

TEACHING THE GENOME GENERATION

cBioPortal Tutorials



cBioPortal Tutorials

Use the following menu to link to a specific topic within this tutorial:

- A. [Locating a cancer study](#)
- B. [Demographic data](#)
 - i. [Sex](#)
 - ii. [Race](#)
 - iii. [Age of diagnosis](#)
 - iv. [Median age of diagnosis](#)
 - v. [Age of diagnosis range](#)
- C. [Genetic data](#)
 - i. [Mutation count](#)
 - ii. [Median mutation count](#)
 - iii. [Mutation count range](#)
 - iv. [Mutated genes](#)
- D. [Individual patient data](#)
- E. [Sorting data within a study](#)
 - i. [Sorting by demographic data](#)
 - ii. [Sorting by gene name](#)
- F. [Interpreting mutation data](#)

A. Locating a cancer study

1. Navigate to the [cBioPortal](https://www.cbioportal.org/). This main page to the site lists the cancer studies entered in this database. Each study is slightly different, but all of them include genetic or genomic data about patients diagnosed with a specific type of cancer. Each patient underwent a form of genetic testing to be included in a study. The studies are organized by type of cancer, which means by tissue or organ.
2. In this tutorial, we will explore one study of patients with pancreatic cancer. Use the menu on the left side of the page and locate “Pancreas.” Once you narrow your search to studies involving the pancreas, you can then select a specific study “Pancreatic Cancer (UTSW, Nat Commun 2015).” Click the box to the left of the study name and then click the blue button at the bottom of the page “Explore Selected Studies.”

The screenshot shows the cBioPortal interface. At the top, there is a navigation bar with links for Data Sets, Web API, R/MATLAB, Tutorials/Webinars, FAQ, News, Visualize Your Data, About, and cBioPortal Installations. Below this is a search bar with a 'Query' button and a 'Quick Search Beta!' button. The main content area is titled 'Select Studies for Visualization & Analysis:' and shows a list of studies filtered by 'Pancreas'. The 'Pancreas' category is highlighted in the left sidebar. The list of studies includes 'Acinar Cell Carcinoma of the Pancreas', 'Cystic Tumor of the Pancreas', 'Pancreatic Adenocarcinoma', and 'Pancreatic Neuroendocrine Tumor'. The study 'Pancreatic Cancer (UTSW, Nat Commun 2015)' is selected, indicated by a blue checkmark. At the bottom of the page, there are two buttons: 'Query By Gene' and 'Explore Selected Studies', with a purple arrow pointing to the 'Explore Selected Studies' button.

3. Welcome to the summary page for the pancreatic cancer study. You will now see several visuals including charts, graphs, and tables about the study, including information about the patients as well as the genetic data collected about the patients. The rest of the tutorial will focus on different parts of the data and how to analyze and sort the data within a study.

See also, [cBioPortal Tutorial Series: Introduction & Homepage Navigation](#) video

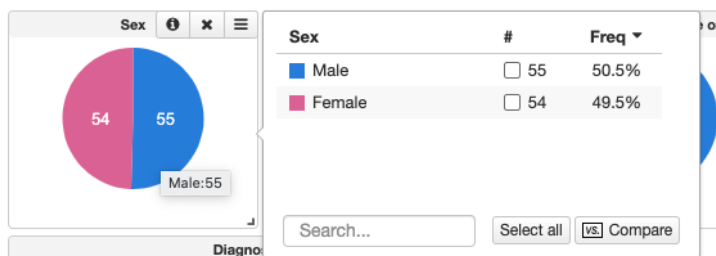
B. Demographic data

This section will demonstrate how to locate and analyze demographic study data. This tutorial will begin on the summary tab of the following [pancreatic cancer study](#).

