Name That Gene, Disease and Protein!

By Sharlene Denos (UIUC) & Kathryn Hafner (Danville High)

What You Will Be Doing

In this assignment you will be introduced to the field of bioinformatics. This is a field of biology that has arisen recently, as more and more organisms' entire genomes are being sequenced and stored into online databases. You will access one of these databases, known as GenBank, to search for the gene that corresponds to a short DNA sequence that we will give you. The genes, you will find, are all associated with genetic diseases, meaning that there is at least one type of mutation in the gene that can lead to a disease in humans.

Once you find the gene that your sequence came from, you will spend time learning more about it and the genetic disease that results from mutations in the gene. You will also find out what protein your gene codes for (what is its name) and where (at which amino acid position) the mutation occurs in this protein. There may be several possible mutations in this gene/protein (as well as other genes/proteins that give rise to this disease). You're required to describe at least one of the mutations in your gene for your report.

Finally, you will search for the protein your gene codes for in the Protein Data Bank (PDB). There, you will view the structure of your protein and view the site of the mutation. Can you guess why a mutation at that site would disrupt the protein's function?

Your Report Must Contain The Following:

1. The name of your gene and its location in the human genome (i.e. give the chromosome number and position, such as "very top of the short arm on chromosome 6" or "close to the centromere in the short arm of the x chromosome").

2. The name and description of the disease associated with mutations in your gene.

3. The name & normal function of the protein your gene codes for and a description of where the mutation occurs (may be more than one place).

4. Describe how the mutation changes your protein's structure and/or function (i.e. "the mutation is in the active site and replaces a key amino acid involved on the catalysis with one that cannot perform the catalysis").

5. A VMD image which illustrates number 4.

PROTOCOL

1. Connect to the Internet.

2. Find the home page for NCBI (National Center for Biological Information) at www.ncbi.nlm.nih.gov.

3). Click on the word "**BLAST**," located on the blue bar at the top of the page.

4). Under the heading,"**Nucleotide**", click on the link,"**Nucleotide-Nucleotide Blast (blastn)**".

5). On the next screen, type (using lower case) the **exact**
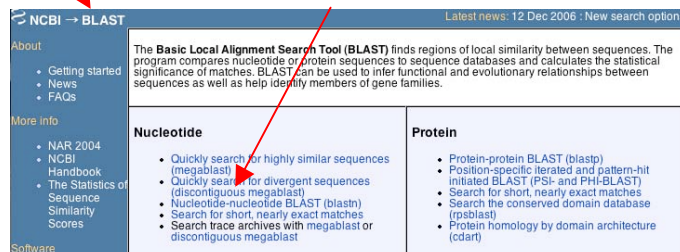
nucleotide sequence you were given in the large empty box labeled "search". Double

check your sequence to make sure it is correct, then click "BLAST!"

6). On the next screen, you should receive the message, "Your request has been

successfully submitted and put into the Blast Queue". Wait here a few seconds and then

click "**Format!**" bar. If your results are still not ready, you will see a screen asking you

to wait for the search to complete. Be patient while formatting takes place.

7). After the search has ended, scroll down to where it says "Distribution of BLAST Hits

on the Query Sequence".

