2 Navigating the NCBI

Introduction

Students navigate parts of the **National Center for Biotechnology Information** (NCBI) website and work independently to explore **databases**, focusing on the *BRCA1* gene and the **bioinformatics** tool *Map Viewer*. Through an analogy that compares two collections of databases (iTunes® and the NCBI), students connect with their own prior knowledge to better understand database structure and function. In *Lesson Two*, students learn how ***veterinarians*** might use bioinformatics tools in their career.

Learning Objectives

At the end of this lesson, students will know that:

- Databases like those available at the National Center for Biotechnology Information (NCBI) are used to organize and search vast amounts of biological information.
- Genetic tests are developed using the biological information available in databases like those at the NCBI.
- Bioinformatics tools are used by people in many careers, including veterinarians.

At the end of this lesson, students will be able to:

- Navigate databases at the National Center for Biotechnology Information (NCBI) to find biological information.
- Explain the need for searchable databases as applied to genetic testing.

Key Concepts

- The NCBI is a central repository for many types of biological data freely available to scientists and the general public including: nucleotide and protein sequences, structures, scientific publications and much more. The many databases at the NCBI can be searched simultaneously using the NCBI's search engine ***Entrez***.
- Companies that offer direct-to-consumer genetic testing rely on biological information available through the NCBI.
- The NCBI resource *Map Viewer* allows users to view and search an organism's complete genome and display chromosome maps.
- Bioinformatics tools are used by and benefit people in many careers, including veterinarians.

Class Time

One class period (50 minutes).

Prior Knowledge Needed

- An introduction to taxonomy.
- An understanding of the relationship between genes and proteins.
- The ability to translate DNA nucleotide codons into amino acids.

Common Misconceptions

- Only humans have BRCA genes.
- BRCA genes are only located in breast tissue.

Materials

Materials	Quantity
Copies of Student Handout—*Careers in the Spotlight* (handed out in *Lesson One*)	1 per student
Class set of Student Handout—*Navigating the NCBI Instructions*	1 per student (class set)
Copies of Student Handout—*Navigating the NCBI Worksheet* [**Note:** This worksheet is for students' answers to lesson questions.]	1 per student
Teacher Answer Key—*Navigating the NCBI*	1

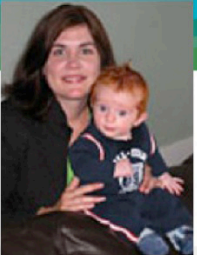| Computer Equipment, Files, Software, and Media |
|---|
| Computer with internet access and projector to display PowerPoint slides. |
| ***Alternative:*** Print PowerPoint slides onto transparencies and display with overhead projector. |
| *Lesson Two* PowerPoint Slides—*Navigating the NCBI.* Available for download at: *http://www.nwabr.org/curriculum/introductory-bioinformatics-genetic-testing.* |
| A student version of lesson materials (minus teacher answer keys) is available from NWABR's Student Resource Center at: http://www.nwabr.org/students/student-resource-center/instructional-materials/introductory-bioinformatics-genetic-testing. |
| Computer lab with internet access for students. |

Teacher Preparation

- Load the classroom computer with the *Lesson Two* PowerPoint slides.
- Make copies of the Student Handout—*Navigating the NCBI Instructions*, one per student. This handout is designed to be re-used as a class set.
- Make copies of the Student Handout—*Navigating the NCBI Worksheet*, one per student. This worksheet is used for students to write their answers to the lesson questions.

Procedure

WARM UP

1. As students enter the classroom, display the PowerPoint **Slide #1**. This slide highlights veterinarian Deborah Tegarden, DVM.

Navigating the NCBI: **Slide #1**

VETERINARIAN
DEBORAH TEGARDEN, DVM

Place of Employment:
Elliot Bay Animal Hospital
Seattle, Washington

"While we often think of genetic testing in humans, vets are seeing more and more tests being developed for animal patients. This is the most exciting time I can imagine in veterinary medicine, when things are getting more and more cutting edge and technology is developing at lightning speed."

2. Have students retrieve Student Handout—*Careers in the Spotlight* from *Lesson One*.

3. Students should think about, and write down, what kind of work a veterinarian might do (Veterinarian Question #1). This will be revisited at the end of the lesson, including how a veterinarian might use bioinformatics in his or her job.

4. Tell students to keep their *Careers in the Spotlight* handout available for future lessons.

PART I: What does the Gene Machine really do?

5. Explain to students the ***aim of this lesson***. Some teachers may find it useful to write the aim on the board.

   Lesson Aim:
   • To understand the role of databases in organizing and searching for biological information.

6. Remind students that, in Lesson One, they watched the play *Meet the Gene Machine* and explored the 23andMe website to learn more about genetic testing.

7. Share with students that in today's lesson, they will explore more about the inner workings of the "Gene Machine" by learning about the role bioinformatics plays in genetic testing. Another goal for today's lesson is to explore how databases, particularly those at the National Center for Bioinformatics Information (NCBI).

8. Ask students, "What does the Gene Machine ***do***?" It may be helpful to walk over to the prop used for the Gene Machine in the play from *Lesson One*, if it is available.

