Student A

Written Qualifying Exam for the Graduate Program in Genetics and Molecular Biology
January 9, 2017
Woodruff Health Sciences Library Computer Lab (Room B65)

This is a closed-book, closed-internet exam.

Read the entire exam before starting.

Answer one from each of the 5 pairs of questions (e.g. 1.1 and 1.2. are a pair). Each question is worth a total of 20 points. The breakdown of points for each question is clearly indicated. Please ensure that your answers address all parts of each question.

Your answers should be concise (but complete) and legible.

Start the answer for each question on a new page.

Label each question clearly and write your Student Letter (provided by Angela), *not your name*, on the first page of each question.

If writing by hand, use a notepad and a pen (not pencil).

Write only on one side of the page. Leave enough space in the margins so your answers do not get cut off on the copy machine.

Clearly indicate which Question and part of a Question any hand-written or drawn material applies to. Write your Student Letter (provided by Angela), *not your name*, on the each page of your answer.

Do not detach pages from your pad; you will turn in the entire pad at the end of the exam.

If using a computer to type answers, the final file(s) should be saved on the USB flash drive provided.

Use your Student Letter as the file title (ex: Student X).

Clearly indicate which Question and part of a Question your answer applies to.

During the exam only a word processing application may be active, with the file open in which you are typing your answers. No other programs may be used to access e-mail, internet or local files such as PowerPoint or PDFs. You are not to consult any notes or other written matter or consult with any person about your answers.

When you are done, put all your materials (exam, notepad, paper, USB flash drive) back in the envelope provided.

Envelopes must be returned to Ken Moberg by 6:00 PM (435 Whitehead). He will collect any exams from the room at 6:00PM but if you finish sooner, please take the exams to him in his office.

The Emory University Honor Code is in effect during the duration of the exam.

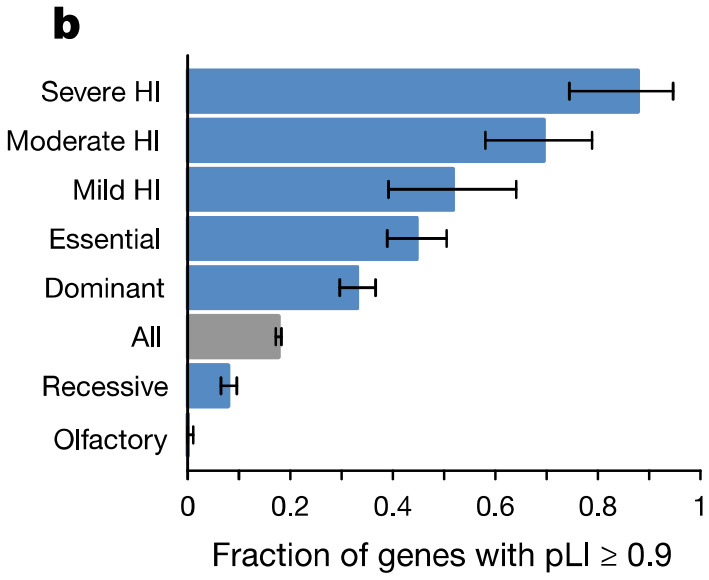
Catered lunch (12-1 PM) will be available in the Calhoun Room (at far end of hallway from B65)

If you have questions, call Ken (cell 404-217-7708) **GOOD LUCK!**

Student A

QUESTION 1.1

This year's Human Genetics class discussed "Analysis of protein-coding genetic variation in 60,706 humans" (Lek et al 2016, *Nature*). **Figure b** from this paper (below) shows the relationship between categories of genes (defined by the types of mutations at those loci, HI = haploinsufficent) as compared to the probability of being loss-of-function intolerant (pLI>0.9).



A. (2 pts) Was the pattern in Figure b expected?

B. (4 pts) Briefly define the categories on the Y axis of Figure b and explain why you expected or did not expect to see the result in the figure.