See also, [cBioPortal Tutorial Series: Locating demographic data](#) video

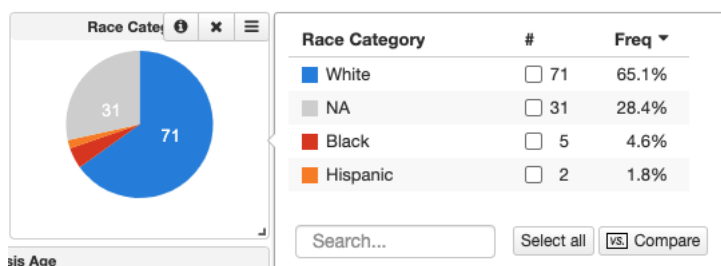
i. Sex

1. Locate the pie chart labeled “Sex,” which gives the breakdown of the assigned sex of the patients in the study.
2. In this study, you can see that there are 55 males (blue pie wedge) and 54 females (pink pie wedge).
3. If you hover over the pie wedges, a box will appear and this box provides the frequency for each assigned sex representing the percentage of patients for each. This study includes 50.5% males and 49.5% females.



ii. Race

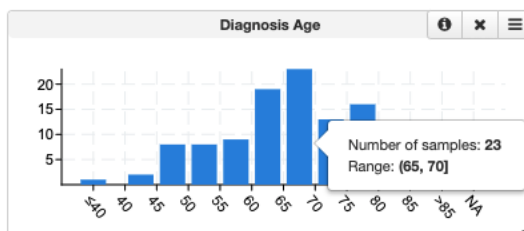
1. On the study summary page, locate the pie chart labeled “Race Category,” which gives the breakdown of the racial demographics of the patients in this study. The numbers within the wedges are the number of study participants who identified their race in a particular category.
2. You can hover over the pie chart to see the breakdown and racial frequencies. You can see that in this study, there are 71 people who identify as White representing 65.1% of the patients. 31 or 28.4% of the patients did not report a race and they are then included as “NA” on the chart and table.



iii. Age of diagnosis

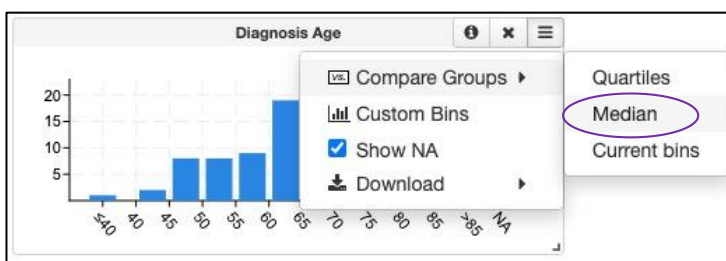
1. On the study summary page, find the histogram called “Diagnosis Age.”
2. First, try hovering over the different bars.
3. If you hover over the tallest bar, the box that pops up indicates that there are 23 people in this study who were diagnosed with pancreatic cancer between the ages of 65 and 70.

4. The bar labeled “NA” on the right represents two patients in the study without a reported age.

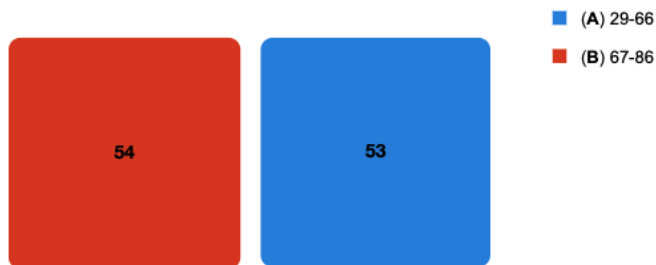


iv. *Median age*

1. Find the “Diagnosis Age” histogram and hover the cursor over the histogram so that a hamburger menu (three little lines) appears in the top right corner. Then hover over the hamburger to bring up a menu and select Compare Groups → Median.

