist can have an associate degree or bachelor's degree.

As you can see, the bioinformatics research team relies on people with all kinds of training and experiences. You can explore careers and skills further with the resources on our Virtual Open House website.

Your Role as a Bioinformatician

Exploring genomics and finding driver mutations

As bioinformatics researchers, an important first part of our team's job is to understand the data with which we are provided. In our portion of the study, we've been given a group of patients who have agreed to let scientists use samples of their tumors to test an experimental melanoma treatment. Our goal is to understand more about the patients and their tumor samples they gave for the study (see **Figure 1**).

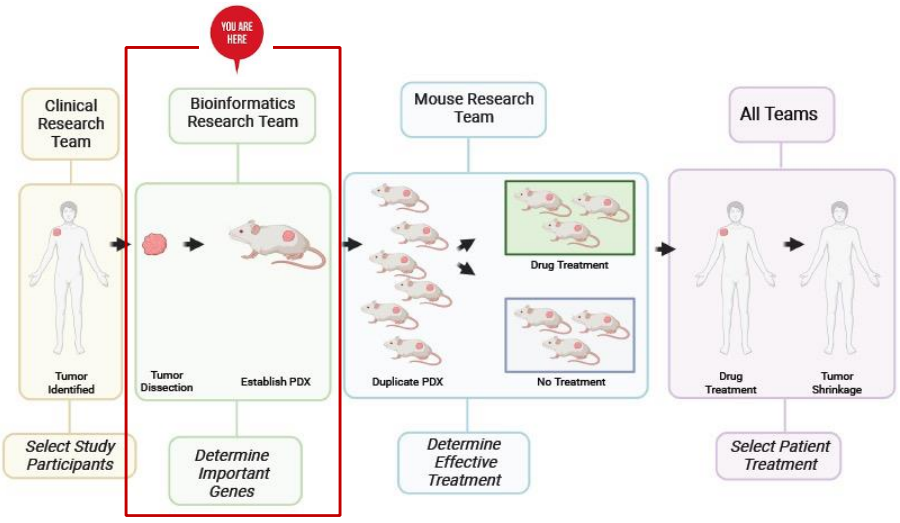


Figure 1. The bioinformatics research team analyzes the genomics of patient tumors.
In collaboration with the clinical and mouse research teams, the bioinformatics team plays an important role in this cancer genomics study by analyzing genomic data to identify mutations in the DNA sequence of patient tumors that could play a role in the progression of melanoma.

Specifically, as bioinformaticians, we are interested in the genetics of the tumors. Because we are equipped to analyze large datasets, we are interested in ALL the genes in a tumor. All the genetic material in a cell or organism is referred to as the *genome* and studying the genome is called *genomics*. One area of genomics is to identify changes in DNA sequence, which we often call *mutation*. Remember, mutations can lead to a cell becoming cancerous.

Researchers have found that a subset of the total mutations involved in cancer, called *driver mutations*, are directly responsible for uncontrolled cellular growth. Recall that cancer cells have genomic instability, so mutations can accumulate but many do not play a role in cancer progression. If a patient has hundreds of mutations, how do we know which ones are important for cancer? How do we identify the driver mutations? Let's explore these questions in our activity.

Bioinformatics Activity

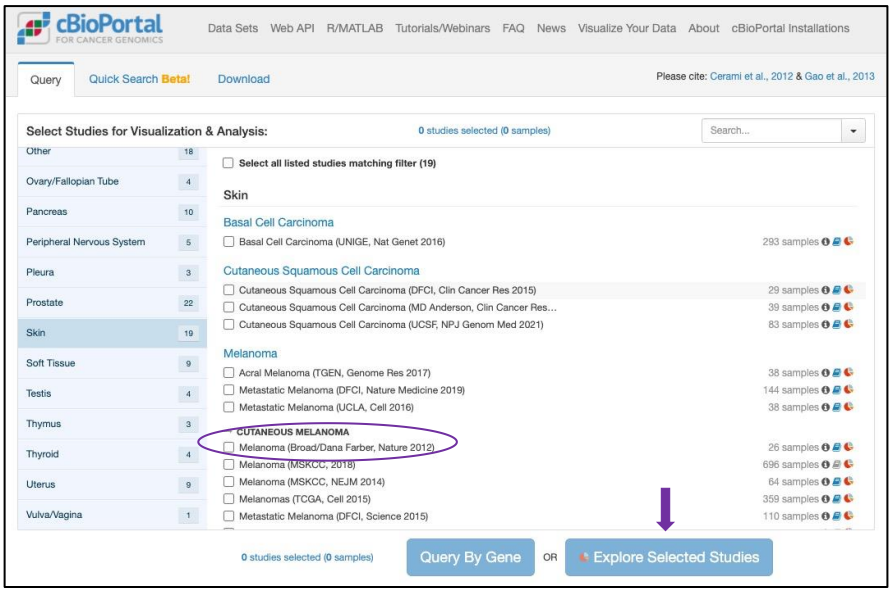
For this lab, use the **Activity Spreadsheet** to keep track of your work and document your findings. If you have already completed another lab, you can continue to use the same spreadsheet for this portion.

