

Biol 322 Fall 2012 Genes We Share

This assignment is worth 15 pts, should be submitted via email and is due on _____

What is HNPCC-MLH1 all about? How could the study of a gene in *E. coli* help us to understand cancer in humans?

F	W	Y	Cancer
+			ABL1
+			Acute Myeloid Leukemia-DEK
+			Adenomat. Polyposis Coli-APC
+			AKT2
+			Ataxia Telangiectasia-ATM
-			BRCA1
-			BRCA2
+			Basal Cell Nevus-PTC
+			B-Cell Lymphoma 2-BCL2
-			B-Cell Lymphoma 3-BCL3
+			Bloom-BLM
+			Burkitt's Lymphoma-MYC
-			CDKN2C
-			CSF1R/C-Fms
+			Chk2 Protein Kinase
-			PDGFB
+			CML-BCR
+			Cyclin D1-CCND1
+			Cyclin Dep. Kinase 4-CDK4
+			EGFR
+			ERBB2
-			ETS
+			E-Cadherin-CDH1
+			Ewing Sarcoma-FLI-1
-			FGF3
-			Fanconi's Anemia A-FANCA
-			Fanconi's Anemia C-FANCC
-			Fanconi's Anemia G-FANCG
+			HNPCC*-MSH2
+			HNPCC*-MSH3
+			HNPCC*-MSH6
+			HNPCC*-MLH1 ←
+			HNPCC*-PMS2

This exercise will introduce you to online resources and databases. We will explore MLH1 (second from bottom of list) as an example. NOTE that in this exercise we will start our ENTREZ search with the general term “colon cancer” and then narrow our hits from there. But you will likely start out with more restrictive search terms in other searches

HOW and WHAT to submit to CT – see next page

Please submit this assignment via email. Each student will send a single word documents. Please include your last name the document name: **SmithGenesWeShare.docx**

NOTES FILE

Open a Word file and use it to take notes and keep track of what you have done. This file should serve as a substitute for recording this exercise in your lab notebook.

If you are instructed to save a figure or record of your analysis in the form of an image, use **Print Screen** which sends a copy of whatever is on the screen to the clipboard. If you want to copy only the active window, hold down the **Alt key**, then hit the **Print Screen** key (Fn + F11). It will copy only that window that is active, not the whole desktop.

🍏 If you are working on an Apple computer use these commands:

Snapshot of Screen: saved to clipboard:

*command (apple) * control * shift * 3*

Snapshot of selected portion of screen: saved to clipboard:

*command (apple) * control * shift * 4*

Here's what you need to submit for this assignment.

BUT, please feel free to browse through info that peaks your interest whether it is "required" for the assignment or not.

WORK-UP FILE Using the information you glean below (starts on pg 3) address these questions.

1. Translation of gene acronym, general biological function of gene product and chromosomal location of gene from OMIM. In your answer be sure to indicate what mut means and explain why this gene name makes sense in the context of the biological role of the gene function and how genes were named traditionally.

2. Answer question posed on pg 4: How could studying *E. coli* possibly help us understand anything about cancer in humans? Explicitly connect the primary phenotype resulting from mutations in this gene with the phenotype of cancer. 4-5 persuasive sentences maximum

If you haven't been introduced to cancer genetics in a previous course, browse through this lecture: *An Introduction to Cancer Genetics*

<http://fire.biol.wvu.edu/trent/trent/11.05.13lecture.pdf>

3. HOMOLOGENE output. Include screen shot of table (see step 12 under ENTREZ) and identify all organisms on the table (those with a proper genus species indicator..). What does the term homologous imply about the mutL genes in these organisms? [Don't know what homologous means, then look it up somewhere and tell me what source you used.] Also, why is E. coli not included in this list?
4. Length in nucleotides of the MTH1 (transcript 1) and length in amino acids of the corresponding protein. Why is the mRNA more than 3X the length of the polypeptide?
5. Review DNA replication and mismatch repair in your genetics text and, using the biochemical info found in your searches, describe the specific role of the product of the MTL1 gene in this process. A labeled diagram with a detailed caption is the best format for this question.
6. Include screen shot of conserved domains. For each domain (2) describe the biochemical function and how it relates to the proteins biological function. (2 sentences max for each domain).
7. Include a list of web sites visited other than the ones explored in this handout.

NOTE explanations appear in dialog boxes that often come up when you scroll over a link or term -- especially in ENTREZ

OMIM (*Online Mendial Inheritance in Man*)

OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. The full-text, referenced overviews in OMIM contain information on all known mendelian disorders and over 12,000 genes. OMIM focuses on the relationship between phenotype and genotype. It is updated daily, and the entries contain copious links to other genetics resources.