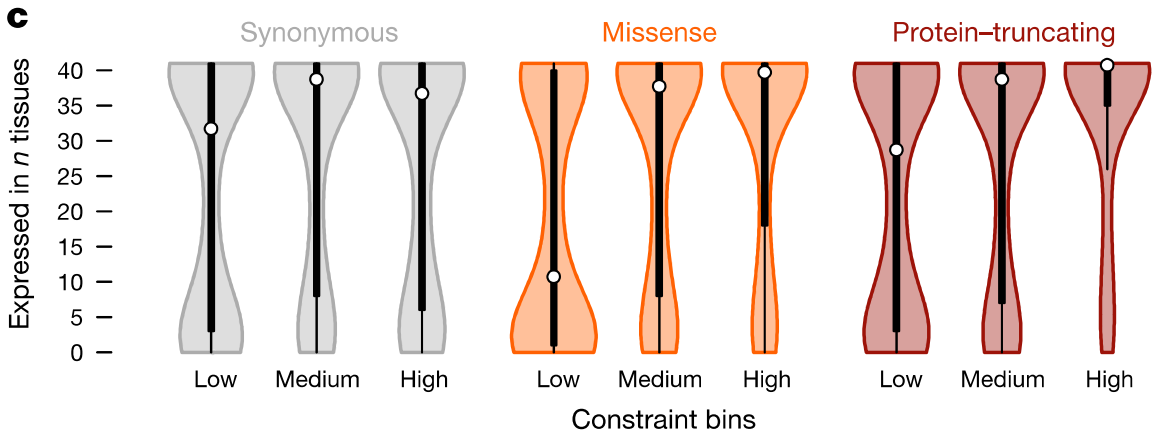


Figure c above is also from the Lek et al 2016 paper we read.

C. (7 pts) Define the three types of mutations contained in Figure c in terms of how they impact the predicted protein.

D. (7 pts) What conclusion can you draw about patterns of evolutionary constraint in these different classes of mutations?

QUESTION 1.2

You recover a mutation "x" in *Drosophila* gene A that maps to chromosome 2 and confers an interesting phenotype in x/+ heterozygotes, which are extra bristles on the dorsal thorax (ie back) of adults. The x/x homozygotes also have this interesting phenotype, although it is even stronger than in x/+ adults. You recover a second mutation "y" that also maps to chromosome 2 and affects bristle abundance. In this case, y/+ does not have a phenotype but y/y animals lack back bristles, which is the inverse phenotype of x/+ animals.

A. (4 pts) In classic genetic terms, what types of mutations do you think “x” and “y” are, and what tests would you perform to test these hypotheses?

B. (4 pts) Provide two (2) distinct models of how the x and y mutations are related to each other in the process of bristle development.

C. (6 pts) Discuss how you could distinguish if x and y are alleles of the same or different genes using classical genetic techniques.

Upon sequencing the gene A locus (proximal promoter/enhancer, introns and exons) in the x/x and y/y animals, you identify a single base change in each background that differs from the wild type sequence of gene A.

D. (6 pts) Assuming that these base changes are causative for the bristle phenotypes and for the genetic behavior of the "x" and “y” mutations, what types of molecular changes do you predict x and y cause within gene A? Please give an example of each molecular change that you choose.

QUESTION 2.1

Your studies have led you to focus your thesis research on a transcription factor that is regulated by stress. You have termed this transcription factor 'Stress-Activated Transcription Factor' or SATF. Experiments you performed in a cultured human cell line reveal that the steady-state level of the SATF protein increases 10-fold when you stress the cells by a variety of techniques, including elevated temperature or lowered O_2 levels.

You wish to define the mechanism through which stress regulates the steady-state level of the SATF protein.

A. (4 pts) Drawing on your knowledge of mechanisms that impact endpoint protein levels, suggest four (4) possible mechanisms that could lead to a increase in the steady-state level of SATF protein. For this question, consider fundamental, central processes that affect protein levels.

B. (2 pts) To follow up on your suggestions from Part (A), you perform a northern blot to examine *SATF* mRNA. Results of this experiment reveal that levels of *SATF* mRNA do not change in response to stress. Of the four (4) mechanisms that you suggested in part (A), which one (or more) of these do you now exclude as the least likely possible modes of regulation and why?

C. (8 pts) Drawing on the remaining possible mechanisms, state one (1) hypothesis for how you think steady-state SATF protein levels are regulated by stress. Describe an experimental approach to test your hypothesis. Be certain to include appropriate controls for your approach. Describe experimental results that would *support* your hypothesis AND also describe results that would *not support* your hypothesis.

