Chromosome Maps

Assignment Goal: To use the Internet-based Genes and Disease site (NCBI) to view the assignment of genes
to chromosomes.
Assignment: Access the Genes and Disease site at
http://www.ncbi.nlm.nih.gov/books/NBK22183/

Under "Contents", select "Chromosome Map" (at the very bottom).
A karyotype will appear.
Click on a chromosome.

1. WHAT CHROMOSOME DID YOU CHOOSE?
Above the chromosome image you will see the number of genes and base pairs on that particular chromosome. **Chromosome 10**

2 & 3. STATE THE NUMBER OF GENES AND BASE PAIRS ON THE CHROMOSOME YOU CHOSE.
Scan the chromosome map. **Genes: over 1400, Base Pairs: Over 130 million**

4. LIST ONE GENE THAT IS LOCATED ON THIS CHROMOSOME. **Ornithine ketoacid aminotransferase (OAT)**

5. STATE THE FUNCTION OF THE GENE YOU LISTED IN #4. **The 10th chromosome is responsible for providing instructions to make proteins**
This is possible by clicking on the gene you stated in #4. It is important that you state the physiological function of the gene product you select, not the disease listed for the gene.

Introduction to Nucleotide BLAST

Assignment Goal: To use the Internet-based site BLAST, Basic Local Alignment Search Tool (NCBI), to search for similarities between nucleotide sequences.
Assignment: Access the BLAST site at http://blast.ncbi.nlm.nih.gov/Blast.cgi

Click on "Nucleotide Blast"
Assume that you found this nucleotide sequence when you cloned a piece of gene in the laboratory in which you work: aattggaagc aaatgacatc acagcaggtc agagaaaaag ggttgagcgg caggcaccca gagtagtagg tctttggcat taggagcttg agcccagacg gccctagcag ggaccccagc

Enter the above sequence (you may copy and paste) into the "Enter Query Sequence" box at the top of the page.
Under "Program Selection" near the bottom of the page, choose "somewhat similar sequence (blastn)"
Click the "BLAST" button at the bottom of the page to run the search.
Give some time for the results of your search to show up.
You will be given significant matches for the sequence that you entered.

6. WHAT IS THE SECOND SEQUENCE DESCRIPTION MATCH FOR YOUR QUERY SEQUENCE? For this answer, you should give the description listed. Do not give the actual sequence. **Homo sapiens CFTR (CFTR Gene), partial cds**

7. WHAT DOES THE ENCODED PROTEIN DO IN THE BODY? To find this information, click on the sequence description and you will get a nucleotide match. To the right of that nucleotide match you will find "Related Information" and under that you will find "Gene". Click on "Gene" to find out more information about this protein. **This gene encodes a member of the ATP-binding cassette (ABC) transporter superfamily. The encoded protein functions as a chloride channel, making it unique among members of this protein family, and controls ion and water secretion and absorption in epithelial tissues. Channel activation is mediated by cycles of regulatory domain phosphorylation, ATP-binding by the nucleotide-binding domains, and ATP hydrolysis.**

8. FOR WHAT DISEASE IS A MUTATED FORM OF THIS GENE RESPONSIBLE?
You should be able to get this information from the description of the gene.
Mutations in this gene cause cystic fibrosis, the most common lethal genetic disorder in populations of Northern European descent..The most frequently occurring mutation in cystic fibrosis, DeltaF508, results in impaired folding and trafficking of the encoded protein.

