Independent Project - Biol 310 - 2025

The main goal of your Independent Project is to determine if there is a **genetic model system**[1] (See Table 1) that could be used to better understand how *your* human **gene of interest** (GOI) functions *and* how it leads to disease when defective. **Each student pair (pairing up is optional) is responsible for completing one Poster.** You will use all the knowledge you have acquired throughout this semester in "Fantastic Genes and Where to Find Them" to complete this assignment.

Table 1.

Species	Organism Type	Advantages
Saccharomyces cerevisiae	yeast	Transparent, rapid life cycle, cheap
Schizosaccharomyces pombe	yeast	Transparent, rapid life cycle, cheap
Caenorhabditis elegans	round worm	Transparent, rapid life cycle, cheap
Drosophila melanogaster	fruit fly	Rich history of genetic analysis
Danio rerio	fish	Most similar to humans but expensive
Mus Musculus	mouse	Most similar to humans but expensive

Find a Human Gene of Interest (GOI) for your Independent Project

For your poster presentation to be successful, you need to choose a human gene that is known to have **pathogenic mutations** that lead to disease. How? I describe two methods in the following sections, but if you already have a gene in mind, you can easily check to see if it is suitable using the following simple method: Type the name of the human gene into the UCSC genome browser Search window using the "Sequence Variants" session link (from Chapter 7). If your gene name does NOT autocomplete, you either have a typo or the gene name you have is not the official name (genes often have alternate names). If your gene name DOES autocomplete, click on it (do not Click Search). A new Browser window will appear, if this gene contains pathogenic mutations you will see at least ONE vertical line, colored red below the gene prediction evidence track. If it has at least ONE pathogenic mutation, you can claim it! That said, it will be more exciting for you if you choose a gene with a missense mutation. Hover over the pathogenic mutations present to see if any are missense mutants. How do you recognize a missense allele? The popup window will include the text: "**Consequence: Missense Variant**." I will check to make sure your gene has an "ortholog" in one of the six genetic model systems listed above.

Use Medline Plus

Click on the following link to explore a list of genetic disorders in Medline Plus. From here you can search for a specific human disease or genetic disorder using the search window or you can browse a comprehensive list. Once you settle on a specific disorder, click the link. A menu of links will likely appear on the left, click on "Genetics". A list of genes will appear. Write them down. If the "Genetics" link does not appear, look within the "Causes" section of the disorder page to determine

[1] **A Model System**: is scientists' jargon for a particular species that has been developed over many years to be experimentally powerful to answer particular questions (Google).

if there is one or more specific genes known to cause the disorder (avoid genetic disorders that do not list specific gene names). Once you have a short list of possible human gene names, search each gene one at a time in the UCSC genome browser using the "Sequence Variants" session link to see if any one of these genes also contains pathogenic mutations (a DNA variant colored red). If you find a gene that has at least ONE pathogenic mutation, you can claim it. If you are trying to decide among several genes, choose the one that has at least one pathogenic missense mutation (i.e. Arg344Leu).

Alternative Method – Use the OrthoDisease 2.0 website.

To find a gene that is not unique to humans AND is linked to a specific disease, go to the OrthoDisease 2.0 website. Click on the "downloads" link then choose one of the six genetic model systems listed in Table 1 from the "Please select..." pull down menu (HINT: Start with yeast, it is the cheapest, most efficient species to model human disease). Once you have chosen a model organism, choose "Display Now" from the "Request as..." pulldown menu then click "Request". A large table will appear. This table lists all the human disease-associated genes with a **putative**[2] ortholog in the model system you chose to search. Repeat with the model system next on the list (Table 1) if you do not find anything of interest.