D. (6 pts) Finally, you'd like to test what role, if any, SAFT plays in regulating cell survival in response to transient stress (for example, 30 minutes in low O_2). Describe an experimental approach that would help you address this question. Be certain to include appropriate controls. Describe the results you would expect if SATF *is required* for cells to survive your chosen stress AND also describe results you would expect if SATF *is not required* for cells to survive your chosen stress.

QUESTION 2.2

Genome-wide association studies in a Scandinavian population have implicated a 100 kb region of chromosome 11 in retinal degeneration and early blindness. The only known gene in this region of the human genome encodes a transcription factor known as KBX12. The *Drosophila* genome contains a single ortholog of human KBX12, and this organism also has eyes with a simple retina consisting of two easily distinguishable photoreceptor types (types I and II).

In studying the role of KBX12 in *Drosophila* eye development, you must consider the following questions:

A. (3 pts) How would you determine experimentally whether *Drosophila* KBX12 (dKBX12) plays a role in retinal development in flies? Do not simply name a technique. Describe any crossing schemes and relevant genotypes in your answer, as well as your interpretations of possible outcomes.

For questions **B**, **C** and **D**, you may assume that genetic tools and/or molecular reagents necessary to carry out your experiments are available.

B. (4 pts) You find that mutations in *dKBX12* affect the specification of both photoreceptor types in the fly eye (types I and II), leading to their incomplete differentiation. Given this observation, you would like to purify dKBX12 and identify any proteins that associate with it. Describe how you would perform this experiment using either *Drosophila* cultured S2 cells or whole flies, including justification of your choice of starting material.

C. (5 pts) You find that dKBX12 protein associates with 6 proteins that are known to be components of a histone methyltransferase complex. You note that 4 of these 6 subunits are encoded by multiple genes, suggesting that several distinct forms of the KBX12 complex may be present in the eye. How would you determine if there are indeed multiple forms of the complex present in the tissue you purified the complex from? How would you determine if such distinct complexes occur exclusively in one photoreceptor cell type or the other?

D. (5 pts) You suspect that this dKBX12-containing complex is involved in the regulation of cell type-specific genes during photoreceptor fate specification. How would you determine which genes dKBX12 acts on and how it affects expression of its target genes?

E. (3 pts) Studies of *KBX12* polymorphisms in human populations worldwide have shown that mutations in this gene are not only associated with retinal defects, but also autism and schizophrenia. Provide a general explanation for how a single gene, such as KBX12, could be implicated in such disparate conditions.

QUESTION 3.1

You are interested in how the spatial organization of the DNA in a nucleus changes as a lineage develops from a pluripotent stem cell into a neuron. You have all of the tools necessary to induce differentiation of stem cells into neurons *in vitro.*

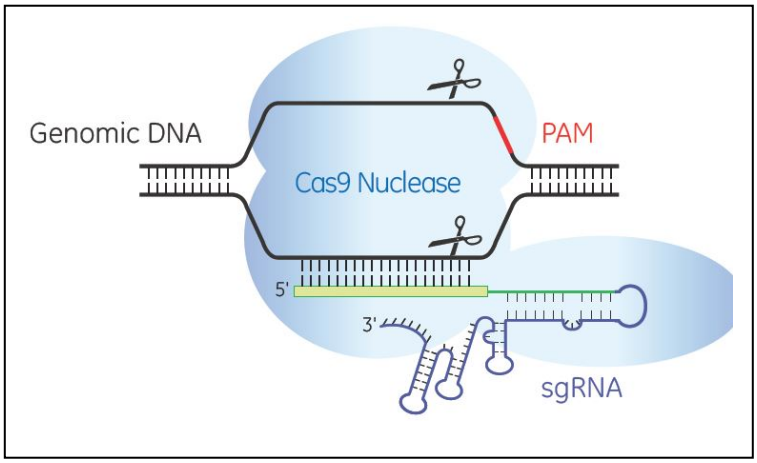
A) (5 pts)
What type of cell would you start with?
How would you obtain this type of cell?
What characteristics would you test to make sure the cells you obtained were pluripotent?

B) (8 pts) Outline the steps of an assay would you use to determine the genome's spatial organization and detect important changes in this organization as the cells differentiate into neurons. Include controls and explain any caveats you might expect from such assays.

