Genomics and Bioinformatics from Biomedicine and Biodiversity

WSU STEM DAY 2018

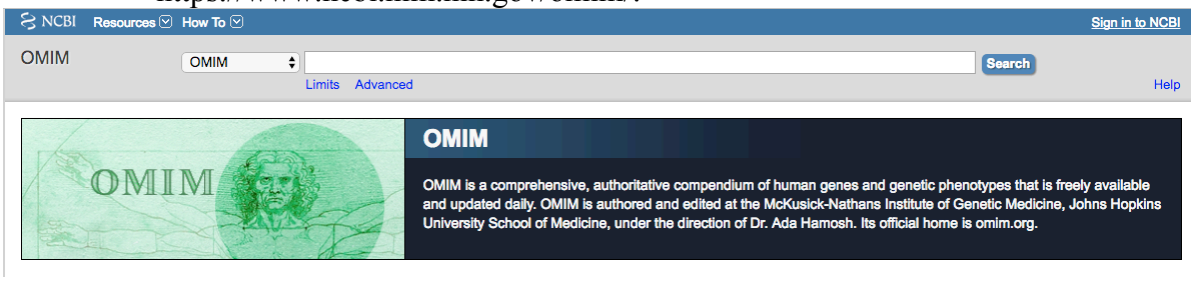
Using Public Genomic Data Bases for Human Genetics and Genomics

Dr. Chuanzhu Fan, Dr. Weilong Hao, Dr. Jared Schrader

Exercise

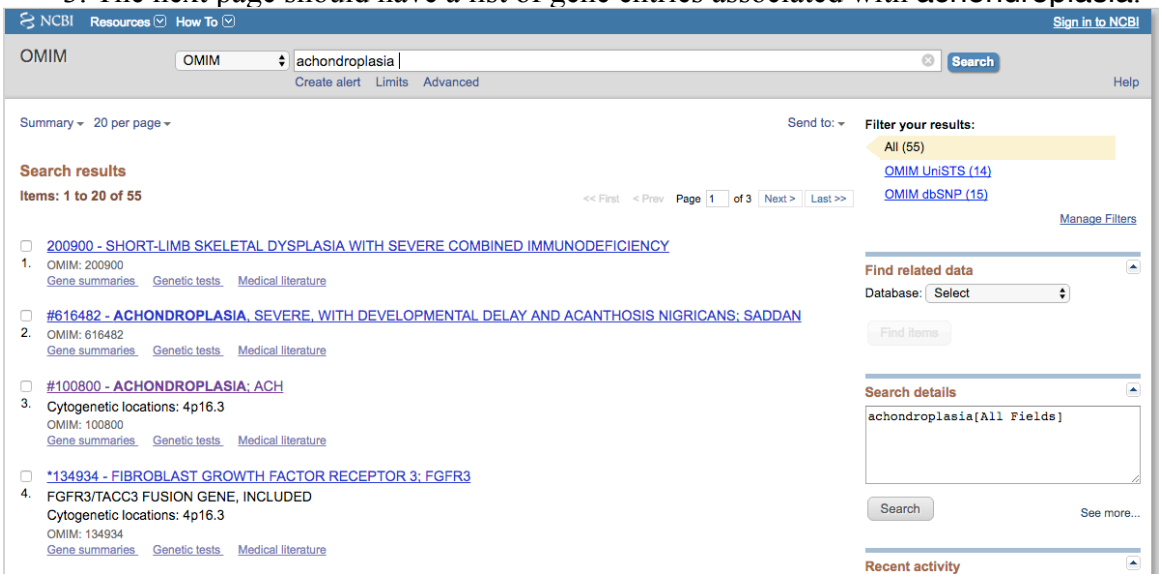
There are quite a number of data bases, maintained through public funding, that are excellent sources for information and for data mining. This exercise will introduce you to these sites and indicate some of the information that you can easily glean. In this exercise, we will go to a site that organizes data on human genes and congenital disorders. The site is called *Online Mendelian Inheritance in Man* (*OMIM*) from NCBI (National Center for Biotechnology Information). We will start out by looking at one human genetic disorder and garner the information listed. Then you will be given a different disorder and you will retrieve the information given on it.