9. Let students brainstorm answers. Students may say that it "analyzes DNA," or "tests for genetic diseases," or "gives a person a printout of his or her genes." For each case, ask students ***how*** this might happen (i.e., "***How*** do you think it analyzes DNA?").

10. Highlight for students that it is not enough to know just the patient's DNA sequences for a genetic test. A genetic test must ***compare*** the patient's DNA sequences to known sequences. These known sequences are called **reference sequences**.

11. Tell students that the fictional Gene Machine from the play represents a number of steps necessary for genetic testing, from purifying and sequencing the DNA to comparing the patient's DNA with a known reference sequence.

PART II: The Need for Databases

12. Share with students the importance of databases in science: **Databases** allow scientists to store, manage, and retrieve information in an organized way.

13. Remind students of how much genetic information was returned for just one person in the play, *Meet the Gene Machine*. How would scientists handle the amount of data if everybody in the class got tested? Everybody in the school? Everybody in the school district? Where are all the reference sequences kept and how do scientists access them?

[**Note:** Teachers may also wish to discuss the Learning Objectives of the lesson, which are listed at the beginning of this lesson.]

Reference sequence: A sequence that has been chosen for the purpose of comparison. At the National Center for Biotechnology Information (NCBI), reference sequences are chosen because they are of high quality and are thought to accurately represent the sequence from the original organism. In genetic testing, a reference sequence is a known and well-studied DNA or protein sequence that does not contain any mutations or changes that are associated with disease. These sequences are used for comparison with patient sequences during genetic testing.

NCBI has a database dedicated to reference sequences, called the RefSeq database. The goal of RefSeq is to accurately represent all naturally occurring DNA, RNA, and protein molecules for major organisms.

Database: A collection of related data that are stored, managed, and retrieved in an organized way.

National Center for Biotechnology Information (NCBI): Part of the National Library of Medicine at the National Institutes of Health (NIH), the NCBI is a collection of biological information organized in over 30 cross-referenced databases related to genetics and molecular biology. It's not just DNA – there are all kinds of data about how cells, organisms, and even ecosystems function.

Entrez: The search engine used to simultaneously search all of the databases at the NCBI. Items stored in different databases are cross-referenced and inter-linked, making it easier to find all the database records that are related to any one subject.

Nucleotide database: One of the databases at the NCBI which contains nucleotide (DNA and RNA) sequences.

14. Tell students the following:

This information, and more, is stored in biological databases such as those at the **National Center for Biotechnology Information (NCBI)**, which is part of the National Library of Medicine, which in turn is part of the National Institutes of Health.

The NCBI houses biological information in over 30 databases related to genetics and molecular biology. All of these databases can be searched using one search engine (**Entrez**). Items stored in different databases are cross-referenced and inter-linked, making it easier to find all the database records that are related to any one subject. When students visit the NCBI in this lesson, they will primarily use the **Nucleotide database**.

Genetic tests like those available through 23andMe could not exist without the information available at the NCBI.

15. Ask students if they are familiar with using databases to search for and retrieve information. It is likely that many students have used iTunes®, a collection of databases related to music, podcasts, videos, and ringtones. iTunes® can be searched in multiple ways and can yield many results from one query, just as students will find at the NCBI.

16. Tell students the following:

Biology is undergoing a tremendous change because of the explosion of data that are being collected.

Because DNA is an information molecule, and we are gathering more and more genetic information, we need new strategies to understand the data.

NCBI can be thought of as a point for "one stop shopping" when searching for biotechnology information. The databases found at NCBI are crucial for scientists conducting biological research.

PART III: What is Bioinformatics?

17. Tell students that **bioinformatics is the application of computer science and information technology to biology and medicine.** Teachers are encouraged to write this definition on the board, and ask students to write it in their notes. The definition can also be found at the top of students' *Careers in the Spotlight* handout.

18. Break this definition down to look at its parts. Bioinformatics is used to:

Create databases: Building cross-referenced, interlinked databases used to store information.

Search databases: Designing a way to search and retrieve information from the databases. When students go to the NCBI and search the databases, they are using the tools of bioinformatics.

Compare sequences: Comparing an individual's DNA to reference sequences that are uploaded and stored in the NCBI databases. These reference sequences have been reviewed by NCBI researchers to confirm that they contain no errors or mutations. They are the result of much collaboration between scientists. The programs used to compare these sequences are the work of bioinformatics.

Represent molecular structures: Identifying and representing the three-dimensional structure of biological molecules such as proteins. Computer programs that determine and represent these 3D structures are the work of bioinformatics.

19. Tell students that bioinformatics offers many different types of tools to use in different circumstances. Like any set of tools, bioinformatics is useful when applied.
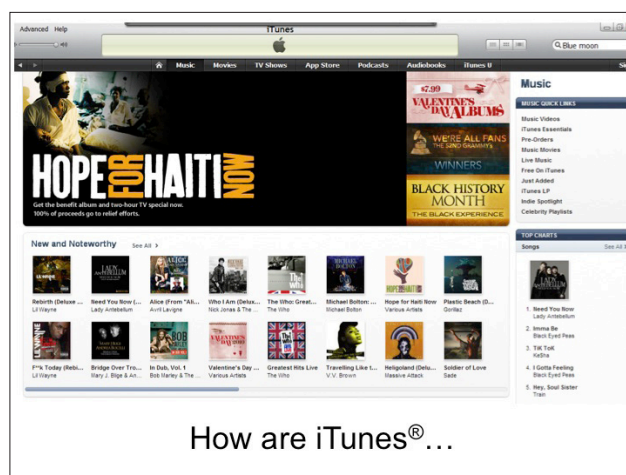
PART IV: Navigating the NCBI

20. Tell students that while they are at the NCBI website, they will focus on a gene they explored in *Lesson One: BRCA1*. Remind student that *BRCA1* plays a crucial role in DNA repair as a tumor suppressor that normally prevents cancer.

