

Gen 677 Semester Long Project

Spring 2009

The goal of this project is an in-depth, genomic and bioinformatic analysis of a human disease gene, which is homologous to one or more genes in one or more model organisms (the bacterium *E. coli*, the single-celled yeast, *S. cerevisiae*, the roundworm *C. elegans*, the arthropod *D. melanogaster*, the mouse *M. musculus*).

Keep in mind that your project should include a “global” analysis of your selected disease gene, in addition to the implied “local” analysis of a single, isolated gene. A global analysis reflects the central themes of the course: structural, functional (both transcriptomics and proteomics), and comparative genomics. A local analysis reflects the gene’s structure and function, and its primary role in the disease etiology. Both analysis types should reflect the intricate relationships among sequence, structure, function, genome organization and expression, and evolution.

Some specifics in your research should reflect the ways in which you have been studying genomics and bioinformatics. For example, any particular gene was once mapped and localized to a chromosome within the human genome; how was the gene discovered and mapped? How was the gene sequenced? What is its sequence? What are the molecular features and properties of this gene (i.e., describe and analyze its gene “anatomy”)? What is the sequence of the protein produced by this gene? What are the important and relevant structural domains and motifs in the protein, which are related to the disease pathology? Has the gene’s expression been characterized by Microarray analysis? How are these domains and motifs related to similar domains and motifs in the other, model organisms? Is the disease gene a member of a conserved gene family? Is the protein’s domain a member of a super-family of protein domains? These questions cover a full range of analysis, from cytogenetics, to gene and protein molecular characterizations, to functional and evolutionary comparisons with other genomes in other organisms.

You will of course be making extensive use of bioinformatics programs and databases found on the Internet (NCBI, etc.). My evaluation of your project will be based in part on the depth of your analyses, and on how well you use these available bioinformatics tools. But a tool is only a tool, and its use generates data that must be interpreted, analyzed, compared, and critiqued from a variety of perspectives. Finally, conclusions must be drawn along with recommendations for future experiments or analyses.

The format of your project should be that of a Web Page, so you will want to use **Weebly.com** (or any other web-based program you are familiar with) to construct the final version with all text and images. The length depends on the amount of information associated with a particular disease gene, and you should strive for brevity and clarity but not at the expense of completeness.

- Find a disease and an associated gene/protein for which they have found a mutation.

Genes and Disease:

<http://www.ncbi.nlm.nih.gov/books/bookres.fcgi/gnd/tocstatic.html>

OMIM (Online Mendelian Inheritance in Man):

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

- You will be working on this gene/protein for the rest of semester as your semester long project.
- You need to write a popular press review about what is known about this disease.
(see Popular Press Article First) <http://gen677.weebly.com/projects.html>
- Review a paper that initially discovered this gene/protein as being involved in this disease.
(search on PubMed for article) <http://www.ncbi.nlm.nih.gov/sites/entrez?db=pubmed>
- Compare and contrast popular press vs. a scientific article of your choosing
- You will present your findings at the end of the semester in a 15minute presentation to the class

Choosing your disease gene/protein?

Go to OMIM or Genes and Disease.

Genes and Disease:

<http://www.ncbi.nlm.nih.gov/books/bookres.fcgi/gnd/tocstatic.html>

OMIM(Online Mendelian Inheritance in Man):

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

ScienceDaily: (great place to find popular press articles)

<http://www.sciencedaily.com>

Popular Press Journals: New York Times, USA Today, your local newspaper, etc.

Example search try: huntington's disease as an example via Google (click on New York Times Article)

Find a diseases/trait and pick a gene that has been found to be associated with this disease. This will be your gene/protein that will be come your semester long project. You cannot choose the same gene/protein as someone else in class. Be sure to pick a disease and gene that has been written up in the popular press as you will need to find an article on it in the newspaper for your web project 2.

Please email me your top 3 choices ASAP. First come first serve. I will confirm with you whether or not you got your first choice via email. You will get your next choice if someone else in class already took it.

<u>Disease:</u>	<u>Examples of Genes Associated with Disease/Trait:</u>
Diabetes	IDDM2, PTPN1
Breast cancer	BRCA1 & BRCA2
Down's syndrome	amyloid precursor protein gene or App
Parkinson's disease	PARK2
Huntington's disease	HTT
Muscle ataxia	ATAXIN 3
Sterility	Dnahc8 (dynein HC 8), MSI,
Autism	NRCAM
Aging	BRCA1
Arthritis	Tnf, SLEB3
Colon cancer	APC
Blindness	HFI
Alcoholism	CREB
Retinitis pigmentosa	RPI
Nearsightedness	MYP3

Assignment:

Select a human genetic disorder, by which we mean a disease or heritable predisposition to a disease condition. It can be a single gene trait, or a polygenic one. In the case of a polygenic disorder, you need only do the genomic analysis for one of the involved genes. Also, you can and should feature that selected gene with greatest detail in describing the evidence that it is part of causation and reviewing the underlying genetic evidence for the claim. Some disorders are not polygenic traits, but you will find that more than one gene can cause or predispose different individuals to the disease (BRCA1 and BRCA2 for heritable predisposition to breast cancer, for example). In a case like this, you would introduce both genes as causing the predisposition, and then select one to focus on in detail for the rest of the web project. Be sure to choose a disorder that has a good popular press article associated with it.