To retrieve the name of a *single* GOI from the list, first browse or search (control *or* command F) column one for a disease/disorder that interests you (i.e. Breast Cancer or Dementia). If you get zero results, check your spelling or choose a simple key word to search (i.e. Cancer). Once you settle on a single row, copy the text from **column five** within that row (NOTE: Column two may look like a gene name but it is a protein name and may or may not be the same as the gene name). ***Column 5 contains the unique ID code for the human gene. It should be in the following format: ENSG00000143341*.** You can use this **systematic gene ID** to search for your gene on the UCSC Genome Browser by pasting it into the search window using the "Sequence Variants" session link from Chapter 7 of "Fantastic Genes . . ". This will take you to the gene in the human genome database. Look to the left of the gene schematic to retrieve ***the official human gene name*** (i.e. HMCN1). Also, confirm that the gene has at least one pathogenic mutation (a DNA variant tickmark colored red positioned below the gene schematic). If you are trying to decide among several genes, choose the one that has at least one pathogenic missense mutation ("**Consequence: Missense Variant**").

What to do first once you choose a gene

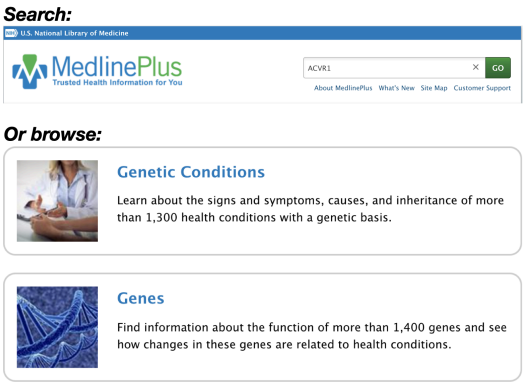
Once you find a GOI, claim it on the class excel spreadsheet (you must be logged into your Horizon account to edit). Insert **your name** (first and last), **official gene name** as provided by the UCSC genome browser *and* the **associated disease (phenotype)**. If your GOI is already taken you need to choose another gene. There are thousands to choose from. No pair of students (from Fall Semester) are allowed to work on the same human gene (**although you are welcome to work on the same disease**). If you are working as a pair, each person still needs to input a single row but with his or

[2] **Putative**: means "generally considered or reputed to be" or "possible". This is a word scientist often use to indicate they are not 100% certain of the claim. i.e. When I say "putative ortholog", this means that I am not 100% certain that the claim is correct: That the so-called ortholog is a true ortholog.

her name. If you are taking this class for a second time you are not allowed to choose the same gene from a previous year *unless you worked alone* AND the gene has not yet been chosen this year.

Learn more about the disease associated with your GOI

To learn more about a gene or disease, use **Medline Plus**. Keep in mind, you ***cannot*** search the **Medline Plus** database using the gene ID (i.e. ENSG00000143341). You must use the official gene name found at the UCSC genome browser site (i.e. BBS1).



When to Work on Your Independent Project

Once you choose a gene/human disease you are encouraged to read the literature to learn more about your gene, the gene product it encodes and the disease associated when mutant. You will not have all skills at the outset to use the UCSC genome browser to complete your project but you can certainly make headway. Then as you complete new chapters in the "Fantastic Genes . . . " manual you can apply what you learned to your gene of interest to begin to build your poster. **Review the "What to include in your poster" section below so you know what to look for.** If you have questions or doubts, you are welcome to ask during slow periods of the **Activity or during office hours. Finally, if you become interested in another gene later in the semester you are welcome to change your mind (up to a point - ask for the deadline). You can also change partners or decide to work alone. *Just change the information on the Excel Spreadsheet.***

How to find more information about your Gene of Choice

TBD

What to include in your poster

Use the Power Point template provided in the Activity folder in the Google Drive. Include the following elements described in detail below (BACKGROUND, QUESTION, RESULTS, CONCLUSION, METHODS and REFERENCES). Note that Optional images are highlighted in green and Required images/tables are highlighted in yellow. Also, see the example poster uploaded to the Google Drive.