21. Acknowledge that students may feel overwhelmed by the amount of information at the NCBI. The purpose of this exercise is ***not*** for students to learn detailed, specific information from the site, but to gain an appreciation for the amazing depth and breadth of information available at the NCBI.

22. Distribute copies of Student Handout—*Navigating the NCBI Instructions* and copies of Student Handout—*Navigating the NCBI Worksheet* and allow students to start working on the activity independently at a computer.

PART V: A Database Analogy: iTunes® and the NCBI

23. After students have worked through Student Handout—*Navigating the NCBI Instructions*, tell them that they will now tie what they have learned about databases at the NCBI to what they already know about the iTunes® database. It may be helpful to ask students if they are familiar with the music database iTunes®, and ask them what they know about the service. For example, students might say that they use iTunes® to download songs, to find out which songs were published by which artist, or to learn which album a particular song was released on.

24. Show ***Slide #2***, "How are iTunes®…"

[**Note:** While working through Student Handout—*Navigating the NCBI Instructions*, students may feel overwhelmed with the information available in the databases. Working through it as a group or as a class may be helpful. It is important to realize that neither teachers nor students are going to understand everything at the NCBI. Students and teachers are encouraged to learn together. The NCBI contains a great deal of information and is updated regularly, making learning an on-going process.]

Navigating the NCBI: ***Slide #2***

How are iTunes®…

25. Show **Slide #3**, "*and the NCBI…similar and different?*"

[**Note:** You may wish to have students follow along with you by taking their own notes. They too can write "**iTunes®**" at the top of one column, and "**NCBI**" at the top of the second column. Rows on the top half of their paper can contain "**Similarities**," while rows on the bottom half of the page can be for noting "**Differences**." Under the columns, they can write down what these two databases have in common, and then how these two databases differ. This is also a useful way for students to take notes during the iTunes® versus NCBI brainstorming session.]

26. In pairs, have students brainstorm ways in which iTunes® and the NCBI are similar and ways they are different. Students could think about search functions, content, cost, and the type of results returned.

27. To review with students the results from their brainstorming session, **Slides #4–6** may be used. Alternatively, teachers may wish to draw tables on the board, and have students supply answers from their brainstorming. If using PowerPoint slides, project **Slide #4** "*iTunes® and the NCBI*" for all to see, and review with students the similarities between the two databases. If writing on the board, write "iTunes®" at the top of one column, and write "NCBI" at the top of the other column, listing the similarities, and then the differences, between the two databases. In addition to student answers, **Table 1** and **Table 2** on the following pages may be used to fill in the "similarities" and "differences;" write on the board only the bolded words found in the tables on the following pages, and discuss with students the ideas in italics.

28. If using PowerPoint slides, next show *Slide #5*, which illustrates additional ways in which the NCBI and iTunes® are similar.

iTunes® and the NCBI *are* similar because they both…

iTunes®   **NCBI**

Yield Large Numbers of Results

Searching for a song title may yield a large number of closely-related results.

Searching for a DNA sequence may yield a large number of closely-related results.

Provide Additional Information

Each song search result includes information about the artist, album, release date, and more.

Each sequence search result includes information about the organism, location of sequence, date of database entry, and more.

Table 1: How iTunes® is Similar to the NCBI.

iTunes®	NCBI
Contains multiple databases *iTunes® houses several organized collections of certain types of "data" (music, ringtones, videos, TV programs, and podcasts).*	**Contains multiple databases** *NCBI houses several organized collections of certain types of "data" (biological information).*
Uses search functions *The iTunes® music database can be searched in a variety of ways, such as by song title, album title, artist, and more.*	**Uses search functions** *The NCBI Nucleotide sequence database can be searched in a variety of ways, such as by topic, DNA sequence, accession number, author, and more.*
Yields large number of results *Searching for a song title may yield a large number of closely-related results.*	**Yields large number of results** *Searching for a DNA sequence may yield a large number of closely-related results.*
Provides additional information *Each song search result includes information about the artist, album, release date, and more.*	**Provides additional information** *Each sequence search result includes information about the organism, location of sequence, date of database entry, and more.*

Lesson 2 – Navigating the NCBI

29. If using PowerPoint slides, next show **Slide #6**, and discuss with students ways in which iTunes® and the NCBI are different.