We will now go through the steps to explore the genetic data of our research study participants. You can use your spreadsheet to keep track of your work. Start by working in the "Bioinformatics (Parts 1&2)" tab.

Part 1. What genes are most frequently mutated?

Some genes are mutated in >50% of cancer patients. On a quest to identify driver mutations, one possibility to consider is if genes are frequently mutated in a specific cancer, they could be contributing to the progression of that cancer. Let's explore this idea and uncover which genes are frequently mutated in our melanoma patients.

1. In order to start our bioinformatics research, we first need to be introduced to the site where we can find genomic data about the patients who are enrolled in our study. To do this, we need to navigate to a cancer database called the cBioPortal.
2. Once you open this link, you are on the main page of the cBioPortal database. If you scroll down the center box, you will see a very long list of studies of many different types of cancers. The cBioPortal houses a lot of useful information within each of these studies, including facts about different patients as well as about their cancers, such as about the genetics of different tumors or response to different cancer treatments.
3. We are interested in skin cancer, so we will navigate to the skin cancer studies using the menu on the left and clicking on "Skin."
4. In the center box, find the study that is directly under "Cutaneous Melanoma" called "Melanoma (Broad/Dana Farber, Nature 2012)." To access this study, check the box next to the study and then click the blue button "Explore the Selected Study." You can also directly access the study using this link.



5. Once on the main page of the study, you can see a lot of information presented as graphs and tables about each of the 25 patients that are part of this study, including the genomic data that is relevant to our bioinformatics study as each patient had genetic testing on their tumors (cancer tissue).
6. In order to find the genes that are most frequently mutated in our melanoma study, locate the table that lists the "Mutated Genes" identified in the study. Record the top three genes* with mutations and the percentage of patients that have each mutation in the "Bioinformatics (Parts 1&2)" tab of your spreadsheet.

   *Genes are given name abbreviations that include a few letters and sometimes numbers. You can simply type the abbreviation of the gene name in the spreadsheet.

7. What do you notice about these genes and the frequencies of mutation? Can we conclude with certainty that these three genes are driving the progression of melanoma?

   We need more info! Let's dive deeper . . .

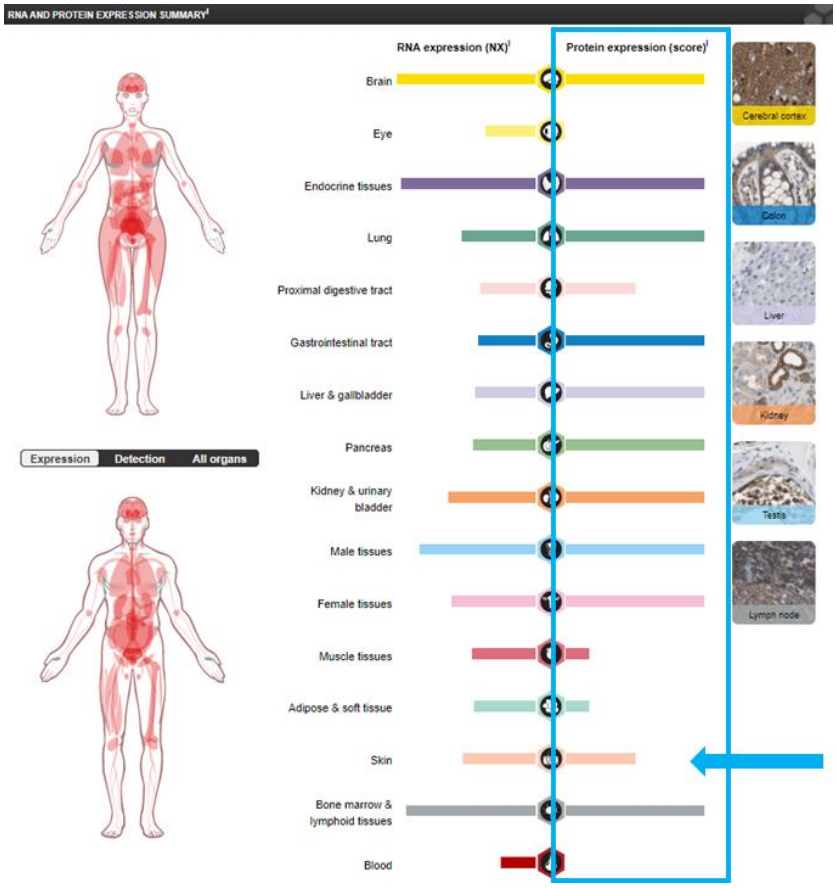
Part 2. Are the most frequently mutated genes expressed in the skin?