BACKGROUND:

Provide information about your GENE and associated disease (including the following):

1. A written description of the gene structure including the following information
    a. Chromosomal location (which chromosome).
    b. Chromosomal Strand (+ strand or – strand, top or bottom strand).
    c. Number of unique mRNAs/isoforms produced by the gene.
    d. Number of exons/introns.
2. A written description of three ***notable*** features from the list below (your choice):
    a. start codon context of the main ORF (Is it optimal? adequate? neither?).
    b. The presence or absence* of **uORFs** (defined as the presence of out-of frame start codons within the 5' UTR).
        i. If present, you can include the start codon context as well.
    c. The presence or absence* of a PAS within 15-30 nt upstream of the cleavage site.
    d. Alternative transcriptional start site(s) supported by TSS seq data
    e. Alternative splice forms supported by RNAseq data but not predicted in the NCBI Refseq Evidence Track.

    Other NOTABLE features that are rare:

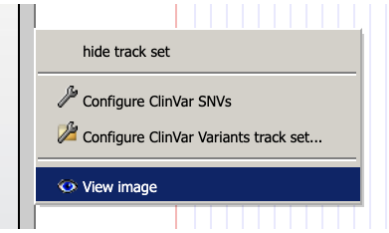
    f. A pathogenic **silent mutation.** Your screenshot should confirm that it is located to a coding exon.
    g. A pathogenic mutation that maps to noncoding (intron – not including splice site mutations, 5' UTR or 3' UTR). Your screenshot should confirm that the pathogenic mutation maps to noncoding.
    h. A noncanonical 5' or 3' splice site (***NOT*** GT or AG). Your screen shot must include a close up of the base position and gene prediction evidence tracks.
    i. The presence of transcriptional start site motif described in Chapter 6 and found in the expected location (rare and notable!)! Your screen shot should include the short match evidence track info and the gene prediction tracks.
    j. A Predicted Start Codon for the main ORF that is not an **ATG**!!! Your screen shot should include a close-up of the gene prediction and base prediction tracks to confirm that you are looking at the start codon for the main ORF.

3. A written description of your human disease of interest ***(Optional: include images if appropriate***)
4. A written description of the protein your gene encodes. How is it thought to function? Search Medline Plus, OMIM, Google Scholar and/or PubMed to learn more about how the gene/protein is thought to function. Keep track of your links and include them in your methods). **Optional: Include an image of the crystal structure if available – search for the gene on http://www.rcsb.org/).**

QUESTION: You are trying to determine if yeast, worms, flies, zebrafish and/or mice might be suitable to study your human gene of interest (find out how it functions) and/or study the human disease that is associated with your gene of interest. In other words, you are trying to determine if you can model your gene or disease of interest using a simpler model organism. ***Rewrite this into the question (See my sample poster). Your question will include your GOI, the human disease associated with this GOI and the genetic model organisms you are specifically examining.***

Figures and Tables:

5. Include a high-resolution image of the gene schematic. You will use this to illustrate your written description above. In a way, you are using it as evidence in support of your description and for the features of note you choose to write about above (See 6 for "how to")
6. HOW? Use the following session link to generate an ***image of your gene structure from the UCSC Genome Browser including the position of the pathogenic mutations and the TSS seq peaks***). **HINT**: To generate the image file, use the session link to find your gene. Right click on any gray evidence track rectangle on the far left of the browser window. Choose "View Image". A new window will open with the browser window only. Download this onto your computer. This is a high quality image of your browser window. This trick can be used to generate any desired image. You can also crop in powerpoint later if needed.



7. **At least three close-up screenshots** providing visible evidence for notable features described in the text (See 1 and 2 above). How do you know if you included the necessary information? If you think that I will have to go to the UCSC genome browser to confirm the presence or absence of these features then the answer is NO. *How can you ***provide evidence for the presence*** OR ***the absence*** of a gene feature? **Use the Short Match evidence track.** Take a screenshot. In your screenshot image, include the results of the short match search (the consensus sequence you search will be displayed in the title of the Short Match evidence track) along with surrounding elements (so you can prove that are searching in the right area).