Examples of what you can find out over the course of the semester:**I Background:**

- A. What is the phenotype of the disorder?
- B What is the genetics of the disorder in humans and what is the evidence?
- C. The genetic epidemiology of the disorder.

II. Genetic Evidence:

Give a thorough but succinct account of the evidence that this gene is causal (or partly causal) for the genetic disorder, including highlights of the most important evidence. Say why this particular evidence is powerful and convincing (if it is). Do you think the assignment of this gene to the disorder could be incorrect? Be critical in your account.

III. DNA Sequence

Go to Genbank and/or use both the human genome browsers to find the locus in human and mouse genomes. What is the sequence? See if you can find additional vertebrate orthologs (hamster, chicken, cow, etc).

IV. Homology and Phylogeny

Find out if your gene has suspected orthologous genes in other vertebrate species (which ones?). Are there similar genes/proteins in invertebrate model organisms (*Drosophila*, *S. cerevisiae*, *C. elegans*, *S. pombe*, *E. coli* etc.)? How sure are you that a candidate ortholog is indeed orthologous and why? Does your gene have candidate paralogous genes in man? What does the literature on orthologs of your gene from other model organisms say, if anything, about the function and evolution of function of your gene? You might want to go to Wormbase and Flybase (as an example) as a complement to information in the literature and in the human browser annotations.

V. Phenotypes & Gene Ontology

Mutant and/or RNAi phenotypes in other organisms
Cellular component, biological process and molecular function
Etc.

VI. Protein sequence and data

What is the amino acid sequence of your protein?
Is any amino acid mutated in your disease? If so, which ones?
What is the protein structure of your protein?
What do the protein domains tell you about what your protein does normally?
Has your protein been crystallized? If so what did they find out?

VII. Protein Network

What proteins are known to interact with your gene? Which ones do you find interesting? Why? Do you think your protein might be modified? If so, how and why?

VIII. Conclusion of your findings

Briefly go over what your main findings over the course of semester were for your gene/protein and how these findings might help you understand or create new approaches to understanding how this gene/protein functions normally and why the disease state might occur when this gene or protein is mutated or altered in some way. Conclude with a brief discussion of what you think is most important thing to find out next about this disorder and the gene(s) that cause it, and discuss why.

IX. Future Approaches

Outline how you would approach this future research, assuming that lab money and manpower are not your limitations.

X. Bibliography

Primary literature references from the journals are desirable when they are really the source: however when you are taking information from a review, text or website that – in turn – cites the primary source, you should cite the review (unless you really go see what the primary research paper says). Over-documenting where the credit comes from for an idea or fact is far better than under-documenting, but we do not want or need inflated bibliographies. If you use a web site, please be sure to credit the site. See Syllabus for how to cite online material.

References and sources that you used:

1. Any texts, journal articles, general articles and publications.
2. Any web sites used in a significant way (i.e., web sites that you used as primary or major sources of information and/or data). List the actual URL in the locator bar of the browser by copying and pasting into your report.
3. Include a separate list of all programs (or algorithms) used, with appropriate web site or other references; this list should be grouped into those programs for structural genomics, those for functional genomics, and those for comparative genomics. Part of the intent of this listing is to make certain that you know the distinctions among these various programs, in terms of their primary purposes. Also indicate on this list which programs have similar functions (e.g., analyzing promoter consensus sequences), but use different algorithms.
4. Within the body of the report (so not here in the bibliography), indicate the parameters used for any search or comparison program. This information is especially important if you adjust or change the default values on any program.

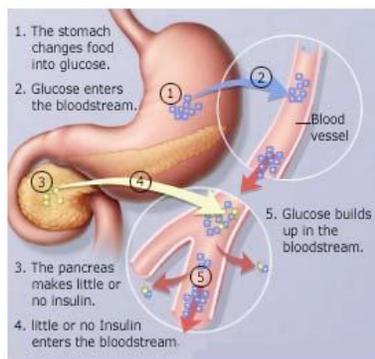
General outline, ideas & deadlines of your semester long project:

1. At the top of every page please put the following:
This web page was produced as an assignment for Gen677 at UW-Madison Spring 2009
2. At the bottom of each page please put the following:
Name, email, date page was last updated
Please link back to the genomics web page <http://www.gen677.weebly.com>

Home Page: (Due Feb 3rd)

Gene & Associated Disease Introduction

- Gene name
- Disease gene is associated with
- Picture of a person or tissue with disease and/or schematic on what is known (if available)



Example of schematic of what is known about diabetes
(<http://generalhealthinfo.wordpress.com/2008/08/14/let-natural-health-deal-with-your-diabetes/>)

Introduction

- Brief introduction to your disease and gene associated with it
- What is the prevalence of your disease in the human population?
- Is there a geographical location where this disease prevalent?
- In the last sentence, state what your project will focus on during the course of the semester.

