STUDENT A

Written Qualifying Exam for the Graduate Program in Genetics and Molecular Biology
January 12, 2018
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 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 Guy Benian by 6:00 PM. One of them will collect any exams from the room at 6:00 PM but if you finish sooner, please take the exams to either Guy (105-E WH) or Anita (RRC1021) in their 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 Guy 404-455-6948 (cell)/404-727-5953 (office) or Anita 404-421-9061 (cell).

GOOD LUCK!

STUDENT A

Question 1.1

The muscle-specific protein that you are studying, muscle-specific protein 1 (MSP1), is readily detectable by immunoblotting of lysates from a C2C12 mouse muscle cell line but is not detectable in lysates from a mouse fibroblast cell line. Analysis of RNA from the same cells indicates that the mRNA encoding MSP1 is also undetectable in the fibroblast cell line but is readily detected in the C2C12 muscle cell line.

A) Propose two models for the apparent cell-type specificity of MSP1. Design an experiment to distinguish between these two possibilities. Be sure to include appropriate controls for your experiment remembering that you have access to two cell lines that show different results. (**5 points**)

B) The transcription factor, TrnM, is the only known trans-activator for *MSP1* and you show by immunoblot analysis that TrnM protein is expressed at equal levels in both muscle and fibroblast cell lines. You hypothesize that TrnM is unable to bind to the sole, well-characterized, TrnM binding element (TBE) in the *MSP1* promoter in the fibroblast cell line. Design an experiment to test this hypothesis. Be sure to include appropriate controls and provide an explanation for how the results support or refute (or help you to modify!) your hypothesis. (**5 points**)

C) The TBE (TCTCGAGA) has been defined by in vitro binding assays using recombinant TrnM protein and fragments of the *MSP1* upstream sequences. Your experiments from (b) indicate that while TrnM can bind the endogenous *MSP1* promoter in muscle cells, there is no binding in the fibroblast cell line. You sequence the *TrnM* gene and promoter region, including the TrnM binding site, in the fibroblast and muscle cell lines and find no polymorphisms or sequence differences that could explain your results. Suggest two possible mechanisms that could explain the results that you have obtained (the **Figure** shows a schematic of the domain structure of the TrnM protein to consider) and design an experimental approach to test your hypothesis. (**10 points**)

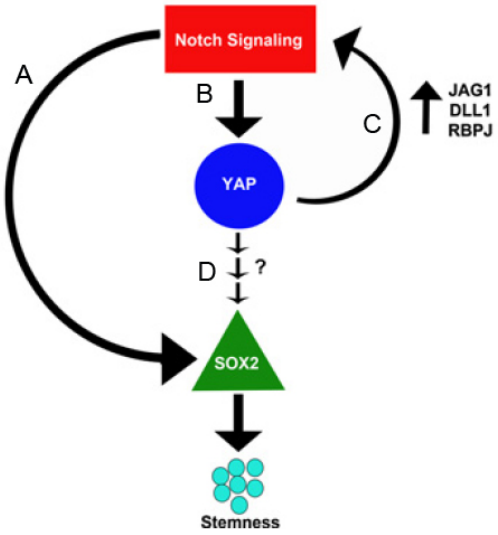
Domain Structure of TrnM Transcription Factor. As with most transcription factors, TrnM contains a nuclear localization signal (NLS) and both DNA Binding and Transactivation domains. The DNA binding domain in TrnM is a winged helix motif.

STUDENT A

Question 1.2

Below is part of an abstract from a recent paper and the model figure from the paper. YAP is a transcriptional co-activator and key effector of the Hippo signaling pathway.

Abstract: The developmental pathways of Notch and YAP have been identified as potent but independent oncogenic signals that support the embryonal variant of rhabdomyosarcoma (eRMS). Here, the cross-talk between these pathways and the impact on eRMS tumorigenesis is reported. [We] found that Notch signaling upregulates YAP activity. Reciprocally, YAP transcriptionally upregulates the Notch ligand genes JAG1 and DLL1 and the core Notch transcription factor RBPJ. This bidirectional circuit boosts expression of key stem cell genes, including SOX2, which is functionally required for [tumor] spheres.



Based on this abstract and figure, answer the following questions:

A) Describe the experiments the authors might have used to demonstrate the connection between Notch and SOX2. Be sure to include controls. **(5 points)**

B) Describe two experiments that you would perform to support the statement "Notch signaling upregulates YAP activity". **(5 points)**

C) Describe the experiments you would perform to support the statement "YAP transcriptionally upregulates the Notch ligand genes JAG1 and DLL1 and the core Notch transcription factor RBPJ". **(5 points)**

D) Describe two experiments you would perform to extend the current study and show a connection between YAP and SOX2. **(5 points)**

STUDENT A

Question 2.1

Heterochromatin in fission yeast is often nucleated at repetitive elements near centromeres and can spread to neighboring regions to occupy a discrete domain within the genome.