8. **Include a table of at least 7 pathogenic mutations** (from the ClinVar short variant evidence track), It should be completed with accurate information concerning the following five columns:
    A. **Variant ID** (i.e. NM_024649.5(BBS1):c.434dup (p.Asp145fs)). This must be included! See template. I split it into DNA Variant ID and Protein Variant ID.
    B. **DNA Variant Type:** Input **transition**, **transversion** or **Indel**. You must figure this out.
    C. **Protein Variant Type:** Input **Nonsense**, **Missense** (**Conservative or nonconservative**), **Frameshift**, **Silent/Synonymous**, **Unknown** (if splice donor or splice acceptor) or N/A (if noncoding).
        a. *If possible, 5 should be pathogenic missense alleles* (conservative or nonconservative). If there are less than 5, **include ALL found.**
        b. **At least one should be a null allele*** (**nonsense**, **frameshift** or **splice site variant**) unless none have been found. Why include a null allele? This helps determine if the missense mutations found are nulls/hypomorphs as well (similar mutant phenotype) or perhaps .
        c. The last one is your choice.
    2. **Location** (i.e. Intron, Exon (near splice site junction or not), 5' Splice site, Promoter etc.). You have to figure this out.

D. **Phenotype:** Copy and paste the phenotype from the detailed information page that pops up when you click on a specfiic DNA variant. Phenotypes are often alleles specific.
E. **Inheritance Pattern**.
    a. *Go to Medline Plus*
    b. *Search for your gene (do not misspell it!)*
    c. *Open the "Health Conditions Related to Genetic Changes" tab.*
    d. *Find your health condition.*
    e. *Read about it but click "More about this health condition".*
    f. *Open the "Inheritance tab"*
    g. *Read what it says . . . . Don't worry if it says the disorder is not inherited. Those are usually caused by somatic mutations acquired during a person's lifetime. Just include the information it provides.*

9. **One RBH analysis per student author (each chooses a different genetic model organism from Table 1 excluding the mouse except see below).** Indicate on the poster who did which RBL. **Each student must do the following for his or her RBH analysis:**
    A. Include both pairwise alignments (First and Reciprocal) including a screen shot of the header information for each. First look in yeast, worms, flies or zebrafish. If no orthologs are found, choose mouse as one of the model organisms.
    B. Annotate one of the pairwise alignments with ALL the missense mutations listed in the Pathogenic mutation table or include a note that annotation is NOT possible because the patient alleles are not included in the alignment.

RESULTS (text):

10. Describe the results of your reciprocal best hit analysis: What is the name of the gene that was the top hit in your model system of choice? When you use this gene as a query in your reciprocal BLAST, what gene if your top hit in humans? In other words, did you find a reciprocal best hit in your model system of choice?
11. Comment on whether the pathogenic missense mutations are conserved (identical or similar) or not (unrelated).

CONCLUSION: Answer your question (ie. Do you think you can use worms and/or zebrafish to model your human disease of interest? Why or why not?) ***Cite concrete evidence that supports your conclusion: What did your reciprocal best hit analysis tell you? Is there an ortholog of your human gene in your model system of choice? Are the pathogenic mutations that cause human disease conserved in this ortholog?*** For inspiration, read *about the success of* yeast models of human PMM2 deficiency *here* https://archive.perlara.com/blog/pmm2-preprint-figures-part-1/ *(This website will not be updated but it is still very informative!).*

METHODS and REFERENCES: Briefly, describe the methods you used to perform a comparative sequence analysis (i.e. reciprocal best hit analysis). Also include all references and webpages.

