

## 2 Navigating the NCBI Instructions

**Aim:** To become familiar with the resources available at the National Center for Bioinformatics (NCBI) and the search engine *Entrez*.



**Instructions:** Write the answers to your questions on the Student Worksheet, in your lab notebook, or on a separate sheet of paper, as instructed by your teacher.

1. Go to the NCBI homepage <http://www.ncbi.nih.gov/>

The screenshot shows the NCBI homepage. On the left, a vertical menu lists various resources such as 'All Resources', 'Chemicals & Bioassays', 'Data & Software', 'DNA & RNA', 'Domains & Structures', 'Genes & Expression', 'Genetics & Medicine', 'Genomes & Maps', 'Homology', 'Literature', 'Proteins', 'Sequence Analysis', 'Taxonomy', 'Training & Tutorials', and 'Variation'. On the right, a 'Popular Resources' section lists 'BLAST', 'Bookshelf', 'Gene', 'Genome', 'Nucleotide', 'OMIM', 'Protein', 'PubChem', 'PubMed', 'PubMed Central', and 'SNP'. A central search box is highlighted with a green border, containing the text 'Search All Databases' and 'Search Clear' buttons. A callout box on the left points to the 'All Resources' menu, and a callout box on the right points to the 'Popular Resources' section.

**Figure 1:** Familiarize yourself with the NCBI homepage. Credit: NCBI.



2. Take a few minutes to look around the site. The goal is to familiarize yourself with a few key components of the NCBI.
  - a. What is the name of one interesting resource or database shown in the blue box on the left? What do you think is its function or purpose?
  - b. What is one interesting resource listed in the Popular Resources menu on the right? What do you think is its function or purpose?
3. Find the search box in the center of the webpage (black box in above image). This search box uses the NCBI search engine *Entrez* to look for your search term (or "**query**") across **all** of the databases at the NCBI.

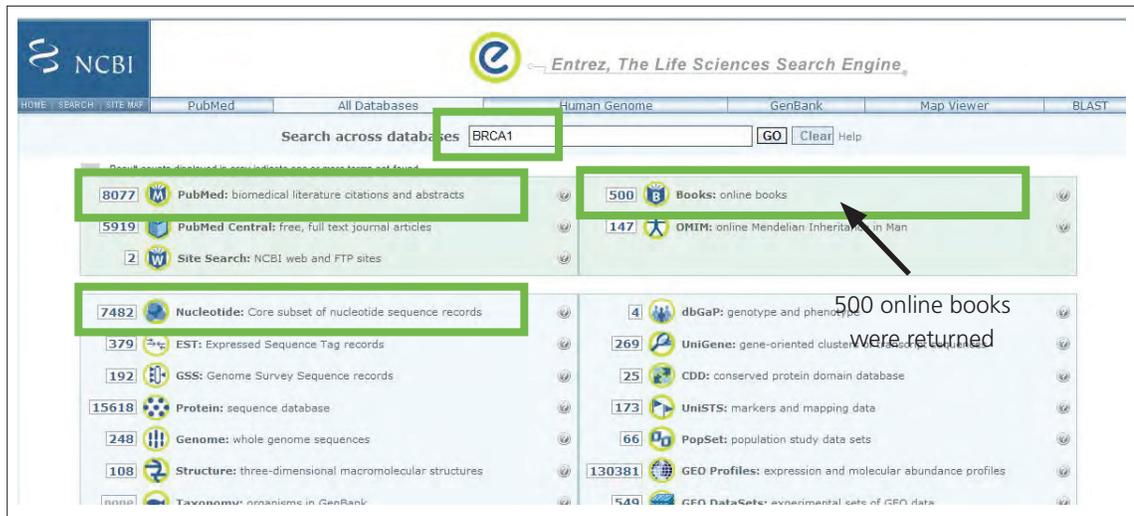


Figure 2: Search for “BRCA1.” Credit: NCBI.



4. Type “BRCA1” into the **Search** box. Make sure there is *no* space between BRCA and 1. Click **Search**.

*BRCA1* is a tumor suppressor gene that normally prevents cancer. Mutations in this gene are associated with increased risk of hereditary breast cancer and ovarian cancer when normal function is lost.