1. Connect to the *OMIM* site via the internet by going to https://www.ncbi.nlm.nih.gov/omim/.

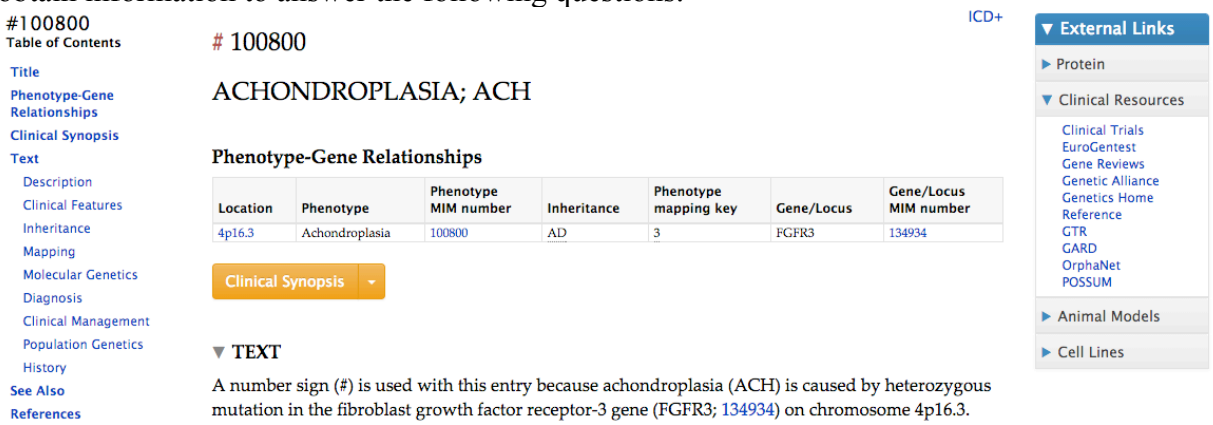


2. Near the top of the page, type in achondroplasia in the **Search** box. Click on **Search** at the right of this box.

3. The next page should have a list of gene entries associated with achondroplasia.



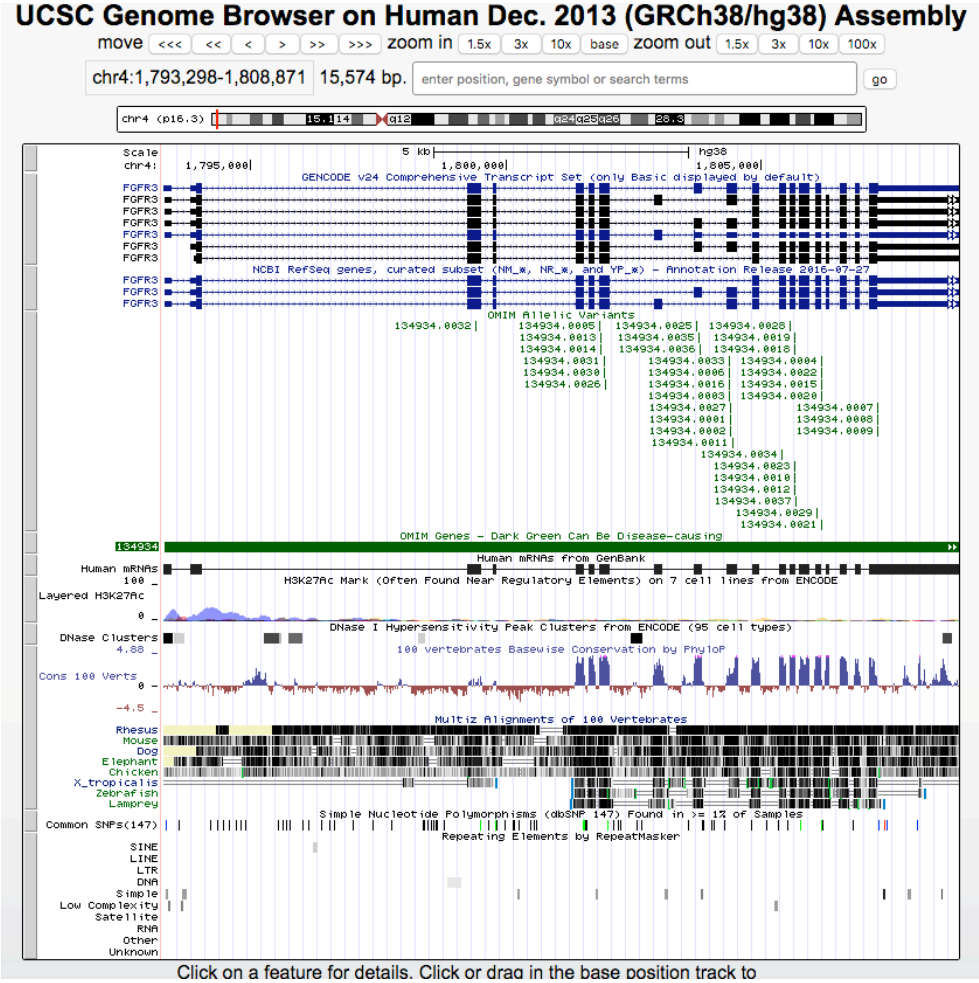
4. On the web page, click on the number in the first line of the first gene entry. You should see a new page entitled "ACHONDROPLASIA; ACH". As you read down the page, you will obtain information to answer the following questions.