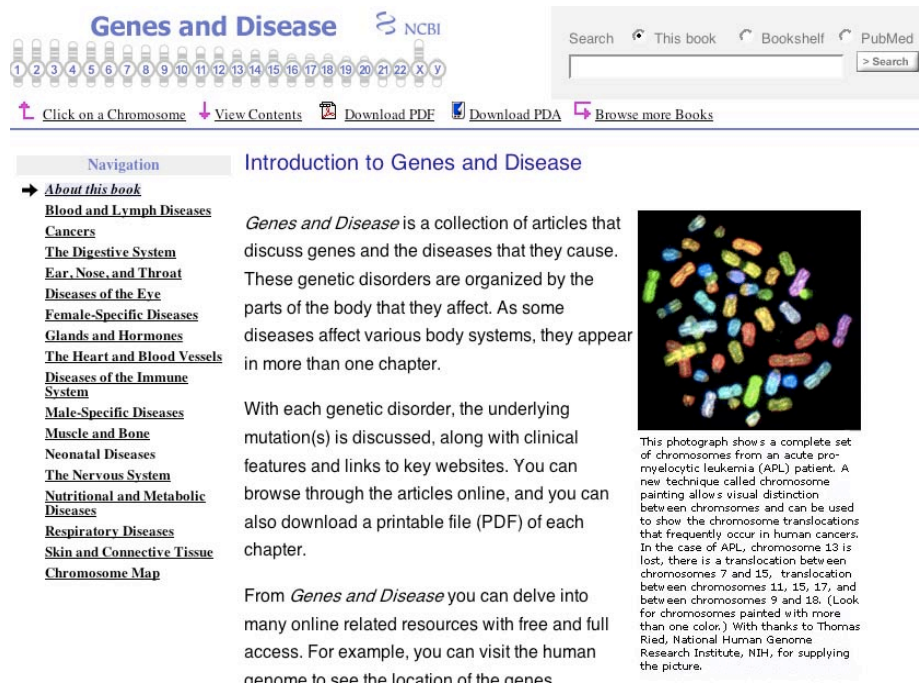Distribution of 19 Blast Hits on the Query Sequence
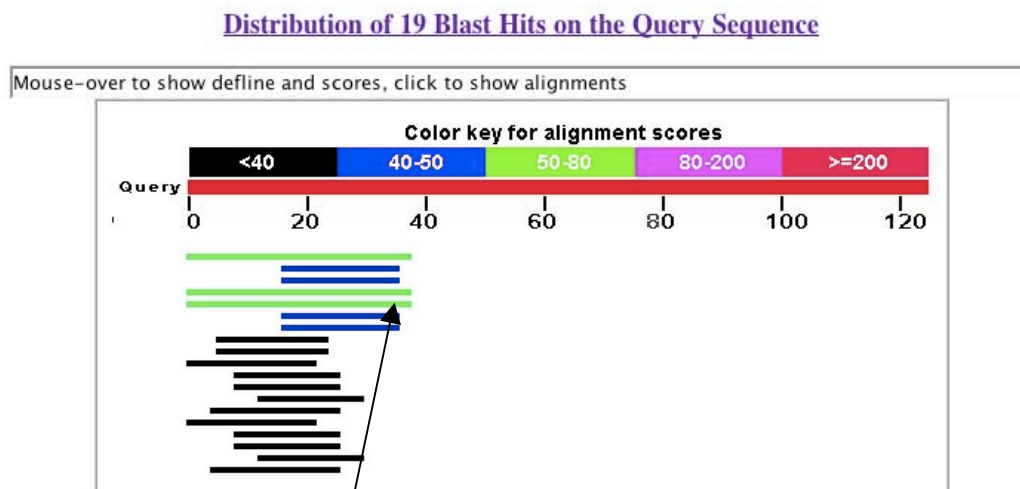
The color key for alignment scores lets you know how well the returned sequences

aligned to the one you entered. Your query sequence is listed just beneath this key. It is

in red since your query must align perfectly with itself! The other matches will differ

from pink (the best) to black (the worst). If you scroll your mouse over these colored

lines, you will see the sequence name, the score ("S"), and the expectation value ("E")

appear in the box above the color key. Click on the line with the highest score and the

lowest expectation value. This is the sequence that gives the best alignment to your

"query". From what organism is this sequence? You should be able to determine what

genetic disease your sequence corresponds to from the name alone. If not, you will need

to do some more digging to find the genetic disease. The alignment of this sequence with

your query is shown at the bottom. How many bases are shown in this alignment? How

many bases were in the original query (the sequence you originally entered)? Is there a

100% match for the bases which are shown? What is the Score for this alignment (in

bits) and what is the Expectation or E-value?

8). Now you will go to the Genes and Diseases part of the NCBI website at

http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=gnd

Read this main page:

9). This site allows you to search genetic diseases by chromosome or by the type

of disease. If you know which chromosome(s) contain mutations associated with your

disease, you can click directly on that chromosome (all 23 chromosomes are located at

the top of your screen). You can also just click through the various chromosomes until

you find your genetic disease. Also, you can easily find your genetic disease if you know

what category it falls under. Is it a type of cancer? Is it a disease of the blood, the eye,

the muscle or bone? If you know this information, you can easily click on the

corresponding category to get to the page describing your genetic disease.