Rubric for grading purposes (How to get 100% or 100 points)

	Pts
You chose a gene by the **deadline** that satisfies the following criteria: ● Has at least one pathogenic mutation leading to a human disease	4
All images are high resolution (See "SUPPLEMENTAL INSTRUCTIONS FOR POSTER PROJECT")	5
Your file includes the gene name, your name and the year in the title (in that order)	1
You provide an ***accurate*** written description of the following information for your gene of interest (GOI): ● chromosomal location ● strand (+ or -) ● number of isoforms (and a description of how they vary) ● number of exons / introns	10
You provide an ***accurate*** written description of what the gene encodes and how the protein function (in general terms).	10
You Provide an ***accurate*** description of one disease caused by pathogenic mutations (if more than one disease is caused by mutations in your gene, state this fact but no need to list them all)	10
You state a question (that makes sense)	5
You download an **image for a figure** that contains the **ENTIRE** gene schematic including the following three evidence tracks (formatted as pack, squish or dense as needed): ● **base position evidence track** (including chromosome information). ● **gene prediction evidence track** (NCBI refseq) ● **TSS seq evidence track** (zoom out to see the TSS seq data before you download image)	10
You provide an ***accurate*** written description (***and close-up image detail as proof***) of at least **three notable features** (including but not limited to): ● The context of the predicted start codon (optimal, adequate or neither). ● The presence of **uORFs, if any** (i.e. start codons in the 5' UTR). ● the context of the uORF start codons (optimal, adequate or neither) ● The presence or absence of a **PAS** within 15-30 nt upstream of the cleavage site (3' end of the transcript). ● The presence of a *noncanonical* 5' or 3' splice sites (***NOT a*** GT or AG but something unusual - this is rare FYI). ● A predicted transcriptional start site NOT supported by TSS seq data ● Alternative splice forms found by RNAseq not predicted via NCBI Refseq. ● A pathogenic **silent mutation!** or a rare pathogenic noncoding mutation (excluding splice site variants - those are not rare). ● The presence of transcriptional start site motifs described in Chapter 5 found in the expected location (as you have seen this is rare too)!	15
Includes a table of **pathogenic mutations,** completed with accurate information concerning the following five columns: ● **Variant ID** (i.e. NM_024649.5(BBS1):c.434dup (p.Asp145fs)). **This must be included!** ● **DNA mutation Type:** Input transition, transversion or Indel. You must figure this out.	10

● **Protein Variant Type:** Input Nonsense, Missense (Conservative or nonconservative), Frameshift, Silent, Unknown (if splice donor or splice acceptor) and N/A (if noncoding). ● **General Location** (i.e. Intron, Exon (near splice site junction or not), 5' Splice site, Promoter etc.) ● **Phenotype:** Copy and paste the phenotype from detailed information page that pops up with you click on a DNA variant. ● **Inheritance Pattern** (only if known – check Medline plus or OMIM but if nothing is found write ND for no data).	
If your gene has 7 or fewer ***pathogenic*** mutations you should include all of them. If your gene has more than 7 pathogenic mutations, you should include ● At least one null allele (nonsense, frameshift or splice site mutation) *unless none are found*. ● If you highlight a silent pathogenic mutation, intron (not splice site) or UTR variant as an interesting feature, you must include the variant in the table. ● Five missense alleles (conservative or nonconservative) unless fewer are found.	10
Includes one RBH analysis per student authors of the poster (If working as a pair, indicate who did what in the reference section of the presentation): ● Display the results of an RBH analysis involving a genetic model from Table 1: Display one pairwise alignment for the first BLAST and one pairwise alignment for the Reciprocal BLAST) including the **header information** (a screenshot of the BLAST results is fine). ● Locate then annotate (i.e. circle then label) **each** pathogenic missense mutation included in your table onto the First BLASTp alignment where Query is human. ● If you cannot find the amino acid at the expected position, you likely did your first BLASTP with the wrong isoform. Search with one of your missense alleles (Paste into the Search window). When the new window appears look for the isoform that has the name blacked out. This is the isoform you should use for the RBH.	10

*Click on the gray bar corresponding to the ClinVar Short Variants evidence track. At the top of the new window, you will see a yellow box. Within the "Molecular Consequences" pulldown menu, you can choose missense, nonsense and/or frameshift . Click submit. This trick allows you to limit the type of mutations you view.