Written Qualifying Exam for the Graduate Program in Genetics and Molecular Biology (GMB)
January 11, 2019
Woodruff Health Sciences Library Computer Lab (Room B65)

Read the entire exam before starting.

Answer one question from each of the five (5) pairs of questions. 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 each part of the question.

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

Start the answer for each question on a new page according to detailed directions.

Label each question clearly and write your Student Letter (provided by Roberta) on the first page of each question.

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

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

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

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.

If writing by hand, use the notepad provided and a pen.

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

Clearly indicate which Question and part of a Question the hand written or drawn material applies to.

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

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

Envelopes must be returned to Guy Benian by 6:00 PM (105-E 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.

A continental breakfast (8:30 AM) and lunch (12-1 PM) will be available in the Biochemistry Connector at the table area near the vending machines and restrooms (between Whitehead & Dental School Building)

If you have questions, call Anita at 404-421-9061 (cell) or Guy 404-455-6948 (cell)/404-727-5953 (office).

GOOD LUCK!

Question 1A

A mutation in worms leads to a loss of all posterior muscle cells. You map the mutation to a small region on Chromosome III. Using the genome sequence, you identify several genes in the region that appear to encode transcription factors, one of which has a homologue in *Drosophila* (*dHox-Z*), which previous studies have shown is involved in posterior cell fate determination.

A) Describe experiment(s) that would demonstrate unambiguously that the phenotype you observe in your mutant worms is due to a mutation in the worm *dHox-Z*. Be sure to include appropriate controls in these and all subsequent experiments. (**4 points**)

B) Assume you have shown that the mutation is in the worm *Hox-Z* (*wHox-Z*) gene. Describe an experimental approach to determine whether the expression pattern of this gene fits with its predicted function. Detail how you generate any reagents you would require for this experiment and describe results that would support a role for wHox-Z in determining the fate of these posterior muscle cells. (**4 points**)

C) You suspect that the function of the gene is conserved in *Drosophila*. Describe an experimental approach to test this idea. Be sure to include appropriate controls. (**4 points**)

D) You strongly suspect that the *wHox-Z* gene product (wHOX-Z) is a transcription factor that regulates the expression of a variety of genes that lead to the specification of posterior muscle cells. Design an experiment that would allow you to identify candidate genes that require wHOX-Z for their expression. (**4 points**)

E) Pick <u>one of the gene candidates</u> you identified in (**D**) and describe an experimental approach to provide evidence that this gene is a direct target of wHOX-Z. Note that your lab has limited resources so you want to test only this specific candidate target gene in addition to any required controls. (**4 points**)

Question 1B

During *Drosophila* development, a certain pool of stem cells gives rise to two distinct differentiated cell types (types A and B). Gene *X* is strongly expressed in cell type A but is silenced in cell type B.

A) Propose a mechanistic model to explain the transcriptional state of gene *X* in each of these two cell types. The model should invoke sequence-specific DNA binding proteins as well as any specific histone modifications that you think could play a role for each cell type, and should account for the observation that the daughters of cell types A and B heritably maintain the parental activity state of gene *X*. (**3 points**)

B) Describe an experimental approach to define the transcriptomes of cell types A and B to determine how they differ from one another. Consider that these cell types are found within the body of the fly, mixed in among many other cell types; assume all reagents and techniques you need or propose are available. (**5 points**)

C) Propose an experimental approach to test a mechanistic model that involves epigenetic modes of gene regulation differences in cell types A and B. Describe an experimental outcome that would strongly support your model. Include all appropriate controls. (**7 points**)

D) Describe an approach that would allow you to identify factors that are genetically required for the differentiation of cell type A. In your description of the approach you choose, include details of any assays required, the phenotypes you are expecting, and controls required for your assay(s). (**5 points**)

Question 2A

In David Bartel's GMB seminar, he described the function of microRNAs (miRNAs), a class of RNAs that he discovered and named.