a. What gene is involved? (The name of the gene will be listed both as an abbreviation with letters and numbers and as a number.)
b. What is the phenotype of affected individuals?
c. Is this a recessive or dominant condition?
d. What is known about homozygous individuals?
e. Are there any cognitive developmental effects?

5. After you have answered the questions above, go back to the top of the page and click on the chromosomal location (i.e., 4p16.3, under the word "location"). You will reach a page that is tabular in form. Click again on the chromosomal location (i.e., "Location (genomic start, cyto location)") and you will reach a page with very detailed genetic maps.

a. In the 5th row in this table is the information on "**FGFR3, ACH**". FGFR3 is the gene that when mutated causes Achondroplasia. Click on the number in the far left column. A highly detailed set of information will appear. The topmost map is a drawing of chromosome 4; a small vertical red bar on the left side of this chromosome shows where the gene for Achondroplasia resides. Underneath this image is a series of very detailed maps drawn as a series of horizontal lines. Clicking on one of these lines will lead you to another window with additional highly detailed information. Here the map is first generated by complete genome sequencing data, and genes are annotated on this map.
b. Underneath these maps is information on the same, or very similar, genes found in other species, e.g. the Rhesus monkey, mouse, dog, and elephant. Again, clicking on any of these lines will open another window with even more detailed information.
c. The next set of maps on this same page appear as lines in various colors. Click on one of these lines to open yet another browser window. Here you will be provided information on genetic variants of these FGFR3 DNA sequences. What do the different colors of lines indicate? How might you use this information if you had a patient with Achondroplasia? Might this information be useful in knowing how your patient's genetic condition might be treated compared to other people who have the same altered DNA sequence?



6. Now go back to the page with the chromosome drawing at the top. Click on the 10X "zoom in" button at the top, middle, of this page. Look carefully at the maps as you do this. What happens? And what happens if you keep clicking on the 10X button? How detailed is the genetic information that you can obtain? What is the smallest unit of genetic information that can be identified here?
7. Now click several times on the 10X zoom out button. What happens to the vertical red line on the left side of the chromosome? What happens to the map information beneath the chromosome? How big can the "big picture" get? Can you see the whole chromosome?
8. Now click on one of the buttons to the right of word "move" that is found to the left of "zoom in". What happens to the red box? How detailed is the information you can obtain on this chromosome? Can you see the level of detail that is available on the FGFR3 gene? And can you see the detail on genes that are found to the left and to the right of the FGFR3 gene? Finally, do you have a feel for the depth and breadth of information that is readily available on the human genome? Remember that there about 30,000 human genes, many of which are associated with one (or even more!) diseases. You should be able to see how a website such as OMIM is essential for storing, accessing, and tracking the information on human genetic diseases.

For this exercise, pick up any one human congenital disease or condition from the list below. Do steps 1 through 8 with conditions. Answer the same questions as you answered as above, but this time it is for diseases you choose. Some of the questions might not be applicable for the diseases you choose, hence only answer the questions that apply to the chosen diseases.

1. Red-Green Color blindness
2. Hemophilia A
3. Cystic fibrosis
4. Albinism
5. Tay Sach's disease
6. Sickle cell anemia / Hemoglobin C disease
7. Phenylketonuria (PKU)
8. Alkaptonuria
9. Huntington disease
10. Alpha thalassemia
11. Galactosemia
12. Beta thalassemia
13. Trichothiodystrophy / Cockayne Syndrome / Xeroderma Pigmentosum complementation group D
14. Duchenne / Becker muscular dystrophy
15. Hunter's syndrome
16. Marfan Syndrome
17. Hypolactasia
18. Azoospermia
19. Retinoblastoma
20. Hutchinson-Gilford progeria syndrome (HGPS)
21. Myotonic dystrophy
22. Breast cancer
23. Werner syndrome
24. Familial Hyperchrolesterolemia
25. Amyloidosis