The white box to the left of each database contains the number of “hits” returned from that database (see screen shot, above). This is like searching in iTunes® without specifying categories like ringtones, podcasts, movies, TV, or songs.

- Why are we searching for *BRCA1*?
  - The **Nucleotide** database has DNA sequences that have been loaded onto the NCBI database. How many times is ‘BRCA1’ cited in the Nucleotide database?
  - The **PubMed** database has the articles that have been published about a specific gene or disease. How many times is ‘BRCA1’ cited in the PubMed database?
  - Compare the numbers you got for Questions **a** and **c**. Do these relative numbers surprise you? What does this tell you about the *BRCA1* gene? Explain.
5. Go back to the NCBI homepage by clicking the **NCBI** logo in the upper left corner of the screen.
- This search shows that there is a lot of information at the NCBI! It can be challenging to try to make sense of it all. Let’s start with something more familiar.
6. Click the “**All Resources**” link from the list of resources on the left side of the screen.
7. Find “**Map Viewer**.” Click on the “Tools” tab and either scroll through the alphabetical list, or use the “Find” feature (PC: “Control+F” Mac: “Command+F”) to Find “**Map Viewer**.” Click on the “**Map Viewer**” link.

The resulting page is called **Map Viewer** and it allows us to search the genomes of many different organisms, including humans.



Figure 3: Click “All Resources” on the left side of the screen. Credit: NCBI.

8. Open the Search menu, select *Homo sapiens* from the pull-down menu, and click "Go."



9. Now we can see the ***Homo sapiens* (human) genome view**. A genome is all of the genetic information in an organism. Each figure you see in the "genome view" represents a pair of chromosomes. Most of the chromosomes are numbered, but a few are not. The abbreviations "X" and "Y" refer to the human sex chromosomes.

a. How many different types of chromosomes do you see?

b. What does "MT" represent?

[Note: you can click the "MT" link to find out.]

c. With the exception of MT, the chromosomes of the human genome are in pairs. X and Y are a pair. Using this information and the information from your answer to Question 9A, how many **pairs** of chromosomes are in the human genome?



10. The Breast Cancer Susceptibility gene *BRCA1* is on chromosome 17 in humans. [Click on the link below chromosome 17.] Explore some of the links and views.

What do you see when you click on chromosome 17? Explore some of the links on the picture, and write down two things you found interesting, such as the description of other genes that are also found on chromosome 17.



11. To find the location of the *BRCA1* gene, type "BRCA1" in the "Search" box at the top left of the screen, and click "Find in This View." Scroll through the Map of Chromosome 17 and locate the *BRCA1* gene, which should be highlighted in pink. "BRCA1" will be found in the list of Symbols. You can also use the "Find" feature (PC: "Control+F" Mac: "Command+F"), which will highlight in yellow every mention of "BRCA1," including the *BRCA1* gene.

Draw a picture of chromosome 17 and show the *approximate* location of *BRCA1* on this chromosome.

Search pull-down menu

Select *Homo sapiens*

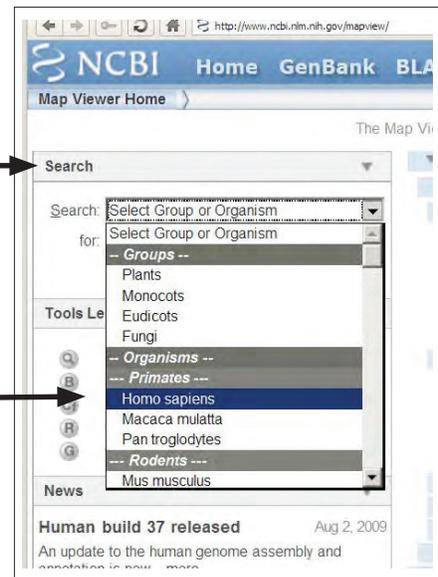


Figure 4: Select "*Homo sapiens*" from the list of groups or organisms. Credit: NCBI.