iTunes® and the NCBI are *not* similar because they both…

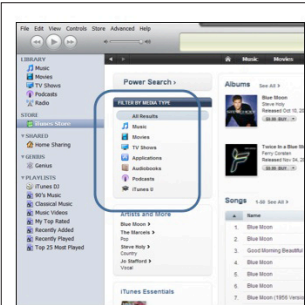| iTunes®	NCBI
Search Scale	
*The user **cannot** search for an exact sequence of notes that occur in more than one song in the music database.*	*The user **can** search for an exact DNA sequence that occurs in many organisms in the DNA Nucleotide database.*
Content	
iTunes® provides audio and audiovisual content.	*NCBI provides biological content.*
Cost	
iTunes® charges a fee for downloading most items.	*Use of the NCBI is free! It is paid for by the federal government through tax dollars.*

Table 2: How iTunes® is NOT Similar to the NCBI.

iTunes®	NCBI
Search scale—can't search by musical note sequences *The user cannot search for an exact sequence of musical notes that occur in more than one song in the music database.*	**Search scale—can search by DNA sequences** *The user can search for an exact DNA sequence that occurs in many organisms or genes in the DNA Nucleotide database.*
Content—audio and visual *iTunes® provides audio and audiovisual content.*	**Content—biological** *NCBI provides biological content.*
Cost—fee-based *iTunes® charges a fee for downloading most items.*	**Cost—free** *Use of the NCBI is free! It is paid for by the federal government through tax dollars.*

30. Show **Slide #7**. Draw students' attention to the first similarity between iTunes® and the NCBI (contains multiple databases). Tell students that the three databases they will be using the most during this unit are the **Nucleotide**, **Protein**, and **Structure** databases. Emphasize for students the types of information that are available in each database.

Two types of searchable databases…
not so different.

- **Nucleotide** database—this contains nitrogenous base sequences written in the form of A (adenine), T (thymine), C (cytosine), G (guanine), or U (uracil). Nucleotides are the building blocks of DNA and RNA. Three nucleotides code for one amino acid.

- **Protein** database—this contains the amino acid sequences written using the one-letter abbreviations for each of the 20 amino acids. For example, the amino acid *glycine* is abbreviated *G* and the amino acid *alanine* is abbreviated A. Amino acids are the building blocks of proteins.

- **Structure** database—this contains the three-dimensional representations of the protein structures. The amino acid sequence for the protein is also included in the structure database. Structures can be viewed with molecular viewing programs like **Cn3D**.

PART VI: Closure—Careers in the Spotlight

31. Check for understanding of the student handout. Make sure to reinforce the major concepts:

- The NCBI is the "wizard behind the curtain" that provides biological information to 23andMe and other direct-to-consumer genetic testing companies, as well as to research scientists and others studying genetic diseases. These companies have their own databases as well, which they add to with patient samples that they analyze.

- NCBI has a number of different databases that can be searched using the search engine Entrez.
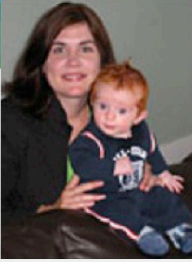
32. Check for understanding about the BRCA genes. Ask, "Do all people have the *BRCA1* and *BRCA2* genes?"

"Yes, we do." It is important for students to realize that all people—and in fact many other species—have both *BRCA1* and *BRCA2* genes—as students saw in Student Handout—Navigating the NCBI Instructions. The BRCA1 and BRCA2 proteins play crucial roles in DNA repair. However, some individuals have inherited BRCA alleles that contain mutations. Those mutations can encode defective proteins and are associated with increased risk for cancer.

33. Inform students that in *Lesson Three* they will read a case study that introduces students to a family with a history of breast cancer, and focuses on some of the ethical principles involved deciding whether to proceed with genetic testing.

34. Return to the picture of the veterinarian from the *Careers in the Spotlight* **Slide #8**.

Navigating the NCBI: **Slide #8**

VETERINARIAN
DEBORAH TEGARDEN, DVM

Place of Employment:
Elliot Bay Animal Hospital
Seattle, Washington

"While we often think of genetic testing in humans, vets are seeing more and more tests being developed for animal patients. This is the most exciting time I can imagine in veterinary medicine, when things are getting more and more cutting edge and technology is developing at lightning speed."

35. Show **Slide #9**, which provides job information for a veterinarian. Review this information with students.

Navigating the NCBI: **Slide #9**

CAREERS IN SPOTLIGHT:
VETERINARIAN

What do they do?
Veterinarians diagnose and treat animals. Some veterinarians specialize in a particular area (such as oncologists who treat cancer), and some perform research to improve animal and human health. Veterinary technicians assist veterinarians in their work.

What kind of training is involved?
Veterinarians complete a Bachelor's degree and a DVM (Doctor of Veterinary Medicine) degree, which requires four years. Veterinary technician training is usually a two year (Associate's) program.

What is a typical salary for a Research Scientist?
Veterinarians: $45,000/year ($22/hour), up to $140,000/year ($67/hour). Veterinary Technicians: $20,000/year ($10/hour), up to $45,000/year ($22/hour).

36. Ask students, "How does a veterinarian's job relate to today's lesson?" Point out that:
   • A number of genetic tests have been developed for animals.
   • Bioinformatics tools can be used to conduct research on opportunities for animal breeding programs.
   • As students saw today, much of the biological information at NCBI is gathered from animals, and veterinarians, as well as research scientists, play a role in this.