Now we have a sense of which mutations might be important based on mutation frequency in our patient group, but we still need to determine if these mutations drive melanoma (are responsible for progression of the cancer). Perhaps we could look at the typical expression of these genes in skin tissue.

One important thing to keep in mind is that even though we have the same DNA in most of our cells, different sections of the DNA are read and used to make protein in each type of cell. If a specific gene is read and used to make protein, we say it is "expressed." It is likely that some of our patients' mutations are not driving skin cancer because the genes that are mutated are not normally expressed in our skin. Let's explore the expression of the most frequently mutated genes in our study.

1. Navigate to the Human Protein Atlas webpage.
2. To use the database, type the name of one of the genes into the search bar and click on the matching gene name that comes up in the search. Then navigate to the tissue atlas.

3. Scroll down to look at the expression levels in different tissues in the body. We're specifically interested in protein expression, so we'll be looking at the bars on the right side of the graph.



4. In the "Bioinformatics (Parts 1&2)" tab of your spreadsheet, note at what level (not expressed, low, medium, or high) the corresponding protein for each gene is expressed in skin and add the level to your spreadsheet.

5. Repeat this for each of the most frequently mutated genes and record your findings.

So why aren't these genes expressed in the skin? Does this mean these mutations are not driving melanoma?

Many of the commonly mutated genes are very big. For example, the gene *TTN* is made up of approximately 224,000 nucleotides, while the average gene is made up of only around 28,000 nucleotides**! In fact, the gene is called "Titan" simply due to its size! Since cancer cells have a lot of genomic instability, larger genes are more likely to be mutated due as the result of chance. So, in this case, looking at frequency alone was not the best metric for identifying driver mutations.

***Source for TTN length and average gene length.*

Part 3. What mutations are driving melanoma in our patients?

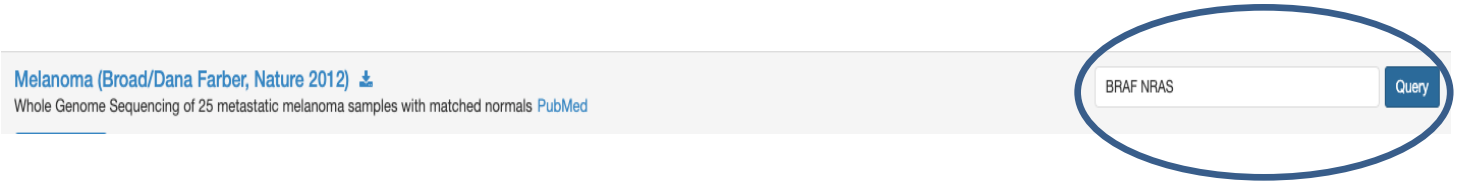
We need a new approach. We can identify genes that are likely driving melanoma by thinking about the function of the proteins that each gene codes for. For example, *BRAF* produces a protein which helps our cells to grow and divide. Since cancer involves uncontrolled cell division, *BRAF* could be driving melanoma in some of our patients. Another example is *NRAS*, which also codes for a protein involved in growth and cell division.

In fact, both *BRAF* and *NRAS* are known as *oncogenes* frequently driving melanoma. When *BRAF* or *NRAS* (or other oncogenes) are mutated, they can speed up cell growth and cause tumors to develop. To learn more about oncogenes and other types of mutations in cancer, you can watch this video from HHMI.

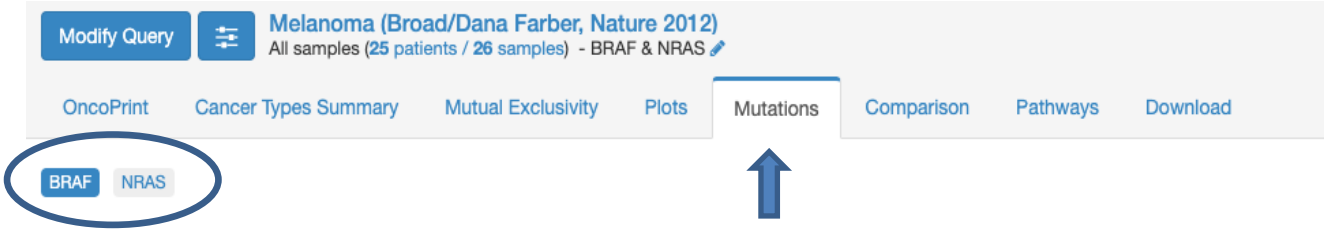
1. Navigate to the "Bioinformatics (Part 3)" tab of your spreadsheet, which contains a table with headers of the commonly mutated drivers of melanoma, *BRAF* and *NRAS*.