10). Once you have found the page that describes your disease, you will need to read

about it and take notes. What chromosome is the sequence from and what is the name of

the gene this sequence is from. What protein does it code for? Find at least one mutation

associated with your disease that is not present in the gene you were assigned. To find

the additional mutations, select "genome view" in the right panel of this page. You

should arrive at a page that shows all 22 autosomes, both x and y sex chromosomes, and

the mitochondrial DNA. Red lines on the chromosomes indicate regions of DNA where

mutations are associated with your disorder.

You only want the genes relevant to your disease, so find the "Quick Filter" box on the right of your screen. Click "Gene" and then hit

"filter". You should now see very few (or one) mutation sites in only a few

chromosomes. Note which chromosomes (& the region of these chromosomes) are

associated with this disease.

11). Now go back to the main description page for your genetic disease. Click the "Entrez Gene" link on the right panel of the page. This will take you to a page containing a list of all genes associated with your disease and their corresponding symbols. Click on these symbols to go to a page which will provide a summary of the function of each gene and the problems associated with mutations. You will need this information for your reports. You will need to find and write down the exact location of your mutation. You may see something such as A83 or Ala83 or Alanine 83. That means the 83rd amino acid in your protein is mutated from alanine to something else.

12). Now go to the Protein Data Bank website, www.rcsb.org. Type in your protein's name in the search bar at the top of the page and press "site search". If you have any difficulty, try typing just one word (i.e. instead of beta hexosaminidase A, just type hexosaminidase). You will get many "hits". Click on the one that best matches your protein's name. Write down the PDB code for this protein, you will need it to load your molecule into VMD.

13). Load VMD from the desktop icon and select "New Molecule" from the File menu in the VMD Main window. Now type in your protein's PDB code and press enter. The "File Type" should automatically change to "Web PDB Download". Click "Load" and see that your protein is loaded in the OpenGL window. Spend a few moments refreshing your memory about how to manipulate molecules in VMD. Use the table below as a reference.

Key pressed	Action Performed
t, T	Enter translate mode for moving entire molecule
r, R	Enter rotation mode, stop rotation of selected molecule
S, S	Enter scale mode for magnifying or shrinking the selected molecule
c	Choose center about which rotation will take place
x, X	Begin rotation about the x axis, rock back and forth about x axis
y, Y	Begin rotation about the y axis, rock back and forth about y axis
z, Z	Begin rotation about the z axis, rock back and forth about z axis
1	Label atoms selected with mouse left click
2	Enters "bond label" mode, which gives the distance between two atoms selected by successive mouse left clicks

14). Select "Representations" under the Graphics menu in VMD Main. You will use the selected atoms window to locate your mutations. Select "Create Rep" in the Representations tab and type resid #, where # stands for the amino acid number of your mutation (from step 11). To make it easier to see, change the Drawing Method to "VDW" and the Coloring Method to "ColorID". If there is more than one chain in your protein, you may want to type "resname NAME and resid #" under selected atoms, where NAME is the 3-letter abbreviation of your amino acid in all caps and # is the resid number. If you still have more than one amino acid highlighted, that's okay, it means the chains are identical and the mutation occurs in more than one place.

15). Mutations that cause disease occur in parts of the protein structure that are especially sensitive to change, such as active sites and binding sites. Is your mutation in an active site or binding site? To help determine this, first select the first representation and type in "protein" and change the Drawing Method to "New Cartoon". You can click "Create Rep" again and type in "not protein and not water" to see if your protein has any

co-factors. This is an easy way to locate the active site. Alternatively, if you took good notes in step 11, you may already know how this mutation disrupts your protein and the amino acid numbers of the active site.

16). Once you have an image in VMD that highlights where your mutations occur and why they interfere with its function, you will need to save. From the file menu select "render" and choose render using "snapshot". Browse to the folder where you want to save and type in "YourMoleculeName".tga as the filename and click "Start Rendering".