C) (7 pts) Suppose you observe specific nuclear organizational differences between stem cells and neurons in regions containing genes known to be important for neurogenesis. Propose one hypothesis that explains these differences, and outline an experiment to test your hypothesis.

QUESTION 3.2

The CRISPR/Cas9 gene editing technology has emerged as a powerful technique with many applications in molecular biology and beyond. The illustration below shows a general overview of the mechanism by which Cas9 functions together with a synthetic guide RNA.



A. (2 pts) There have been other genome editing mechanisms used throughout the years (e.g. zinc finger nucleases) that showed promise. What is it about the CRISPR/Cas9 system that offers the potential for such exquisite site-specific editing that was not a component of prior approaches to genome engineering?

B. (2 pts) A modified version of the Cas9 protein exists in which a single nuclease domain (indicated by the scissors) is inactivated. What would be the possible advantage of using such a form of Cas9 (2 points)?

C. (2 pts) Another modified version of the Cas9 protein exists in which both nuclease domains (indicated by the scissors) are inactivated, so called “dead” Cas9 (dCas9). Suggest one experimental application for this modified dCas9 protein.

D. (4 pts) Name two specific challenges or concerns about using CRISPR/Cas9 as a potential therapeutic tool for genome editing. What is the basis for these challenges/concerns?

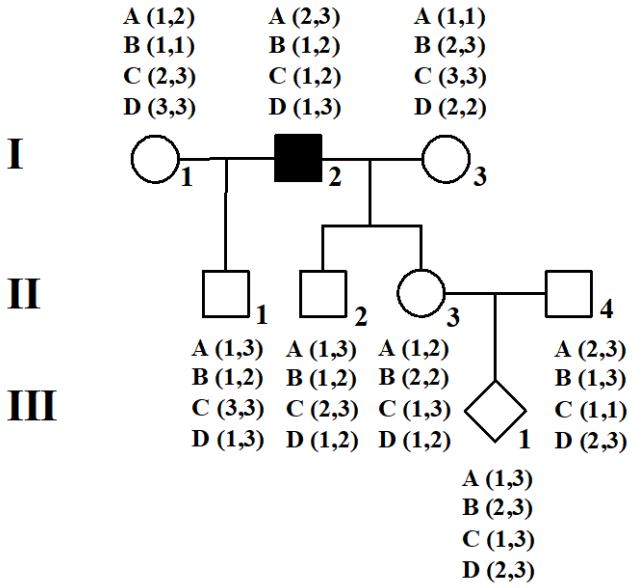
E. (4 pts) If you were treating patients with CRISPR/Cas9, would you choose to target a disease of the blood or a muscle disease? Provide the rationale for your choice.

The first attempt to use CRISPR/Cas9 genome editing for treating patients was carried out by Chinese scientists in the fall of 2016. These scientists isolated T-cells from a patient with cancer and used CRISPR/Cas9 genome editing to target the PD-1 gene and remove PD-1 protein in these cells. These cells, which are the patient’s own cells, were re-injected into the patients. The PD-1 protein modulates the immune response of the T-cells so these modified T cells should have the ability to mount a strong immune response against the cancer in these patients.

F. (6 pts) Suggest a therapeutic approach in which you could implement the power o**f** CRISPR/Cas9. Provide details of what gene you would target, how you would edit the genome, and how you would actually introduce the modification to your patients. What cells are you targeting and why? Provide rationale for why you have opted to target this specific disease/condition.

QUESTION 4.1

The pedigree illustrated below depicts three generations of an extended family; this pedigree was drawn by a genetic counselor from information provided by person II-3 during a recent appointment. DNA studies reveal the genotypes of the represented individuals at loci A, B, C, and D, as indicated. The genotype at each locus is presented in ascending numerical order; the order of presentation is not intended to reflect which alleles are in phase. All four loci (A, B, C, D) map to the short arm of chromosome 4.



A. (8 pts total) For each person <u>in the second generation</u> (II-1, II-2, II-3, and II-4) fill in the table provided to assign presumed parent of origin for each allele at each of the four loci (A, B, C, and D) and explain how you reached this assignment. If no assignment is possible, explain why.