2. Return to our specific melanoma study in cBioPortal. Now we will look specifically at the genetics of the 25 patients included in this study.

3. Search for the gene names by typing both names into the search bar on the top right corner of the page and then hit "Query."



4. On the search results page, navigate to the "Mutations" tab. Record the patient ID numbers (patients have been deidentified and given numbers instead of including their names) for the patients that have a mutation in either of the driver genes *BRAF* and *NRAS* in the table on the "Bioinformatics (Part 3)" tab of your spreadsheet. You can use the gene name buttons on the left to toggle between the two genes.



5. You should now have two lists of patients, one with *BRAF* mutations and one with *NRAS* mutations.

Conclusions

Now we know what genes are frequently driving the progression of melanoma and which patients have each of these driver mutations. Next, let's think about our overall study goal – to find effective treatments to help the patients with these tumors. Our mouse research team will be testing two different therapies on each of the tumors isolated from 10 of these patients identified by the clinical research team.

Based on the genomics data we have just gathered about the study participants, we can make some predictions about the responses to treatments.

- What predications can we make about the responses of each group to each treatment based on this genomic data? Will the patients with the same driver mutations respond to the same treatments? Why or why not?

- What other information would you want to know to more accurately make a prediction of treatment response?

Awesome work! Our bioinformatics team really made a lot of research progress. We learned the types of jobs we could have on bioinformatics research teams, which range from different education paths and different skills. We also learned how to explore the genetics of specific patients by looking for driver mutations present in their tumors by navigating a real cancer database, the cBioPortal. You also learned how to correlate your findings with an independent line of evidence and narrow down possible driver mutations using the Human Protein Atlas database.

Reminder: Don't forget to check out the conclusions tab of your spreadsheet! Once you complete all three lab segments, you can see a summary of your combined work.

Extension into Ethics

Do genetics determine everything? Once a person gets their DNA sequenced, can others exploit this information? How will it be kept confidential?

Genetic testing is becoming more and more frequently used. For some people, like those in our research study, identifying variations in DNA sequence can be important for finding successful medical treatments. However, some individuals use direct-to-consumer genetic testing as a way to better understand their ancestry or find biological relatives. With the explosion of genetic testing companies over the last several years and increased access to these services, it's important to consider the ethical implications including how

this potentially sensitive information will be kept confidential as well as how to educate individuals on what the knowledge of their DNA sequence really means to their health.

Links to articles about genomics and ethics:

Genetic Information Non-discrimination Act of 2008
Informed Consent for Genomic Research

Case study 1: Megumi the data analyst

Megumi is a data analyst for a research lab that uses genomic data from hundreds of patients with a specific degenerative disease. At a scientific conference, Megumi meets a member of a collaborating lab who is interested in mining the genomic data samples Megumi studies for new gene variants potentially associated with a related, often co-occurring disease. Megumi hesitates and then remembers that all the identifying information about the patients, their names, ages, *etc.*, has been removed from the genomic sequences. Megumi does not see a problem with sharing the data with the collaborator.

Questions to Consider:

- Do you think that it is okay for Megumi to share the DNA sequences with her collaborator since the identifying information from the patients has been removed?
- What else should Megumi consider when making this decision?

Case study 2: Clyde the IT specialist

Clyde is a 24-year-old IT specialist with a large extended family. Over the last ten years, three of his aunts and uncles on his father's side have been diagnosed with pancreatic cancer that is difficult to treat. Clyde's father does not have cancer. Clyde's family doctor recommends at a routine physical examination that he go for genetic testing to see if he carries any gene variants associated with this cancer. The results of the genetic testing are mailed to his house and indicate that Clyde does indeed have an inherited variation in a gene increasing his risk for pancreatic cancer. Ever since he received his test results in the mail, Clyde has not been able to sleep and has been up all night worrying that he may get pancreatic cancer.

Questions to Consider:

- Who should Clyde talk to about his results?
- What other things can Clyde do to process this news?
- Do you think that, because he has a gene variant associated with pancreatic cancer, Clyde will definitely get cancer?

Case study 3: Cecile the postdoctoral researcher

Cecile is a postdoctoral researcher and for her birthday this year, her mother got her a direct-to-consumer genetic testing kit so she can learn about her ancestry as well as about her health. The test will provide information such as risk for conditions such as diabetes, heart disease, and cancer based on DNA sequence data. Cecile hesitates to use the kit

because she recently read that genomic data is frequently shared with health insurance companies and she is worried that she will not be able to get health insurance coverage if her DNA sequence reveals that she is at risk for one or more of these conditions.

Questions to Consider:

- Does Cecile have a valid concern?
- Should a health insurance company be able deny her coverage because of her DNA sequence?
- What can Cecile do before making a final decision?