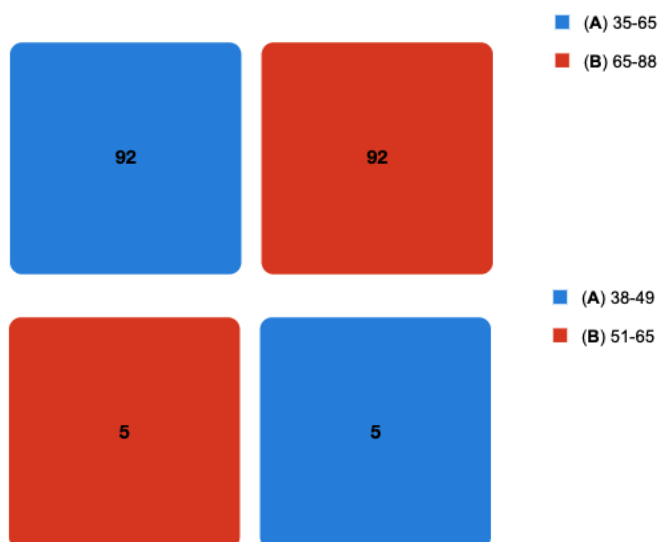


2. Selecting median then loads a new browser tab showing the individuals separated in two non-overlapping groups. One group, represented by the blue square includes 53 patients between the ages 29 and 66 years old. The second group, represented by the red square, represents 54 patients between the ages of 67 to 86 years old.



3. Notice that the median value is not listed anywhere but must be gleaned from this image. Because there are an odd* number of patients who reported age in this study (107 total), the median will be one of the individuals from larger group size. Because the larger group (red square) is the older-aged individuals, the youngest person in this group is the median. Therefore, the median age for this study is 67.

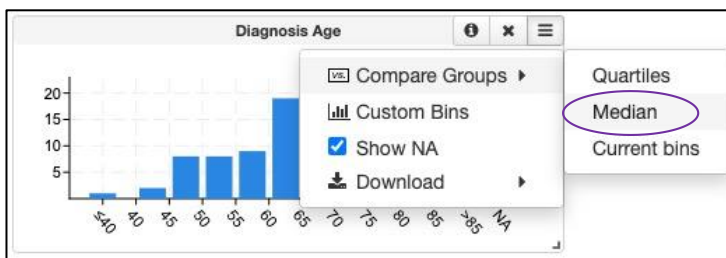
*If the study had an even number of patients, the calculation of the median would involve taking the average age of the youngest patient of the older group and the oldest person of the younger group. There are two examples below; the top one shows the median age is 65 with no calculation necessary. For the second one, the median is calculated by averaging 49 and 51, and therefore, is 50.



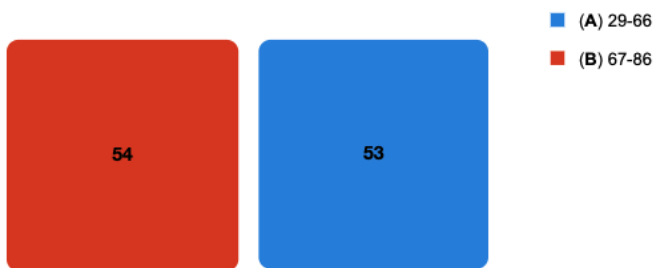
See also, [cBioPortal Tutorial Series: Calculating median diagnosis age](#) video

v. *Age range*

1. Find the “Diagnosis Age” histogram and hover the cursor over the histogram so that a hamburger menu (three little lines) appears in the top right corner. Then hover over the hamburger to bring up a menu and select Compare Groups → Median.



2. Selecting median then loads a new browser tab show the individuals separated in two non-overlapping groups. One group, represented by the blue square includes 53 patients between the ages 29 and 66 years old. The second group, represented by the red square, represents 54 patients between the ages of 67 to 86 years old.



- To find an accurate range for the age of diagnosis for a study, find the youngest age of the younger group (blue square) and the oldest age of the older group (red square). The range of this study is 29 to 86 years old.