37. Ask students to answer Veterinarian Question #2 on their *Careers in the Spotlight* handout, which has students explain how this lesson has changed their understanding of the kind of work a veterinarian does.

38. Ask students to also answer Veterinarian Question #3 on their *Careers in the Spotlight* handout, which has students explain how a veterinarian might use bioinformatics in his or her work.

39. Tell students to keep their *Careers in the Spotlight* handout available for future lessons.

Homework

The following are suggested homework activities to follow this lesson. It is highly recommended that Student Handout—*Case Study: A BRCA Genetic Testing Dilemma* from *Lesson Three* be provided as homework to allow more time in class to discuss the case study.

A. As homework, ask students to write about the activities they learned in *Lesson Two* in their lab notebooks, on another sheet of paper, or in a word processing program like Microsoft Word® or Google Docs which they then provide to the teacher as a printout or via email. This can serve as an entry ticket for the following class. Have them complete these prompts:

   a. Today I learned that…

   b. An important idea to think about is…

   c. Something that I don't completely understand yet is…

   d. Something that I'm really confident that I understand is…

[**Note:** Suggested scoring for homework: +5 points if all 4 prompts are complete.]

B. Have students return to the NCBI homepage and search for something of interest to them, such as a disease, an animal, or a gene they've heard about in class or in the news. Based on the Entrez search results, what can students conclude about how much this topic has been studied? Do any of the search results surprise them? Encourage them to explore some of the links to their search results, such as the Nucleotide sequences or scientific articles found in **PubMed** Central [which are all freely available], and describe what they've found.

[**Note:** Suggested scoring for homework: Up to +10 points.]

C. Distribute Student Handout—*Case Study: A BRCA Genetic Testing Dilemma* from *Lesson Three* as homework before the class discussion session. Encourage students to read through the case study and discuss the issues with family members and/or friends. Answers to the Homework Questions can serve as the entry ticket to participate in the class discussion the following day.

Glossary

Bioinformatics: Bioinformatics is the application of computer science and information technology to biology and medicine. Bioinformatics makes it possible to analyze large and complex biological data and can be used to search biological databases, compare sequences, and draw molecular structures. Bioinformatic techniques are used to design and carry out the computer-based portion of genetic tests.

Database: A collection of related data that are stored, managed, and retrieved in an organized way.

Entrez: The search engine used to simultaneously search all of the databases at the NCBI. Items stored in different databases are cross-referenced and inter-linked, making it easier to find all the database records that are related to any one subject.

National Center for Biotechnology Information (NCBI): Part of the National Library of Medicine at the National Institutes of Health (NIH), the NCBI is a collection of biological information in over 30 cross-referenced databases related to genetics and molecular biology.

Nucleotide database: One of the databases at the NCBI which contains nucleotide (DNA and RNA) sequences.

Reference sequence: A sequence that has been chosen for the purpose of comparison. At the National Center for Biotechnology Information (NCBI), reference sequences are chosen because they are of high quality and are thought to accurately represent the sequence from the original organism. In genetic testing, a reference sequence is a known and well-studied DNA or protein sequence that does not contain any mutations or changes that are associated with disease. These sequences are used for comparison with patient sequences during genetic testing.

Credit

Tegarden, Deborah. Personal Interview. 6 November 2009.

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/

List of **All Resources** or Tools Available at the NCBI

List of the **Most Popular** Resources at the NCBI

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.
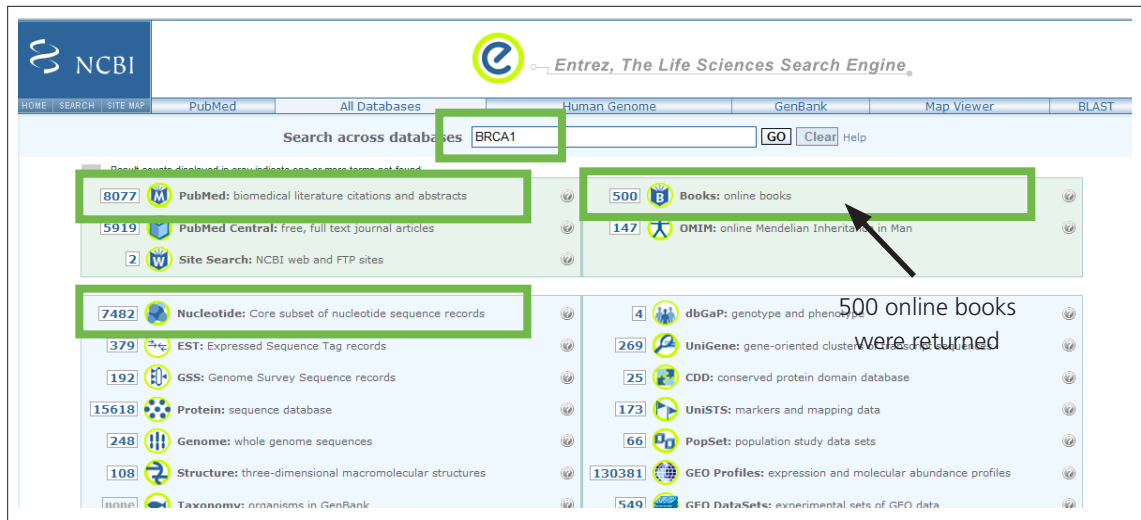
Figure 2: Search for "BRCA1." Credit: NCBI.