(2pts) Person II-1:

Locus	Allele	Parent of origin (mat, pat, or unknown)	Comments
A			
B			
C			
D			

(2pts) Person II-2:

Locus	Allele	Parent of origin (mat, pat, or unknown)	Comments
A			
B			
C			
D			

(2pts) Person II-3:

Locus	Allele	Parent of origin (mat, pat, or unknown)	Comments
A			
B			
C			
D			

(2pts) Person II-4:

Locus	Allele	Parent of origin (mat, pat, or unknown)	Comments
A			
B			
C			
D			

B. (2 pts) Person II-3 is 12 weeks pregnant with her first child. Her father (person I-2) was just diagnosed with Emory Disease, a very rare late adult onset autosomal dominant disorder known to result from mutations that map to chromosome 4 between loci C and D. The population risk for Emory Disease is 1 in 4,000,000 for the relevant population. Emory Disease is fully penetrant with no new mutations reported.

In a panic over the risk to her unborn child, Person II-3 has CVS performed; the presumed fetal genotype results are indicated on the pedigree (**the fetus is person III-1**).

Fill in the table below to assign parent of origin for each allele identified in the CVS sample.

Locus	Allele	Parent of origin (mat, pat, or unknown)	Comments
A			
B			
C			
D			

C. (10 pts total) From the information presented, what is the risk for each of the following individuals of eventually having Emory Disease? Explain your reasoning.

(2pts) Person II-1?

(2 pts) Person II-2?

(2 pts) Person II-3?

(4 pts) The unborn child (person III-1)?

QUESTION 4.2

Cardiomyopathies are characterized by mechanical and electrical dysfunction of heart muscle, with accompanying heart failure and risk of sudden death. A significant portion of cardiomyopathies are thought to have a genetic cause, but the genes involved in these cases are not well defined. Mutations in genes encoding protein components of the sarcomere can act as autosomal dominant 'cardiomyopathy genes', but these account for less than half of the genetic cases of the disease. Thus, there is a need to identify new cardiomyopathy genes.

Exome sequencing of a large number of cardiomyopathy patients and unaffected controls has identified an amino acid variant (a valine-to-aspartate (V-to-D) change) in a sarcomeric protein expressed in the heart called MLP2. You have a family in which the father and the son have the V-to-D variant as heterozygotes, and each have cardiomyopathy. However, you would like to seek evidence that the V-to-D variant actually causes the disease. *C. elegans* does not have a heart, but its body wall muscle is striated muscle like human cardiac muscle. Moreover, it has a single clear ortholog of human MLP2, called nematode MLP-2 (or nMLP-2), with a valine residue at the relevant position within the protein sequence.

A. (4 pts) Using *C. elegans*, how would you test the hypothesis that the V-to-D change results in muscle dysfunction? In your answer, please explain how the mutation would be generated AND how you would assess muscle structure and function.

B. (4 pts) List two mechanisms in which the V-to-D form of nMLP-2 could act dominantly AND what *C. elegans* genetic tests you can use to distinguish these two possibilities?

C. (4 pts) Choose one (1) of your mechanisms in Part (B) and explain how a missense mutation like V-to-D could cause it. Assume that a specific antibody is available to nMLP-2. What biochemical experiment can you do that would support or refute your mechanism? Explain your answer.

D. (6 pts) You determine that heterozygous and homozygous worms for the *nmlp-2*$^{V\text{-}to\text{-}D}$ allele have the same, easily observed defect in body wall muscle. Beginning with the homozygote, explain two (2) ways in which you could conduct a suppressor screen to identify loss-of-function mutations in other genes that suppress (i.e. rescue) this muscle phenotype. Explain how you would identify the gene(s) involved. Where necessary, assume that the modifiers themselves will have easily detected recessive phenotypes in a wildtype background.

E. (2 pts) Briefly explain one (1) possible mechanism of suppression by a modifier allele from Part (D).

QUESTION 5.1

A novel gene named *AMD* (anxiety movement disorder) consists of 5 exons that encode a protein of 1000 amino acids. The structure of the *AMD* gene is highly conserved between mouse and humans, and is shown below.