Page 2: (Due Feb 12th)

Popular Press review:

Your first project is to select a trait that the popular press has dubbed "the gene for _____" (fill in the blank). For example the gene for deafness, Usher's Syndrome Type I (USH1B). The trait may be "a disease gene" but it does not have to be. You may choose your trait from any species as long as it has been published. You should find both a popular press citation and a scientific publication when choosing your trait/gene pair. I strongly encourage you to choose a species whose genome has been sequenced (at least in draft form if not finished completely). It probably will be easier to search for the popular press source first.

A popular press review should be from an science article in a magazine or newspaper such as *Discovery*, *ScienceNews.com*, *Scientific American* and/or *New York Times*, *Wisconsin State Journal*, etc. Popular press articles should address the relevant background information, the methods of the study, the results, and their importance to science. Be sure to cite the article title, author and reference. Paste the URL if there is one. If you can link the PDF of the article to your page here then great!

The purpose of this web page is to explore the commonalities and differences when a new gene discovery is announced. In your description, you should tell your readers about the trait as well as the gene/protein. Be sure to explore any other genes that may play a role in the trait you have chosen since this is one of the central features of genomics.

Before you write your web page, you should read this paper that describes what a good popular press science story should do: [Eunice Kua, Michael Reder, and Martha J. Grossel. 2004. Science in the News: A Study of Reporting Genomics.](#)

Public Understanding of Science, 13: 309–322. This paper can be found at: <http://gen677.weebly.com/projects.html>

Page 3: (Due Feb 26th)

Scientific Article review:

A scientific article review should be from a primary scientific journal such as *Science*, *Nature*, *JCB*, *Development*, etc. Summarize the article in your own words (do not repeat the abstract), what YOU learned from it and YOUR opinion of it. Then, go through the paper one figure at a time to explain to me the main points of each figure. This should not be a super long review of the paper but I want you to demonstrate your facility with the material and the big picture addressed in the paper. You are welcome to paste some of the figures (can copy figures as long as you reference them correctly) and describe their data in your own words.

Page 4: (Due Mar 5th)

Popular Press vs. Scientific Article

Here you should critique the popular press article and the scientific article and describe the differences between what the public knows and what you found out by reading the primary literature. It should be no more than a few paragraphs long.

Results and Methods:

In the next following pages can be up to you. Be creative! You need to find information about your gene/protein as you can throughout the semester and week by week as we go through these topics. You can use the databases we go over in the computer lab or find some on your own. It is up to you how thorough you are with this. Your grade will depend on how well you do your research. 5% points will be given for overall originality and aesthetics of website! Please be sure to reference the website (paste URL and type name of databases used) you got the info and figures from properly. Use figure legends please.

For example:

Describe what techniques/web sites/databases you are using and why you are using them. What did you find out from using a particular website? For example: You found the genomic sequence of your gene and aligned it using BLAST. You found that your protein is 56% identical to another protein in *C. elegans*. State why what you discovered is important and what this might mean in the context of your overall project. Suggest some organismal models that you might use to figure out what this gene/protein does.

Page 5: (Due on February 19th)

Gene Sequence, homology, phylogeny, motifs, etc.

Genomic Sequence

Homologous proteins in other organisms (similarity & identify %s)

Family tree

Sequence motifs

Etc.

Page 6: (Due on March 26th)

Phenotypes & Gene Ontology

Mutant and/or RNAi phenotypes in other organisms

Cellular component, biological process and molecular function

Etc.

Page 7: (Due on April 14th)

Protein data

Protein Sequence

Protein Structure

Protein Domains

Etc.

Page 8: (Due on April 21st)

Protein Network

If you can determine if your protein is in a network with other proteins using Osprey or other databases such as SMART, paste your data here and discuss what you found. Any interesting proteins that you found that bind to your protein? What are they? Any protein associated with your disease or others? Give few examples of how 2 of these proteins may function and what is known. How does this relate to what you already know about your protein. Discuss why you think this is interesting.

Pages 9: (Due on April 30th)

Misc. Data

Anything you think is relevant to understanding your gene/protein function. You can combine this data on to one page. Please separate with bold headlines.