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.

   a. Why are we searching for *BRCA1*?

   b. The **Nucleotide** database has DNA sequences that have been loaded onto the NCBI database. How many times is 'BRCA1' cited in the Nucleotide database?

   c. 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?

   d. 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.
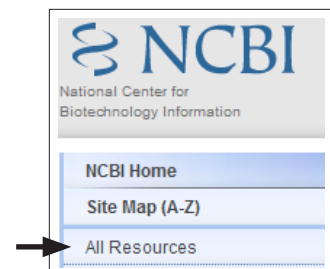

Figure 3: Click "All Resources" on the left side of the screen. Credit: NCBI.

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?

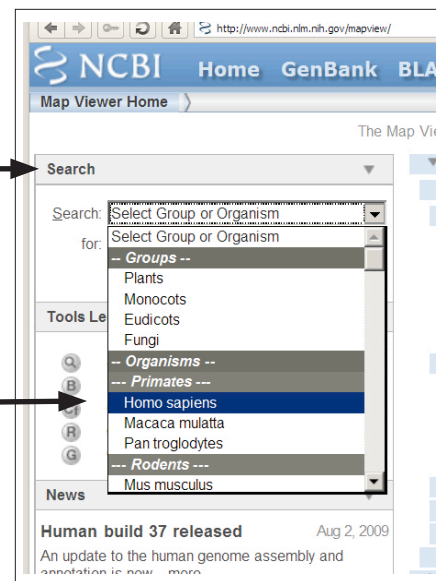Search pull-down menu →

Select *Homo sapiens* →

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

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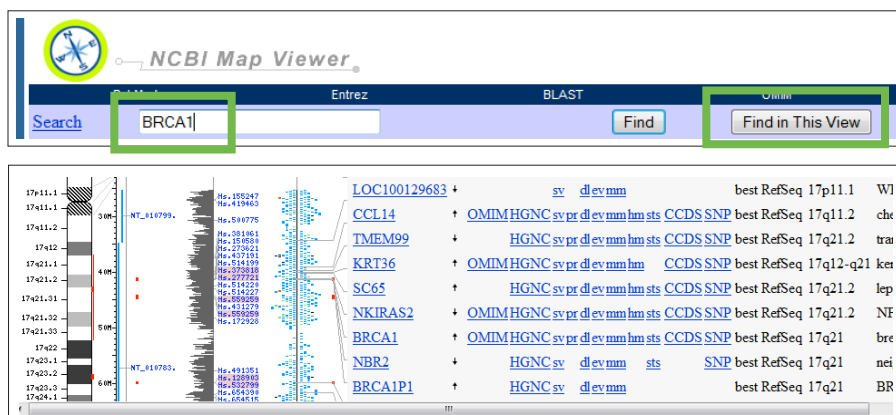
Figure 5: Find the location of the *BRCA1* gene by using the search function. Credit: NCBI.

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."

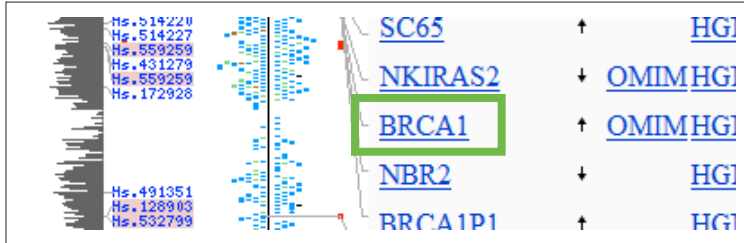List three of the functions that the BRCA1 protein performs.

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

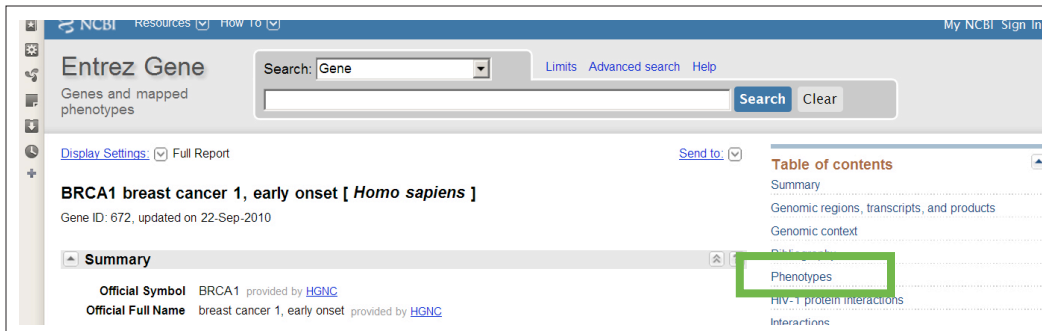*Figure 7:* Scroll down to find the "Gene Ontology" section. Credit: NCBI.