A) You have recently discovered a new miRNA. You discovered this miRNA in an RNA-Seq experiment to discover small RNAs. How could you begin to identify candidate mRNAs that could be regulated by this miRNA? (**2 points**)

B) Dr. Bartel talked about how miRNAs typically regulate mRNA targets by triggering decapping and degradation. However, he described how miRNAs can also regulate mRNA at the level of blocking translation. Describe an experimental approach to determine which of these two mechanisms the miRNA you have discovered uses to regulate one of the mRNA targets that you have identified. Provide a description in words or pictures (ideally both) of key experiments and their outcomes that would support one or the other regulatory mechanism between your miRNA/mRNA pair. Be sure to include appropriate controls to ensure the rigor and reproducibility of your results. (**10 points**)

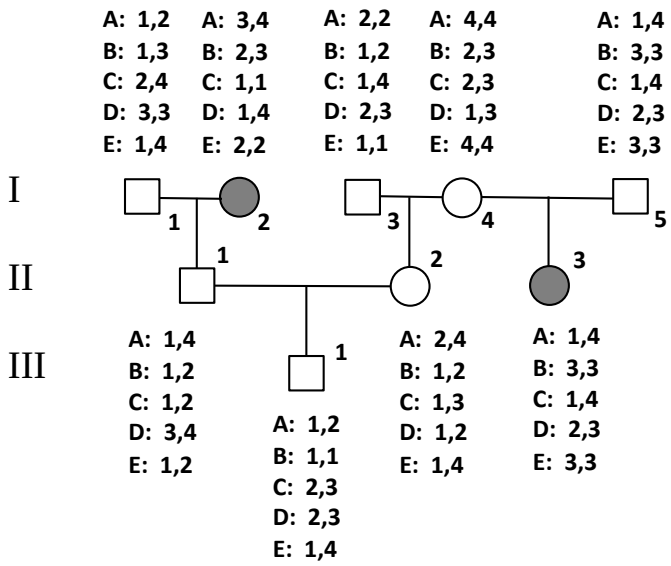
C) You now want to extend your results to provide evidence that the miRNA you have discovered **directly** regulates your candidate mRNA target. Describe an experimental approach that would test the hypothesis that the miRNA you have discovered regulates the candidate mRNA target through direct pairing with the target mRNA. Be sure to include controls that will allow you to robustly test for this mechanism of direct regulation. (**6 points**)

D) Dr. Bartel said that knockout mice for individual members of larger miRNA families often have broad phenotypes with multiple tissues affected. Suggest a model for why loss of a single miRNA, or even multiple members of a miRNA family, could have broad consequences. (**2 points**)

Question 2B

The two shaded symbols in the pedigree represent individuals affected by an extremely rare autosomal recessive, adult onset disorder called Bird's Disease (BD). Assume there is no other known family history for BD, all biological relationships are as presented, BD shows full penetrance by age 25, and there are no new mutations. Everyone in the pedigree except III-1 is over 25; III-1 is 2 years old.

Next to each symbol in the pedigree, you will see the genotype information for that individual at each of five (5) loci labeled A, B, C, D, and E. The gene responsible for BD has yet to be cloned but is known to be tightly linked to locus D.



A) Indicate below what percentage of alleles are shared, identity by descent (IBD), by each the indicated pairs of people from this pedigree. Assume none of the reproductive pairs are consanguineous. **(4 points)**

I-1 and II-1
I-1 and III-1
II-2 and II-3
II-3 and III-1

B) Given the information provided in the problem, what is the risk that person II-1 is a carrier for a BD mutation? Explain. **(4 points)**

C) Given the information provided, what is the risk that person II-2 is a carrier for a BD mutation? Explain. **(4 points)**

D) If you had the pedigree but did not have the genotype information, what would you estimate is the likelihood that person III-1 will be affected with BD when he is 25? Show your work. **(4 points)**

E) Does having the genotype information provided in the figure raise or lower your estimate? Explain. **(4 points)**

Question 3A

Your lab investigates a rare form of muscle wasting disease that is caused by loss of function mutations in the *MWD1* gene. Your studies take advantage of a mouse model of this muscle wasting disease where the mice model the human disease and lack any function Mwd1 protein.

Previous studies have demonstrated that the MWD1 protein is highly expressed in motor neurons and localization studies reveal that the MWD1 protein is enriched within neurite extensions in cultured mouse or human primary motor neurons. Based on your knowledge of local protein expression and RNA trafficking, you suspect that the *Mwd1* mRNA may be trafficked by interacting with specific RNA binding proteins.