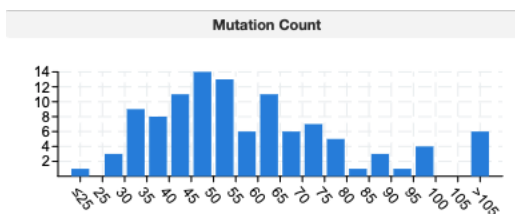
See also, [cBioPortal Tutorial Series: Calculating range of diagnosis age](#) video

C. Genetic data

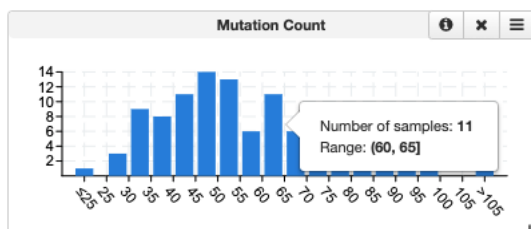
This section will focus on locating genetic data within a study. Begin on the summary tab of the following [pancreatic cancer study](#).

i. Mutation count

- Locate the “Mutation Count” figure on the study summary page. This histogram displays the mutation count for all 109 patients in this study. On this histogram, the number of mutations is labeled on the bottom, and the height of the bar indicates the number of patients who have the number of mutations listed. For instance, 15 patients have 45-50 mutations.

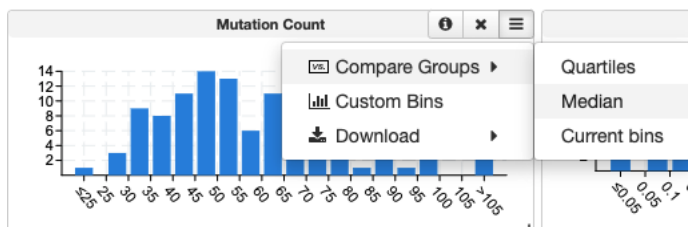


- To better visualize the number of patients who fall into each mutation count range represented by the bars of the histogram, hover over the bars to bring up a pop-up box. There are 11 people who have between 60 and 65 mutations.

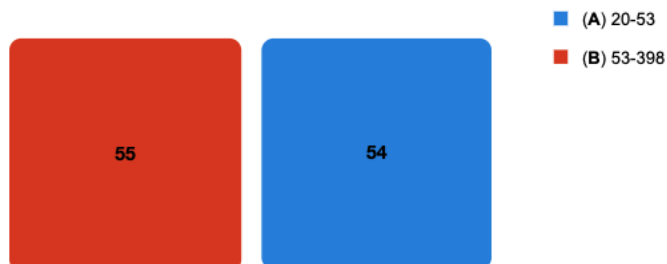


ii. Median mutation count

- Find the “Mutation Count” histogram and hover the cursor over the histogram so that a hamburger menu (three little lines) appears in the top right corner. Then hover over the hamburger to bring up a menu and select Compare Groups → Median.