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*.

    a. Based on what you've learned in class, what is a **phenotype**?

    b. 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).
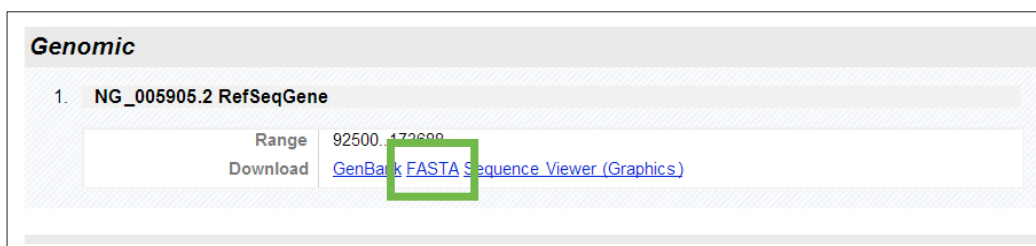
Figure 9: Click "FASTA" to obtain the FASTA sequence for *BRCA1*. Credit: NCBI.

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.
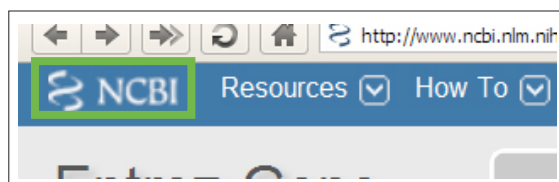

Figure 11: Return to the NCBI homepage by clicking on the logo. Credit: NCBI.

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.
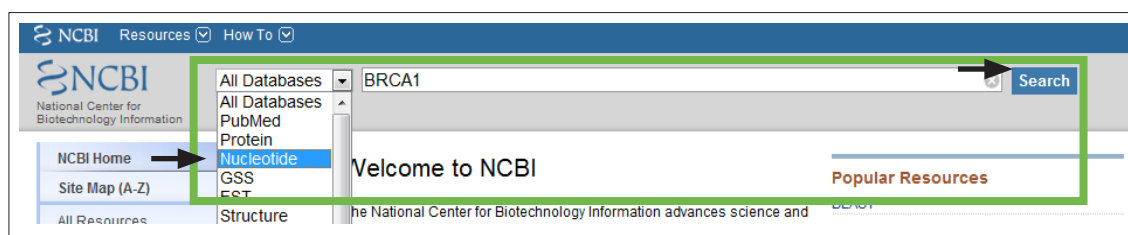

Figure 12: Type "BRCA1" into the Search box and choose "Nucleotide" from the pull-down menu. Credit: NCBI.

59

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.

Name _____ Date _____ Period _____

2 Navigating the NCBI Worksheet

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

Instructions: Use Student Handout—*Navigating the NCBI Instructions* to complete this worksheet.

2a. 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?

2b. What is one interesting resource listed in the Popular Resources menu on the right? What do you think is its function or purpose?

4a. Why are we searching for *BRCA1*?

4b. The **Nucleotide** database has DNA sequences that have been loaded onto the NCBI database. How many times is 'BRCA1' cited in the **Nucleotide** database? _____

4c. 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? _____

4d. Compare the numbers you got for Questions B and C. Do these relative numbers surprise you? What does this tell you about the *BRCA1* gene? Explain.

9a. How many different types of chromosomes do you see? _____

9b. What does "MT" represent? [**Note:** you can click the "MT" link to find out.]

9c. 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. 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.
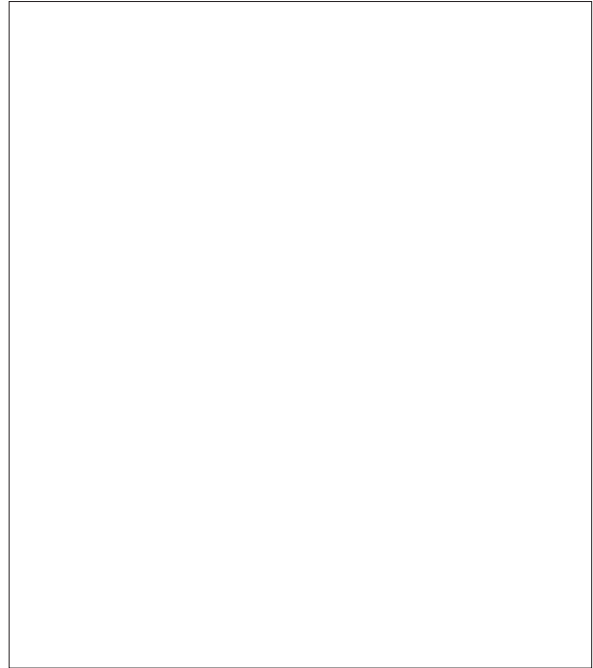
   1:

   2:

11. Draw a picture of Chromosome 17 in the box to the right and show the *approximate* location of *BRCA1* on this chromosome.

12. List three (3) functions that the protein produced by the *BRCA1* gene performs.

   1: Function 1:

   2: Function 2:

   3: Function 3:

13a. Based on what you've learned in class, what is a **phenotype**?

13b. What **phenotypes** are associated with mutations in the *BRCA1* gene? (You don't need to click the links.)

16a. What four letters make up this long sequence? _____

16b. Based on what you've learned in class, what do these letters represent?

19. List three organisms other than humans that have *BRCA1* genes.

   1: Common Name: _____

     Scientific name: _____

   2: Common Name: _____

     Scientific name: _____

   3: Common Name: _____

     Scientific name: _____

20. Does it surprise you that so many organisms share the *BRCA1* gene? Explain. (Hint: Look back at the functions of the BRCA1 protein (Question 12).)