A) Describe an approach to identify candidate RNA binding proteins that could bind to and regulate the *Mwd1* mRNA transcript. Be sure to include appropriate controls to ensure that the candidate RNA binding protein has specificity for the *Mwd1* transcript and is not just a general RNA binding protein. (**4 points**)

B) Your approach that you described in (A) identifies an excellent candidate, a putative RNA binding protein called Brawny. While there is no structural information about the Brawny protein, Brawny does contain a well-defined RNA binding domain consisting of two RNA Recognition Motifs (RRMs). Your initial characterization of Brawny reveals that Brawny localizes to the dendritic tips of motor neurons. This information taken together with your finding that Brawny binds to the *Mwd1* mRNA suggests to you the hypothesis that Brawny protein could be required for localization of Mwd1 mRNA to the dendritic tips of motor neurons. Design an experiment that would determine if Brawny is required for *Mwd1* mRNA localization in motor neurons. Be sure to describe appropriate controls. (**5 points**)

C) The 3'UTR of the *Mwd1* mRNA has a complex structure (**Figure 1A**). Design an experiment to test whether Brawny binds specifically to the *Mwd1* 3'UTR. Assume that you have access to purified, recombinant Brawny protein and you need to test for direct binding between Brawny and the *Mwd1* 3'UTR. (**5 points**)

D) The *Mwd1* 3'UTR is predicted to form stem loops (see **Figure 1A**). How would you test which stem loop is required for Brawny binding? How would you test whether the stem loop you identify is sufficient for binding to Brawny? Be sure to describe appropriate controls. (**4 points**)

E) Design an experiment to test whether Brawny regulates steady-state *Mwd1* mRNA levels. Be sure to include appropriate controls. (**2 points**)

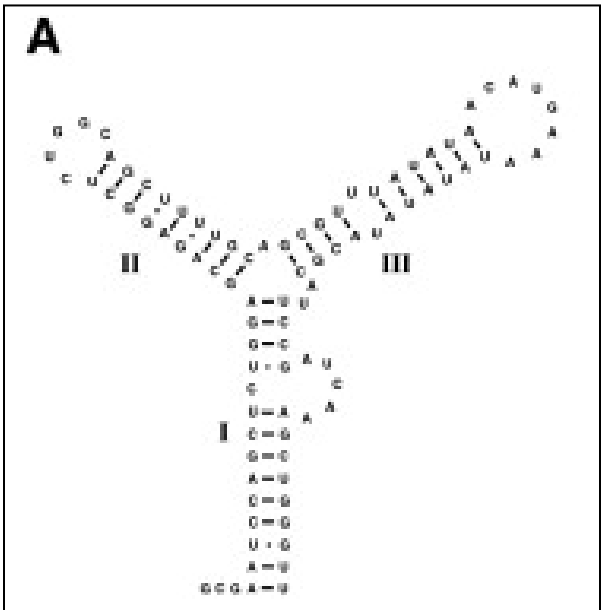


Fig. 1A: *Mwd1* mRNA predicted 3'UTR structure.

Question 3B

Frequent mutations are found in a gene called *TUEL* that can cause tumors or embryonic lethality. Recently you cloned the *TUEL* gene. Using probes generated from the genomic DNA of the *TUEL* gene, you detect two transcripts of 3.0 kb and 4.5 kb on northern blot. While your experiments demonstrate that the TUEL gene is broadly expressed in different tissues and some cultured cell lines, the relative intensity of the bands corresponding to these two transcripts varies in different tissues. Assuming you have sufficient funds to generate all necessary reagents, answer the following questions. Be sure to provide all necessary controls required for your experimental designs to obtain definitive answers.

A) You produce an antibody against the C-terminal domain of the predicted TUEL protein sequence. Immunoblotting reveals that your antibody detects two protein bands. How could you experimentally confirm that both proteins you detect are specific proteins produced by the *TUEL* gene? (**2 points**)

B) Propose a hypothesis to explain the presence of the 3.0 kb and 4.5 kb mRNA transcripts produced from the *TUEL* gene. Be sure to clearly state your hypothesis (Note that Parts C-F of this question draw on this hypothesis for your answers so you might want to read those questions before settling on a hypothesis). (**2 points**).