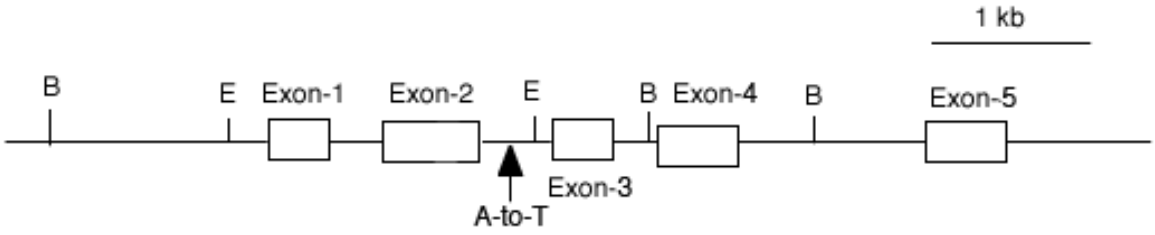


Figure legend: The black arrow indicates an A-to-T mutation within the 2nd intron of the *AMD* gene that has been found in human patients with a heritable anxiety disorder. Restriction enzyme digestion sites are also shown: B: Bam HI, E: EcoRI.

A. (4 pts) AMD is thought to be specifically expressed in the mammalian brain. However, there is no available anti-AMD antibody to test this hypothesis. Please describe two alternate methods to verify *AMD* expression in the mouse brain. If you need to use any probes, please indicate the location of the probes within the *AMD* gene.

B. (6 pts) The function of AMD is unknown. However, the A-to-T mutation within the 2nd intron of the *AMD* gene leads to an anxiety phenotype in heterozygous patients. Homozygous patients display a movement disorder in addition to anxiety. Please provide one (1) potential explanation for how this A-to-T intronic mutation affects *AMD* expression AND why it causes different phenotypes in heterozygous and homozygous patients.

C. (4 pts) Select one (1) genetic engineering approach to create a mouse that models the AMD anxiety disorder. Do not simply state the name of your approach. Explain it in sufficient technical detail to highlight the key steps and reagents that are central to its implementation. If needed, provide pictorial models to show key elements of your approach.

D. (2 pts) Describe one (1) experiment you could perform to molecularly verify that your new allele affects *AMD* expression in the mouse brain in the manner you predicted in Part (B).

E. (2 pts) Describe one (1) technique you could use to test the phenotypic effect of your new *AMD* allele.

F. (2 pts) Assuming that the A-to-T allele acts as an AMD 'loss of function' in vivo, describe a genetic approach to testing the effect of inactivating AMD *only* in neuronal cells in the mouse brain.

QUESTION 5.2

You are interested in identifying genetic associations with Disease X. You perform a GWAS (genome-wide association study) of 50,000 SNPs on a set of 2000 unrelated patients and 2000 unrelated controls.

A. (1 pt) To maintain an experiment-wide type I error rate of .05, what α-level (significance threshold for p-values) should be used? Show any work. (**1 pt**)

Note that the appropriate α-level derived in a) corresponds to a chi-square statistic cutoff of 23.9; this will be useful in answering the rest of this question.

Scenario 1:
Chi-square statistic for most significant SNP is 24.0
Genomic control inflation factor (λ_{GC}) = 1.2

B. (4 pts) What problem in your dataset is suggested by the genomic control inflation factor? Explain briefly how this problem could lead to genomic inflation.

C. (3 pts) How would your answer to (B) change if the GWAS had been performed on parent-child trios using the Transmission Disequilibrium Test (TDT) instead of the unrelated case/control design? Explain your reasoning.

D. (4 pts) If you use the method of genomic control to adjust the test statistics for inflation, will the top most significant SNP with the chi-square statistics of 24.0 reach genome-wide significance? Explain you answer and show any work.

Scenario 2:
Chi-square statistic for most significant SNP is 40.2
Genomic control inflation factor (λ_{GC}) = 0.98

E. (4 pts) Although your GWAS included several SNPs in high LD (linkage disequilibrium) with your most significant SNP ($r^2 > 0.9$), none of these other SNPs show strong association with the disease (chi-square statistics < 8). What possible problem is indicated by this pattern? Explain your answer.

F. (2 pts) Propose a strategy to check the validity of the result for your most significant SNP using the data that you already have (i.e., using the case/control SNP genotypes).

G. (2 pts) Propose a strategy to check the validity of the result for your most significant SNP that involves generation of additional data.