Go to OMIM, search for MLH1 and follow top link:

<http://www.ncbi.nlm.nih.gov/sites/entrez?db=OMIM>

1. Decode the MLH1 gene name (an acronym) and note chromosomal location of the gene as well.
 2. Note the biological process that this gene product is involved in and explain what **mut** stands for.
 3. Browse this page and note information that has special bearing on this question: ***How could studying E. coli help us understand anything about cancer in human?***
 4. On the OMIM page notice External links to the right. Browse through these links to see what type of information is available at your fingertips. Note the biochemical information available on the Protein/uniprot link -- especially the summary under **Ontologies**
-

ENTREZ is a data retrieval system developed by NCBI that provides integrated access to a wide range of data domains, including literature, nucleotide and protein sequences, complete genomes, three-dimensional structures, and more.

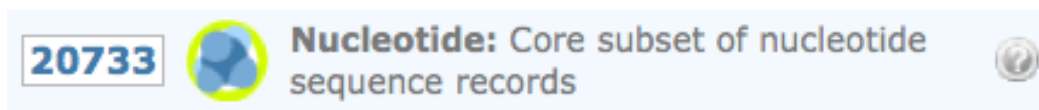
<http://www.ncbi.nlm.nih.gov/sites/gquery>

1. Enter “Colon Cancer” in the search field and select GO



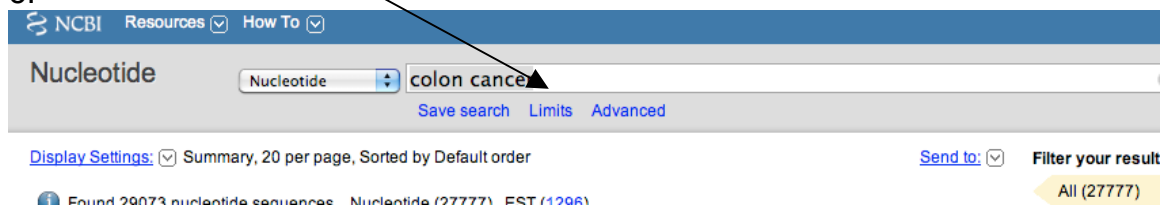
2. You should note that there are over 96,000 PubMed entries, >28,000 Nucleotide entries, and over 2700 Protein entries. Take a moment to familiarize yourself to the databases assayed by Entrez. We will learn how to narrow the search

3. Click on the “Nucleotide” field



4. Select the Limits tab to narrow the search.

- 5.



6. Limit the search to the Title field, to RefSeq (a curated database), and exclude all of the these...



- Hit Search. These limits reduce the number of hits to ~249--still too many unless you are really motivated. To further limit the search, return to the Limits page, search for human after selecting

Search for

“Organism” for the search field.

- Fields -

- Select the “Advanced” tab, Note your search history numbers and combine the title and organism searches

Builder

7 [Show index list](#)

AND 6 [Show index list](#)

AND [Show index list](#)

or [Add to history](#)

History [Clear](#)

Search	Add to builder	Query	Items found	
#7	Add	Search human Limits: exclude STSs, exclude working draft, exclude TPA, exclude patents, Source database: RefSeq Field: Organism	64670	1:
#6	Add	Search Colon Cancer Limits: exclude STSs, exclude working draft, exclude TPA, exclude patents, Source database: RefSeq Field: Title	241	1:

This limits the search to ~ 19 entries. We will pursue **NM_000249(MLH1 transcript variant 1)**—scroll down the list

- Browse through the information on the left side of the page. Scroll down and note gene summary information and amino acid and nucleotide (cDNA) sequences. Note that this page does not include the genomic sequence of this gene.
- Click on CDS (coding sequence) and the translated portion of the mRNA (cDNA) sequence is highlighted. Note the codons that define translation start and translation stop.


11. Browse through LINKS on the right side of the page.
12. Go to **Homologene** (homologous genes in other species) under Related Information . Check box #1 and click Download. TAKE A SCREEN SHOT OF THE TABLE THAT APPEARS
13. Go to Protein under Related Information:

[NP_000240](#) Reports
MutL protein homolog 1 [Homo sapiens]
gi|4557757|ref|NP_000240.1|[4557757]

Note: # of amino acids. Note: at the bottom of the page is the AA sequence.

13 Click on Identify Conserved Domains at the top right

Analyze this sequence
[Run BLAST](#)
[Identify Conserved Domains](#)
[Find in this Sequence](#)



Conserved Domains are comprised of protein sequences that code for a common functional unit, i.e. the active site of a protein, or a transmembrane domain, etc. Click on the domains to see their biological function.

TAKE a SCREEN SHOT OF THE GRAPHICAL SUMMARY OF THE CONSERVED DOMAINS

Scroll over the domain to see a dialog box describing its biological function.