9. ON WHAT CHROMOSOME IS THE GENE LOCATED? You should be able to get this information by looking at the "Genomic context" which is just below the description of the gene. **Chromosome 7**

10. Return to the original nucleotide sequence alignment descriptions. CHOOSE A SPECIES (STATE THE SCIENTIFIC NAME) OTHER THAN HOMO SAPIENS THAT

ALSO HAS A 100% IDENTITY (Per. Ident) FOR THIS SEQUENCE? There will be about 3 species that you could name, so choose one of them. **Pan troglodytes**

11. WHAT IS THE COMMON NAME FOR THIS SPECIES? **Eastern Chimpanzee**

12. DOES IT SURPRISE YOU THAT THIS SPECIES ALSO HAS A 100% SIMILARITY IN IDENTITY? WHY OR WHY NOT? **No, because chimpanzees are very closely related to humans**

13. Return to the original nucleotide sequence alignment DESCRIPTION. Find the first match that has less than 100% similarity identity. Click on the description to answer this question.

a. WHAT IS THE GENUS AND SPECIES WITH THIS NUCLEOTIDE SEQUENCE?
Sapajus apella

b. WHAT IS THE COMMON NAME?
Tufted capuchin

c. HOW MANY GAPS OCCUR BETWEEN THE TWO SEQUENCES (THE ONE YOU ORIGINALLY SUBMITTED AND ONE THAT HAS LESS THAN 100% QUERY COVER)? The gap information is just above the sequence alignment. **1/119 (0%)**

14. WHAT IS A GAP IN SEQUENCE ALIGNMENTS? (This is something you'll have to search for on the Internet.) **The absence of a region, which is present in another sequence.**

You can also do BLAST searches using an accession number that has been assigned to a particular sequence when entered into the database.
Go back to the Blast home page (www.ncbi.nlm.nih.gov/BLAST.cgi) and again choose "Nucleotide Blast".
Look up the following sequences using the given accession numbers.
(Under "Program Selection" near the bottom of the page, choose "somewhat similar sequence (blastn)".
(Again, click on the "BLAST" button at the bottom of the page after you have entered the accession number.)

FOR EACH, STATE WHAT THE GENE IS (#15-18). (Again, give the description of the gene or gene product, not the nucleotide sequence.)

15. NM_145556. **TAR DNA binding protein. Enables RNA polymerase II cis-regulatory region sequence-specific DNA binding activity and pre-mRNA intronic binding activity. Involved in positive regulation of protein import into nucleus; regulation of circadian rhythm; and regulation of protein stability. Acts upstream of or within RNA splicing. Located in the nucleus. Is expressed in several structures, including branchial arch; central nervous system; early conceptus; genitourinary system; and heart. Used to study Grn-related frontotemporal lobar degeneration with Tdp43 inclusions; amyotrophic lateral sclerosis type 10; and frontotemporal dementia. Human ortholog(s) of this gene implicated in Parkinson's disease; amyotrophic lateral sclerosis; amyotrophic lateral sclerosis type 10; and motor neuron disease. Orthologous to human TARDBP .**

16. NM_013444. **UBQLN2. This gene encodes an ubiquitin-like protein (ubiquilin) that shares a high degree of similarity with related products in yeast, rat and frog. Ubiquilins contain a N-terminal ubiquitin-like domain and a C-terminal ubiquitin-associated domain. They physically associate with both proteasomes and ubiquitin ligases; and thus, are thought to functionally link the ubiquitination machinery to the proteasome to affect in vivo protein degradation. This ubiquilin has also been shown to bind the ATPase domain of the Hsp70-like Stch protein.**

17. NM_001010850. **FUS. This gene encodes a multifunctional protein component of the heterogeneous nuclear ribonucleoprotein (hnRNP) complex. The hnRNP complex is involved in pre-mRNA splicing and the export of fully processed mRNA to the cytoplasm. This protein belongs to the FET family of RNA-binding proteins which have been implicated in cellular processes that include regulation of gene expression, maintenance of genomic integrity and mRNA/microRNA processing. Alternative splicing results in multiple transcript variants. Defects in this gene result in amyotrophic lateral sclerosis type 6.**

18. KJ174530. **SOD1. The protein encoded by this gene binds copper and zinc ions and is one of two isozymes responsible for destroying free superoxide radicals in the body. The encoded isozyme is a soluble cytoplasmic protein, acting as a homodimer to convert naturally-occurring but harmful superoxide radicals to molecular oxygen**

and hydrogen peroxide. The other isozyme is a mitochondrial protein. In addition, this protein contains an antimicrobial peptide that displays antibacterial, antifungal, and anti-MRSA activity against E. coli, E. faecalis, S. aureus, S. aureus MRSA LPV+, S. agalactiae, and yeast C. krusei. Mutations in this gene have been implicated as causes of familial amyotrophic lateral sclerosis. Rare transcript variants have been reported for this gene.