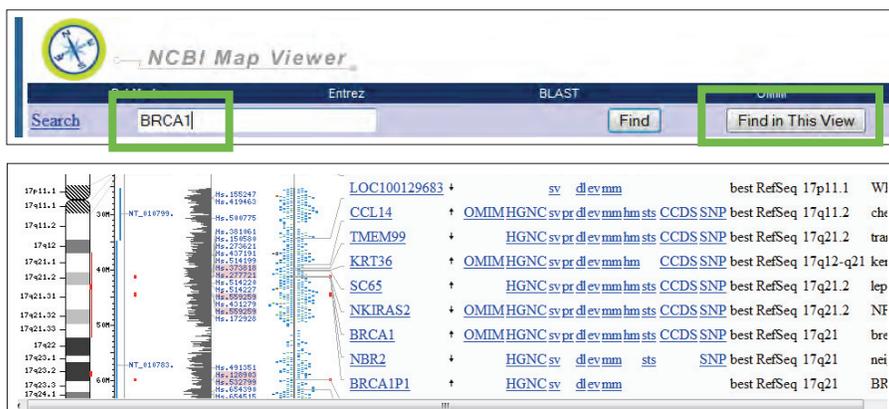


Figure 5: Find the location of the *BRCA1* gene by using the search function. Credit: NCBI.

12. Click on the **BRCA1** link. This will take you to **Entrez Gene**, which provides a summary of the information available at the NCBI for *BRCA1*. Scroll through the webpage and explore some of the information available. Scroll down the webpage to the section titled “Gene Ontology.” There is a table titled “Function.”



List three of the functions that the BRCA1 protein performs.

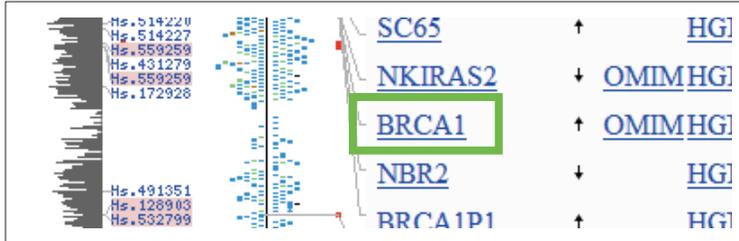


Figure 6: Click on the “BRCA1” link to launch **Entrez Gene**. Credit: NCBI.

Gene Ontology provided by [GOA](#)

Function	Evidence	
	Evidence Code	Pubs
<a href="#">DNA binding</a>	IEA	

Figure 7: Scroll down to find the “Gene Ontology” section. Credit: NCBI.

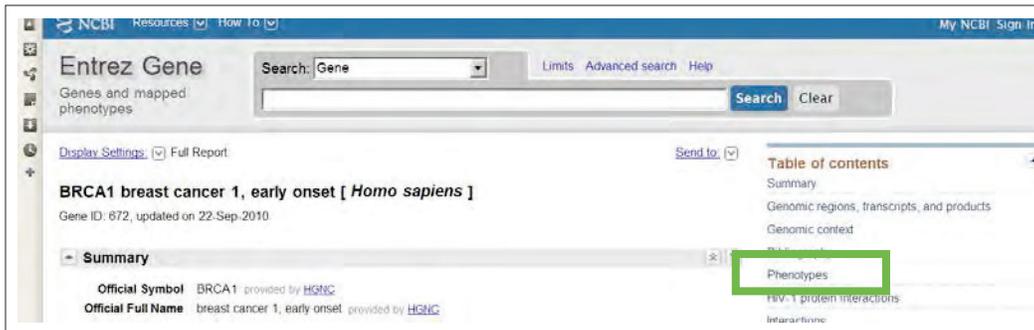


Figure 8: Select “Phenotypes” from the Table of Contents. Credit: NCBI.



13. To learn about all of the **phenotypes** associated with mutations of *BRCA1*, return to the top of the web page and from the “**Table of Contents**” on the right, select “**Phenotypes**.” This will bring you to the portion of the web page that contains the phenotype information for *BRCA1*.

- Based on what you’ve learned in class, what is a **phenotype**?
- What **phenotypes** are associated with mutations in the *BRCA1* gene? (You don’t need to click the links.)