C) Describe an experimental approach to test your hypothesis from (**B**). For your experimental results, describe results that would SUPPORT your hypothesis **and** results that would NOT SUPPORT your hypothesis. Be sure to include appropriate controls to ensure the rigor and reproducibility of your results. (**6 points**)

D) Taking your hypothesis from (**B**) for how the two distinct *TUEL* mRNAs are produced into consideration, describe an approach to help you determine whether the two *TUEL* mRNAs are likely to encode the two distinct protein products that you detect by immunoblotting. (**2 points**)

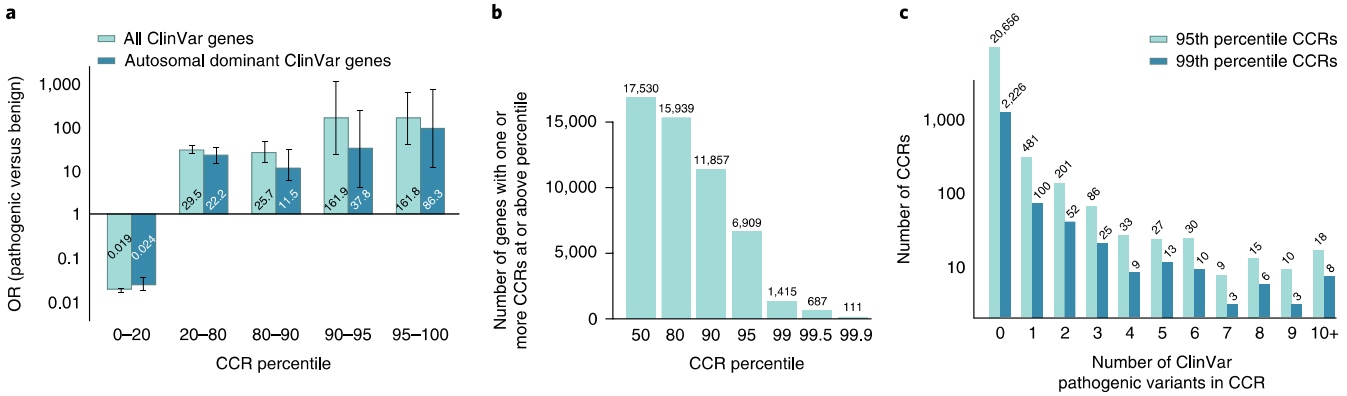
E) As mentioned above, you have found that various tissues show different relative levels of the 3.0 kb and 4.5 kb *TUEL* transcripts. Taking your hypothesis from (**B**) for how the two distinct *TUEL* mRNAs are produced into consideration, suggest a mechanistic model that could explain differences in the relative levels of these two transcripts in different tissues. Describe an experimental approach to test this model being sure to define your experimental system and include appropriate controls. (**6 points**)

F) While normal glial cells express approximately equal amounts of the 3.0 kb and 4.5 kb *TUEL* transcripts, you found that a glioma cancer cell line that expresses primarily the 4.5 kb mRNA but barely detectable levels of the 3.0 kb mRNA. You suspect that this change in expression is due mutations in the genome of the cancer cell line. Taking your hypothesis from (**B**) into consideration, suggest how a mutation (or mutations) could affect the expression of the 3.0 kb transcript with little or no effect on the 4.5 kb transcript. (**2 points**)

Question 4A

Havrilla et al. 2018 recently published a paper entitled, “A map of constrained coding regions in the human genome.” Based on your knowledge of human and population genetics, interpret the figures below. CCR is an abbreviation of a *C*onstrained *C*oding *R*egion.

Figure 2 from Havrilla et al 2018



A) Interpret the pattern of genetic variation contained in Panel a above? Does this pattern match your expectations? Explain why or why not. **(5 points)**

B) Provide a summary of the data shown in Panel c above. Explain whether this pattern matches your expectations or if it does not. **(5 points)**

C) In Panel c, the largest number of CCRs are found with zero (0) pathogenic variants. Provide an explanation for this observation. **(5 points)**

D) The data in Havrilla *et al.* 2018 are derived from assessing patterns of genetic variation within 123,136 sequenced human genomes. Suggest an experiment in the mouse that would allow you to test the function of CCRs with no pathogenic variants and provide interpretations of alternative outcomes. **(5 points)**

Question 4B

When analyzing the function of a new gene and its gene product, a hardcore Mendelian geneticist will always want a series of mutant alleles (aka an "allelic series") for complete analysis, especially in a multicellular organism. You run a developmental genetics lab, and you have collaborated with a human disease lab to identify a new gene, *geneA*, that appears to be mutated in an inherited disease in humans. The analysis of multiple families with this disease suggests an autosomal recessive pattern of inheritance.