19. Search Google to answer the following: WHAT DISEASE IS ASSOCIATED WITH MUTATIONS OF THE GENES REFERENCED IN #15-#18? **ALS (Amyotrophic lateral sclerosis)**
WHAT IS A "COMMON NAME" OF THE DISEASE?
Lou Gehrig's disease (The name of a person; Hint, hint...Baseball season just finished...)

20. BLAST is possible because of the submission of DNA sequences to GenBank.
WHAT IS GENBANK? (You can do an Internet search to find this information.)
GenBank sequence database is an open access, annotated collection of all publicly available nucleotide sequences and their protein translations.

Introduction to Protein BLAST

Assignment Goal: To use the Internet-based site BLAST, Basic Local Alignment Search Tool (NCBI), to 1) translate cDNA and 2) to search for similarities between amino acid sequences. Assignment: Access the BLAST site at http://blast.ncbi.nlm.nih.gov/Blast.cgi

Click on "tblastx". (This is a smaller button next to "Nucleotide BLAST".
21. First, answer this question: WHAT IS cDNA? (Again, you can do an Internet search to find this information.) **cDNA (short for copy DNA; also called complementary DNA) is synthetic DNA that has been transcribed from a specific mRNA through a reaction using the enzyme reverse transcriptase**

Enter the following cDNA sequence on "blastx":
ACATTTGCTTCTGACACAATTGTGTTCACTAGCAACCTCAAACAGACACCATG
GTGCATCTGACTC
CTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGA
AGTTGGTGGTGAGGCCCTGGGCAG

22. WHAT IS THE SEQUENCE MATCH? **Hemoglobin**

Assignment: While still on BLAST, choose “Protein BLAST”. Check the box which says “Align two or more Sequences”. Copy and paste the following sequence into the “Enter Query Sequence” box: Person 1/Sequence 1:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGD
AVELSCPP
PGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRV
LCHFSVRV
TDAPSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPA
AGNPTPSIS
WLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTY
TLDVLERSP
HRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTP
YVTVLKTA
GANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEAD
EAGSVYAG
ILSYGVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMS
SNTPLVRIA
RLSSGEGPTLANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDK
DRAAKPVT
VAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLLGACTQGGPLYVLVEY
AAKGNLRE
FLRARRPPGLDYSFDTCKPPEEQLTFKDLVSCAYQVARGMEYLASQKCIHRDLAA
RNVLVTED
NVMKIADFGLARDVHNLDYYKKTTNGRLPVKWMAPEALFDRVYTHQSDVWSF
GVLLWEIFT
LGGSPYPGIPVEELFKLLKEGHRMDKPANCTHDLYMIMRECWHAAPSQRPTFKQ
LVEDLDRVL
TVTSTDEYLDLSAPFEQYSPGGQDTPSSSSGDDSVFAHDLLPPAPPSSGGSRT

Copy and paste the following sequence into the “Enter Subject Sequence” box:
Person 2/Sequence 2:

MGAPACALALCVAVAIVAGASSESLGTEQRVVGRAAEVPGPEPGQQEQLVFGSGD
AVELSCPP

PGGGPMGPTVWVKDGTGLVPSERVLVGPQRLQVLNASHEDSGAYSCRQRLTQRV
LCHFSVRV
TDAPSSGDDEDGEDEAEDTGVDTGAPYWTRPERMDKKLLAVPAANTVRFRCPA
AGNPTPSIS
WLKNGREFRGEHRIGGIKLRHQQWSLVMESVVPSDRGNYTCVVENKFGSIRQTY
TLDVLERSP
HRPILQAGLPANQTAVLGSDVEFHCKVYSDAQPHIQWLKHVEVNGSKVGPDGTP
YVTVLKTA
GANTTDKELEVLSLHNVTFEDAGEYTCLAGNSIGFSHHSAWLVVLPAEEELVEAD
EAGSVYAG
ILSYRVGFFLFILVVAAVTLCRLRSPPKKGLGSPTVHKISRFPLKRQVSLESNASMS
SNTPLVRIA
RLSSGEGPTLANVSELELPADPKWELSRARLTLGKPLGEGCFGQVVMAEAIGIDK
DRAAKPVT
VAVKMLKDDATDKDLSDLVSEMEMMKMIGKHKNIINLL
GACTQGGPLYVLVEYAAKGNLREFLRARRPPGLDYSFDTCKPPEEQLTFKDLVSC
AYQVARG
MEYLASQKCIHRDLAARNVLVTEDNVMKIADFGLARDVHNLDYYKKTTNGRLP
VKWMAPEA
LFDRVYTHQSDVWSFGVLLWEIFTLGGSPYPGIPVEELFKLLKEGHRMDKPANCT
HDLYMIMR
ECWHAAPSQRPTFKQLVEDLDRVLTVTSTDEYLDLSAPFEQYSPGGQDTPSSSSS
GDDSVFAHD
LLPPAPPSSGGSRT

Click on the “BLAST” box.

After you receive results, choose the “Description” box for more information about the alignment.

23. DO YOU SEE ANY DIFFERENCES BETWEEN THE TWO AMINO ACID SEQUENCES? (Look for a space between the same amino acid comparison for both sequences.) **YES**

24. IF YOU SAW DIFFERENCES, WHAT WERE THEY? **There is a dash/missing letter**

25. ARE THERE ANY GAPS IN THE SEQUENCE ALIGNMENT? **YES**

Return to the BLAST home page (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Return to the Protein BLAST search to identify the polypeptide which you have been analyzing. (You may use either sequence.). Copy and paste the sequence into the "Enter Query Sequence box. Be sure that the box "Align two or more sequences" is NOT checked. Choose "Quick BLASTP" under the Program Selection. Click the "Blast" button at the bottom.

26. WHAT GENE ENCODES FOR THE POLYPEPTIDE YOU WERE ANALYZING?
Click on "Description", then choose "Gene" under "Related Information" to answer the following two questions. **This gene encodes a member of the fibroblast growth factor receptor (FGFR) family, with its amino acid sequence being highly conserved between members and among divergent species. FGFR family members differ from one another in their ligand affinities and tissue distribution. A full-length representative protein would consist of an extracellular region, composed of three immunoglobulin-like domains, a single hydrophobic membrane-spanning segment and a cytoplasmic tyrosine kinase domain. The extracellular portion of the protein interacts with fibroblast growth factors, setting in motion a cascade of downstream signals, ultimately influencing mitogenesis and differentiation.**

27. WHAT IS THE FUNCTION OF THIS PROTEIN? **This particular family member binds acidic and basic fibroblast growth hormones and plays a role in bone development and maintenance.**

28. WHAT HUMAN DISEASE IS CAUSED BY A MUTATION IN THIS GENE?
Again, this information can be gained by clicking on "Description" and then "Gene" under "Related Information" to the right to get this information. **Mutations in this gene lead to craniosynostosis and multiple types of skeletal dysplasia.**

Finally, but not required, if you click on the "AlphaFold Structure" under "Related Information" you will be given a structural diagram of this protein.

29. WHAT IS THE CONNECTION AMONG THE FOLLOWING: NIH, NLM, NCBI, and HHS?
All medical resource sites that have an abundance of information.

30. REFLECT ON ONE THING THAT YOU LEARNED FROM DOING THIS ASSIGNMENT. **I learned that for the disease ALS, so many genes play a role.**