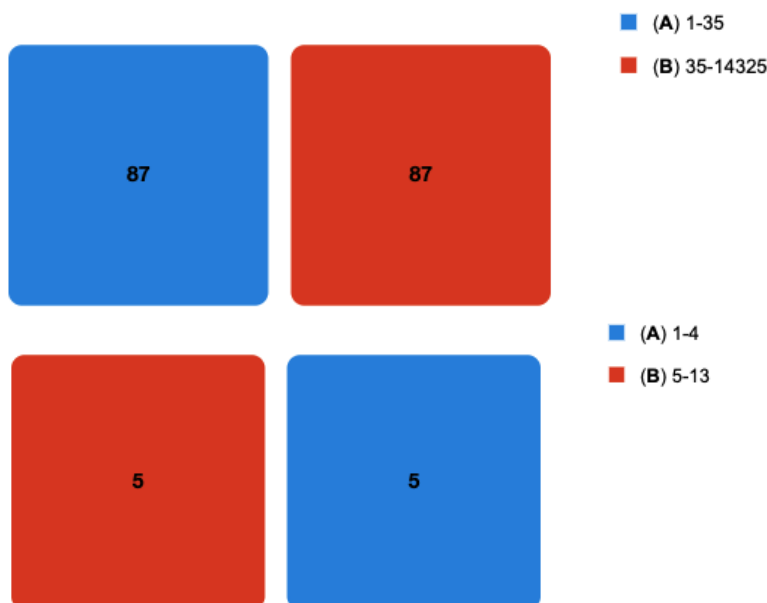


2. Selecting “Median” then loads a new browser tab showing the individuals separated in two non-overlapping groups. One group, represented by the blue square includes 54 patients with fewer mutations and counts of 20 to 53 mutations. The second group, represented by the red square, represents 55 patients with more mutations and counts of 53 to 398.



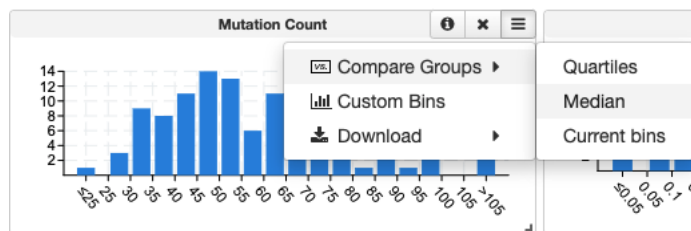
3. Notice that the median value is not listed anywhere but must be gleaned from this image. Because there are an odd* number of patients in this study (109 total), the median will be one of the individuals from larger group size. Because the larger group (red square) is the individuals with more mutations, the one with the least mutations in this group is the median. Therefore, the median mutation count for this study is 53.

*If the study had an even number of patients, the calculation of the median would involve taking the average of the lowest number of the group with more mutations and the highest number of the group with fewer mutations. In the two examples below, the median for the top one is 35 mutations, no calculation necessary. For the bottom one, the median would be the average of 4 and 5, or 4.5 mutations.

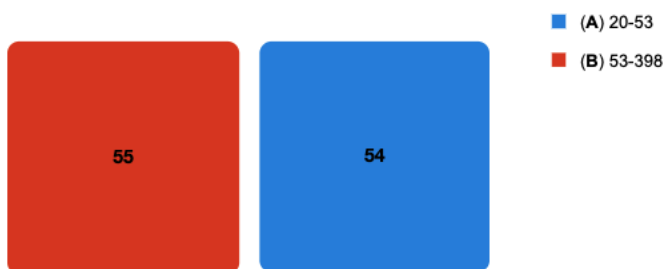


iii. *Mutation count range*

1. Find the “Mutation Count” histogram and hover the cursor over the histogram so that a hamburger menu (three little lines) appears in the top right corner. Then hover over the hamburger to bring up a menu and select Compare Groups → Median.



4. Selecting median then loads a new browser tab showing the individuals separated in two non-overlapping groups. One group, represented by the blue square includes 54 patients with fewer mutations and counts of 20 to 53 mutations. The second group, represented by the red square, represents 55 patients with more mutations and counts of 53 to 398.



2. To find an accurate range for mutation counts, find the lowest number of mutations in the fewer mutations group (blue square) and the highest number of mutations of the higher mutation group (red square). The range of this study is 20-398 mutations.

iv. *Mutated genes*

1. Locate the table labeled “Mutated Genes” to find a combined list of all the genes that were found to be mutated in the patients in this study.