A) What kind of mutant allele of *geneA* do you predict is present in these families? (**2 points**)

To model this new disease, you obtain a single mutant allele of the candidate homolog of *geneA* in the organism of your choice (call the allele *geneA-1*). Assume that you have all of the genetic tools available for that organism and that *geneA-1* gives an obvious phenotype.

B) Using genetic tools and your knowledge of genetics, describe one approach you would use to ascertain if *geneA-1* is a null, a hypomorph, or a hypermorph allele. Draw your crossing schemes. (**4 points**)

C) Design and describe a screen to obtain multiple new alleles of *geneA* in your organism, and describe how you would identify them and place them within an allelic series. Draw your crossing schemes and provide enough detail so that another scientist could replicate your general approach from your written notes. (**6 points**)

D) Describe how you would map *geneA* in your chosen organism using standard genetic tools - note you have not yet identified the gene. Draw your crossing schemes. (**4 points**)

E) Once you think you have identified *geneA*, describe one approach that you would use to unambiguously verify that mutations in *geneA*, and only *geneA*, are causative for the phenotype(s) you described for the original mutation. (**4 points**)

Question 5A

While exploring the rain forests of Costa Rica you discover a new species of yeast. You decide to identify all the "essential" genes in this new species. For the purposes of this question, assume an essential gene is one that is required for survival and growth of cells under permissive laboratory conditions – normal temperature, rich medium, etc.

You do not have a sequence of this yeast's genome (nor the resources to do so at least until Part C of this question), but you find that all the lab techniques that work for *S. cerevisiae* also seem to work well for this new species. Like *S. cerevisiae*, this species can exist in stable haploid and diploid states.

A) Using standard approaches you first create a set of recessive nutritional markers that confer convenient auxotrophies for tryptophan, leucine, and histidine; you name these alleles *trp1*, *leu2*, and *his 3.* Describe your approach to generate these auxotrophic mutants including your approach to confirm that you have generated the mutants that you want (**4 points**).

Next, you perform the necessary manipulations to introduce these markers into haploid cells of both mating types (a and alpha).

B) Now describe an experimental approach that would allow you to identify and clone large numbers of essential genes in your new yeast species. Assume you have a well-equipped lab and lots of colleagues to help do the work. List your proposed steps and explain the purpose of each, including any controls (**12 points**).

C) Assuming that you did have the ability to sequence the genome of your new yeast strain, how would you use that information to begin to guess which genes might be essential in this species? (**2 points**)

D) Once you had identified a candidate gene that you hypothesize is essential, what approach could you use to convincingly test whether the gene is essential? Recall that a negative result is not convincing. (**2 points**)

Question 5B

This is the abstract to a published paper describing a series of experiments in *Drosophila*:

Abstract

Mechanisms that define segmental elements of the metazoan body plan are often highly conserved, and play key roles in controlling tissue-specific patterns of gene expression. Using a collection of balanced P-element alleles distributed across the entire *Drosophila* genome, we have identified seven (7) separate alleles that cause a wing-to-leg transformation when homozygous. Genetic experiments with Deficiency lines suggest that these seven P-element alleles are loss-of-function alleles, and represent two distinct genes, which we have named *millipede* and *centipede*. Transgenic assays show that *millipede* and *centipede* are each required autonomously within primordial wing cells to prevent transformation into leg tissue. The *millipede* gene is required for *centipede* expression, and epistasis tests indicate that *millipede* acts upstream of *centipede* in maintaining wing fate. In sum, we have identified two novel factors, millipede and centipede that define a novel circuit in wing fate determination.

Describe in sufficient detail a genetic experiment that was done to support each of the following statements in the Abstract, and how the results support the statement. You are free to invoke whatever *Drosophila* reagents (alleles, antibodies, transgenes of any kind, etc.) you will need to carry out the experiment.

A) "Using a collection of balanced P-element alleles distributed across the entire *Drosophila* genome, we have identified seven (7) separate alleles that cause a wing-to-leg transformation when homozygous" (**3 points**)

B) "…experiments with Deficiency lines suggest that these seven P-element alleles are loss-of-function alleles" (**3 points**)

C) "…and represent two distinct genes" (**2 points**)

D) *"…millipede* and *centipede* are each required autonomously within primordial wing cells to prevent transformation into leg tissue" (**4 points**)

E) "The *millipede* gene is required for *centipede* expression." (**2 points**)

F) "genetic epistasis tests indicate that *millipede* acts upstream of *centipede* in maintaining wing fate" (**6 points**)