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

   1:

   2:

   3:

2 Navigating the NCBI
Teacher Answer Key

[**Note:** Suggested point values are included after each question, and are intended to provide general guidelines for the weight each question could be given. Using these suggested point values, the total value for this worksheet is **30 points**.]

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

2a. 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?

> This will vary, as more resources are added to the NCBI, but students should list at least one Resource (such as "Nucleotide" and one function, such as "DNA sequences."
>
> (+1 for listing a resource and +1 for attempting to name its function/purpose.)

2b. What is one interesting resource listed in the Popular Resources menu on the right? What do you think is its function or purpose?

> This will vary, as more resources are added to the NCBI, but students should list at least one Resource, such as "Nucleotide," and one function, such as "DNA sequences."
>
> (+1 for listing a resource and +1 for attempting to name its function/purpose.)

4a. Why are we searching for *BRCA1*?

> To learn more about this gene and why mutations in this gene can lead to breast and ovarian cancer. (Students should draw on their experiences in *Lesson One*.)
>
> (+1 for referring to understanding how it causes breast and ovarian cancer.)

4b. The **Nucleotide** database has DNA sequences that have been loaded onto the NCBI database. How many times is 'BRCA1' cited in the **Nucleotide** database?

> New sequences are added every day. As of May 19, 2011, there were 7482
>
> (+0.5 for number.)

4c. 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?

> New articles are added every day. As of May 19, 2011, there were 8077
>
> (+0.5 for number.)

4d. Compare the numbers you got for Questions B and C. Do these relative numbers surprise you? What does this tell you about the *BRCA1* gene? Explain.

> Students should note that these numbers are close, but not all published studies will contain new *BRCA1* sequences, and some studies will contain more than one sequence.
>
> (+1 for response to relative numbers and +1 for explanation of response.)

9a. How many different types of chromosomes do you see? _____ (+0.5 for 25: 22 pairs, plus X, Y and MT.)

> 25

9b. What does "MT" represent?

> MT is the mitochondrial genome/chromosome (+1 for mitochondrial).

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

9c. 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? _____ (+0.5 for 23.)

> 23

10. 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.

> Answers will vary. Students will see many genes listed, with links to other NCBI resources.
>
> (+1 for each thing listed up to +2.)

11. Draw a picture of chromosome 17 in the box to the right and show the approximate location of *BRCA1* on this chromosome.
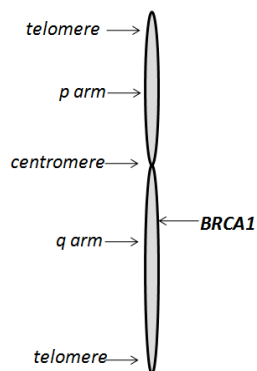
(+1 for drawing and +1 for BRCA1 label in approximate location.)

[**Note:** Labeling of the centromere, telomeres, p and q arms is optional.]

12. List three (3) functions that the protein produced by the *BRCA1* gene performs.

> 1. DNA binding
>
> 2. RNA binding
>
> 3. Androgen receptor binding
>
> 4. Protein binding
>
> 5. Tubulin binding
>
> 6. Transcription activator activity
>
> (+1 for each correct function listed, up to +3.)

13a. Based on what you've learned in class, what is a **phenotype**?

> A phenotype is an observable characteristic or trait (+1.)

13b. What **phenotypes** are associated with mutations in the *BRCA1* gene? (You don't need to click the links.)

> Breast cancer; ovarian cancer; breast-ovarian cancer; pancreatic cancer susceptibility; Papillary serous carcinoma of the peritoneum
>
> (+0.5 for each phenotype listed, up to +2.)

16a. Based on what you've learned in class, what four letters make up this long sequence? _____
   (+1 for including all 4 letters.)

> A, T, C, G

16b. What do these letters represent? A = adenine; T= thymine; G = guanine; C= cytosine

> These are the bases of the *BRCA1* gene. (+1 for either listing all 4 bases or for stating 'the bases.')

19. List three other organisms that have *BRCA1* genes. (+1 for each organism; – 0.5 for common name and 0.5 for scientific name; up to +3.)

> Answers will vary, as almost all animals have *BRCA1* genes, and new sequences are being added every day, Possible answers include:
> 1. Mus musculus (mouse)
> 2. Bos taurus (cow)
> 3. Macaca mulatta (rhesus macaque/monkey)
> 4. Sus scrofa (pig or boar)
> 5. Lagopus lagopus (willow ptarmigan, a type of bird)

20. Does it surprise you that so many organisms share the *BRCA1* gene? Explain. (Hint: Look back at the functions of the BRCA1 protein (Question #12).)

> All organisms need to bind DNA, RNA and other proteins.
>
> (+1 for statement of surprise and +1 for reference to functions of BRCA1 protein.)

Lesson 2 – Navigating the NCBI

67

21. Summarize what you have learned today by listing three types of information found at the NCBI. (+1 for each type listed, up to +3.)

1. Scientific publications
2. Scientific books
3. Gene/nucleotide sequences
4. Protein sequences
5. Protein structures
6. Information about gene functions