Mutated Genes (109 profiled samples)			
Gene	# Mut	#	Freq
KRAS	103	100	91.7%
TP53	55	55	50.5%
SMAD4	22	21	19.3%
ZNF814	20	20	18.3%
MUC16	20	17	15.6%
TTN	19	15	13.8%

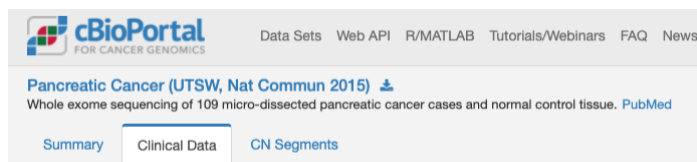
- To analyze this table, first look at the left-hand column labeled “Gene.” This is the gene name. Genes are typically written with an abbreviation of a combination of a few letters and/or numbers. The next column, “# Mut,” indicates the number of mutations in each gene identified in this study. The column “#” represents the number of patients in the study that have a mutation in that gene. The last column labeled “Freq” indicates the percentage of patients who have a mutation in that gene.
- The reason the value for “# Mut” may not match the “#” value is because more than one mutation may be present in one gene in the same patient. For instance, look at the gene KRAS. There were 103 mutations identified in 100 patients combined, meaning a few patients had more than one mutation in KRAS. 91.7% of the patients in the study had at least one mutation identified in KRAS.

See also, [cBioPortal Tutorial Series: Locating gene variants within a study](#) video

D. Individual patient data

This portion of the tutorial will focus on finding data for individual patients in a study. Begin on the main page of this [pancreatic cancer study](#).

- In addition to the study summary data, the cBioPortal houses data included about the individual patients included in this study. To locate this data, click on the second tab “Clinical Data” on the top-left corner of the study main page.



- This reveals a list of all the individual patients who are identified by a patient ID, such as “PDA_001,” the first patient listed in this pancreatic cancer study. Depending on the study, the information included in the clinical data will vary. In this particular study, the clinical data summary page includes information such as mutation count, demographic data, and other details of each patient’s cancer such as tumor stage.








Patient ID	Sample ID	Mutation Count	Fraction Genome Altered	Sex	Race Category	American Joint Committee on Cancer Tumor Stage Code	Nodal Status	Radiation Therapy	TNM Stage	Diagnosis Age
PDA_001	PDA_001	36	0.2004	Female	Black	pT3	1	Yes	2	51
PDA_002	PDA_002	35	0.0880	Male	Black	pT4	1	Yes	3	43
PDA_003	PDA_003	53	0.5361	Male		pT3	1	Yes	2	64
PDA_004	PDA_004	380	0.2701	Male		1	1	Yes	2	77

- To access more information about a specific patient, click on the patient ID number. If you select patient PDA_001 and click, this opens a new browser tab with more

information about this patient. The first thing you will see is a summary of the genomic data for this patient or the mutations identified in this patient's tumor.

- The "Summary" tab includes a list of all the mutations identified in this patient.

36 Mutations (page 1 of 4)

Gene	Protein Change	Annotation	Mutation Type
KRAS	G12V	   	Missense
KMT2D	L388Rfs*14	 	FS del
IL21R	R275Q		Missense

- The "Clinical Data" tab lists more detail about the patient and the "Pathways" tab shows the pathways in the cell that are affected by the mutations that this patient has.

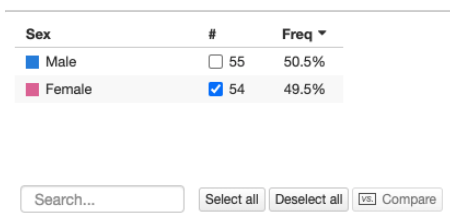
See also, [cBioPortal Tutorial Series: Locating individual patient data](#) video

E. Sorting data within a study

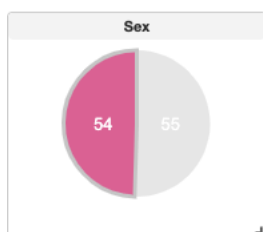
This section of the tutorial will focus on how to sort the larger study by specific criteria such as biological sex, race, or gene mutation using this [pancreatic cancer study](#).

i. Sorting by demographic data

- In order to limit the included information to a specific portion of patients within the study, it is necessary to sort the data.
- To sort by demographic data, such as biological sex, start on the main page of the study and locate the "Sex" pie chart.
- If we want to sort the study data so that it only includes data for assigned females in the study, hover over the pie chart and click the small box next to the number 54 for the females in the study.