Final Presentations (April 23rd-May7th)

Be wise and carefully organize and save your images and data elsewhere for your semester long project. You are going to present to the class what you learned and found out about your gene/protein in the last few weeks of class. You should propose the next steps of testing what you learned by suggesting some experiments that would further your work.

Final Web Site: (Due by May 10th)

Your final web project includes all of the assignments plus data that you obtained about your gene/protein over the course of the semester. Using bioinformatics approaches, you are to use what you learned from our in class discussions and in lab hands on experiences about a particular technique and online database to figure out everything there is to know about your gene/protein. By the end of the semester you should have discovered a lot. And by all means you are welcome to venture off to any online database that is available to you to try to understand and possibly come up with some ideas on how to “cure” your disease that you are studying.

At the end of the semester:

You need to have the following on your page, all properly referenced:

Introduction: (Homepage) write an overview of what is known about your gene/protein and what disease it causes and why it is important to understand how this gene/protein functions. (This should be on your home page and revise it if need be).

Results + Methods: (pages 5-9 +) What techniques/web sites/databases you are using and why you are using them? What did you find out from using a particular website? For example: You found the genomic sequence of your gene and aligned it using BLAST. You found that your protein is 56% identical to another protein in *C. elegans*. State why what you discovered is important and what this might mean in the context of your overall project. You can do this for each piece of data you find throughout the semester. Suggest some organismal models that you might use to figure out what this gene/protein does. Etc...

Discussion: (page 10) Briefly go over what your main findings over the course of semester were for your gene/protein and how these findings might help you understand or create new approaches to understanding how this gene/protein functions normally and why the disease state might occur when this gene or protein is mutated or altered in some way.

References: (page 10) Put the references to papers and sites that you used over the course of the semester on your project. Figures should be referenced as you use them on each page in a figure legend. Organize your references by each page in this section, use headings such as Page 1, Page 2, and/or Homology, Genomic Sequence, etc).

Genetics 677 Web site Construction

These assignments are designed to let you practice what you have learned during the semester. They are open-ended so you can experience the same uncertainty researchers do when initiating research on a new gene/protein. To maximize the educational value of these assignments, you may work in groups on the process of your *in silico* research, but the writing must be done on an individual basis. There are certain guidelines I want you to follow for all your web pages for this course:

Construction of websites: All web pages can be made with **Weebly** (<http://www.weebly.com>)-a free Website hosting site. If you have experience this other sites you can use them (google pages, etc) but I can not help you configure or troubleshoot pages in other programs. You must first sign up for Weebly. All you need is your email address.

Name your site: LastNameGen677S09 (ex skopgen677S09). This is for record keeping purposes. It let's me know what year you made your page and for which course.

Top of the page: At the top of every page, add this line; "This web page was produced as an assignment for Genetics 677, an undergraduate course at UW-Madison."

Bottom of your page: At the bottom of each page, you must put your contact information (Name, email, date page last updated) and a link back to the genomics web page <www.gen677.weebly.com>.

Backgrounds & Animations: Do not have any fancy backgrounds on any page. In general, it is better to keep web pages simple. Do not have any animations on any page other than your first page, unless it is a part of your presentation. See Bioinformatic Databases pages as an example.

Text Color: The color of text should be black on a light-colored background, for easy reading (see how bad this is?).

Font: I prefer Arial or Helvetica (sans serif fonts) as they are easier to read.

References: All references (e.g. Skop, et al., 2004) to WWW resources must be written in the proper format (as described here: <http://www.bedfordstmartins.com/online/cite8.html#1>) and must be hyperlinks if they are web references. You must use the (Name, Year) form of CBE style citations. Do put a comma between the name and the year, and do not say "and others" but rather I want you to use "et al. ". Do **not** use the style (Skop, page 1000). Full citations should be located at the bottom of each page and not on a separate reference page.

Figures: when appropriate reference figures (Fig. 1, Skop, 2004). Please be sure to list the references of all of your figures you use. You need to seek permission to use images. Please reference images that are not your own.

Grading of your page: The text is the most important part of your web page so concentrate on that rather than fancy layout. I will grade its functionality (i.e. do all the links work, are all the images visible, etc.) as well. Finally, you should know that I will weigh heavily on overall content, originality and aesthetics are a minor part of your grade. If you spend time making fancy website, that is OK as long as it enhances your ability to communicate to the reader.

Before you start: Before you create your first web assignment, you should take a few minutes to think about your overall web site. You are going to produce a website for your semester long project. To make it easier for everyone to read your pages, you should have a "front page" that you must call "home". From this home page, you can gradually add relative links to other pages to your home page as you produce them. This home page is where you may put any fancy things that you want to have fun with your web page.

Email me the URL of your site once it's set up. I will link the Gen677 website to your site and it will be easier for me if all of you use the same file name for this page.