14. Return to the **Table of Contents** at the top of the page and click “**Reference Sequences**.” This will take you to the portion of the webpage that contains the actual genetic sequence of the *BRCA1* gene.

15. **Reference sequences** are DNA or protein sequences that scientists, doctors and genetic counselors use to study genes like *BRCA1*. You can download these sequences in different formats. For this exercise, click “**FASTA**” (which is sometimes pronounced FAST-ay).



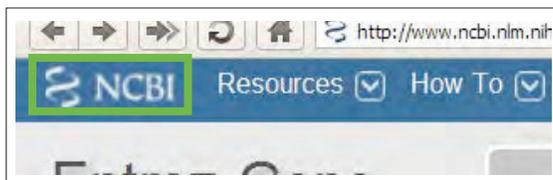
**Figure 9:** Click “FASTA” to obtain the FASTA sequence for *BRCA1*. Credit: NCBI.



**Figure 10:** Take a look at the FASTA sequence for *BRCA1*. Credit: NCBI.

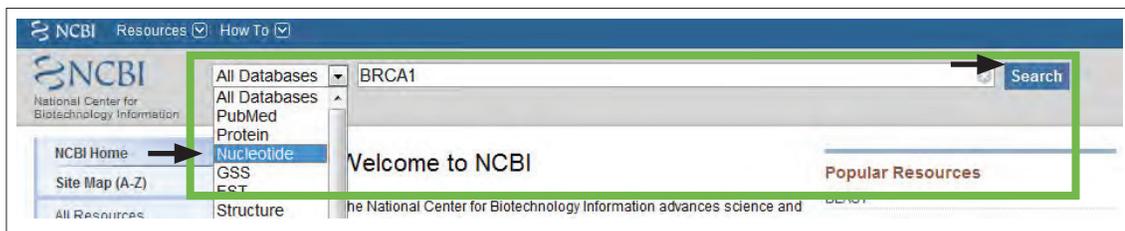


16. This link takes you to the FASTA sequence for *BRCA1*. Scroll through the web page. This gene is very large!
  - a. What four letters make up this long sequence?
  - b. Based on what you’ve learned in class, what do these letters represent?
17. Return to the NCBI homepage by clicking on the NCBI icon on the top left of the web page.



**Figure 11:** Return to the NCBI homepage by clicking on the logo. Credit: NCBI.

18. Type **BRCA1** in the **Search** box and select “**Nucleotide**” from the pull-down menu beside the **Search** box, to limit your search to the database containing all of the DNA and RNA (Nucleotide) sequences. Click the “Search” button.



**Figure 12:** Type “BRCA1” into the Search box and choose “Nucleotide” from the pull-down menu. Credit: NCBI.

19. What other organisms have *BRCA1* genes? You can scroll through the list of organisms, but note that these are listed by the **scientific name** of the organism (Genus and species), not the common name. For example, *Homo sapiens* is the scientific name for humans. Also, the **Top Organisms** (or the organisms with the most “hits”) are listed on the right. **Helpful Hint:** Hold your cursor over the species name to see the common name appear. Alternatively, you can perform an internet search to find the common name(s) of your organisms.



List three organisms other than humans that have *BRCA1* genes. Include both the scientific and common names.

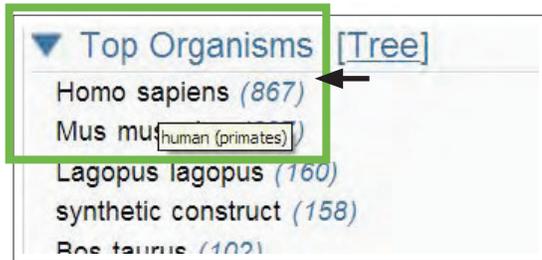


Figure 13: Scroll through the list of Top Organisms. Credit: NCBI.

20. Look back at your list of functions for the *BRCA1* gene (question #12).



Does it surprise you that so many organisms share the *BRCA1* gene? Explain.

21. What kind of information can you find at the National Center for Biotechnology Information?



Summarize what you have learned today by listing three types of information found at the NCBI.