- Once this box is selected, all the charts and tables on the summary page will refresh to display only the data for the assigned females in the study. Notice on the sex pie chart, the male wedge is now gray.



- Similarly, to sort the samples by race, hover over the “Race Category” pie chart and click the box next to the desired race to limit the data to only those patients of that race.
- It is possible to sort the data in several demographic categories at once. For instance, first sort study to include only females and then White people. The summary page of the study will now only display data for people who identify as White females.

ii. Sorting by gene name

- To sort the study data to only patients with a specific gene mutation, find the “Mutated Genes” table.
- To search for a specific gene, type the gene name into the search bar at the bottom of the box. For instance, type “CDKN2A” into the search bar.
- Check the box next to the number 6 and then click the button at the bottom of the table “Select Samples.” The study main page will update and now will only display data from those six patients with *CDKN2A* mutations.

Gene	# Mut	#	Freq
CDKN2A	7	6	5.5%

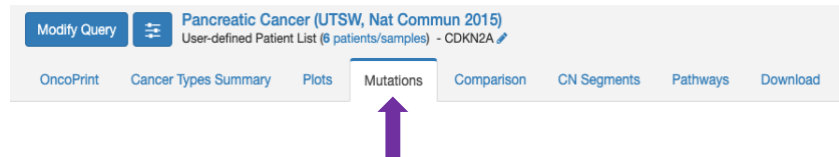
cdkn2a Select Samples

- Another way to look at all the patients with one mutation is to use the search bar in the top right corner of the study main page. First, return to the [unsorted study main page](#) and then go ahead and type “CDKN2A” into the search bar and then hit the “Query” button.

CDKN2A Query

samples Custom Selection Charts Groups

- This will bring up a new window with information about this gene relating to this pancreatic cancer study. By clicking on the “Mutations” tab, you will see a summary of the mutations identified in the patients in this study.



You will see a diagram of the gene and the locations of the mutations as well as a list of the mutations found in each patient.

See also, [cBioPortal Tutorial Series: Sorting a single study by gene](#) video

F. Interpreting mutation data

1. It is also possible to see exactly where in the gene the mutations occur. Sort the study using the “Query” search bar on the top right of the study main page (Tutorial Part E ii) to look only at *CDKN2A* mutations.
2. Use the “Mutations” tab to look closer at the specific mutations identified in this gene.
3. In biology, mutations in the coding regions of genes (the parts of a gene that make protein), are denoted in a conventional format that indicates the specific amino acid in the protein that was altered by the mutation and what the amino acid was altered to. Let’s look at a few examples for *CDKN2A*.

Missense mutations:

For patient PDA_031, the protein change listed is *H83Y*. This means that amino acid H in the 83rd position of the protein was mutated into amino acid Y. Therefore, this would be a missense mutation changing a histidine into a tyrosine amino acid*.

*Remember that mutations are in the DNA sequence, but in the case of a missense mutation, a change in DNA sequence changes the amino acid that piece of DNA codes for.

Nonsense mutations:

A nonsense mutation is indicated by an asterisk because the mutation creates a STOP codon. Looking at patient PDA_074, the protein change listed is *R80** meaning that the arginine amino acid at position 80 is now a STOP codon.

Frameshift mutations:

Frameshifts are a little more complicated and are indicated with “fs” for mutation type. For patient PDA_100, the frameshift mutation is listed as *E2Sfs*24* meaning that the mutation occurred at amino acid position 2, changing a glutamic acid to a serine amino acid and then amino acid 24 became a STOP codon (*).