Constitutive heterochromatin is often nucleated at repetitive elements and can spread to neighboring regions to occupy a discrete domain within the genome. Using your favorite model organism:

A) Describe a mechanism that is thought be involved in the *initiation* of heterochromatin establishment at specific regions, and an experiment that provides support for the mechanism you describe. **(4 points)**

B) Describe a mechanism by which heterochromatin is maintained at a particular region, and detail an experiment through which you could show that maintenance is disrupted if the mechanism you choose is defective? **(6 points)**

C) Design an experiment to identify factors that promote or repress heterochromatin spreading in any model system. **(6 points)**

D) You have identified a factor that represses heterochromatin spreading. This protein contains a chromodomain. Propose a model for how this protein counteracts spreading and describe how you could test this model. **(4 points)**

STUDENT A

Question 2.2

An undergraduate student in the lab accidentally transfected a differentiated neuronal cell culture with a plasmid encoding a trypanosome-specific histone variant that was intended for use in a different project. To his surprise, this caused the neurons to de-differentiate to a stem cell-like state.

A) Propose a mechanistic model for how this, or any, histone variant might cause epigenomic reprogramming of neuronal cells. **(4 points)**

B) Propose an experiment that would directly test your model. Include sufficient detail to address feasibility and all appropriate controls. **(4 points)**

C) You find that transfection of the trypanosome variant into neurons causes a rapid displacement of the endogenous H3.3 variant from chromatin. Propose a hypothesis that explains, mechanistically, why this displacement happens and detail an experiment that directly tests your hypothesis. **(6 points)**

D) You suspect there may be a general effect on genome-wide transcription caused by the H3.3 displacement by the trypanosome histone variant. How could you determine this effect and, assuming you find support for this hypothesis, suggest a) why this displacement might be occurring, and b) why this displacement could lead to the observed phenotype. **(6 points)**

STUDENT A

Question 3.1

Your medical geneticist colleague approaches you to describe her ascertainment of two proband siblings (colored red in Figure 1) with short stature, particularly of the hands and feet, and insulin resistance. During the course of her examination, your colleague takes a family history and draws the pedigree show in **Figure 1**.

Assuming this disorder is genetic in origin, state whether the following types of inheritance are consistent with the pedigree and provide 1-2 sentences in support of each answer:

A) Recessive **(2 points)**

B) Dominant **(2 points)**

C) X-linked **(2 points)**

D) *de novo* mutation **(2 points)**

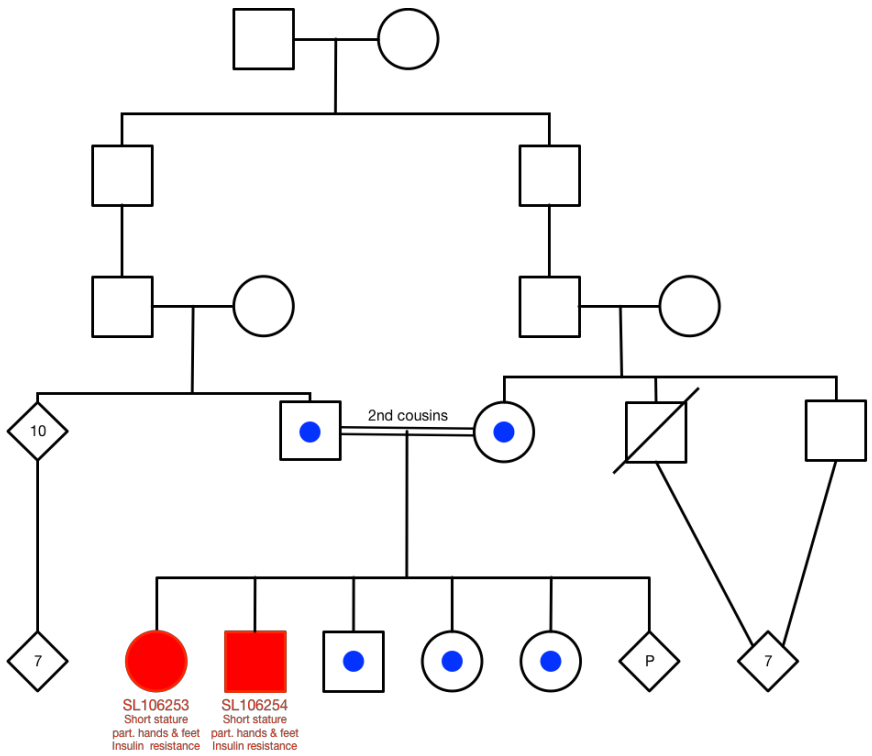


Figure 1: Pedigree of an extended family. Probands (red) are affected. Other family members (with blue dot) are not affected.

E) State the most likely mode of inheritance drawing on your answers for A-D. Provide all possible genotype configurations of the probands based on that mode of inheritance. **(6 points)**

F) You next want to understand the genetic basis of this disorder. Describe two different approaches you would use to find the mutation(s) underlying this disorder. **(6 points)**

STUDENT A

Question 3.2

You are interested in the molecular basis of intellectual disability. Through sequencing a large pedigree, you discovered a deleterious mutation within a gene encoding an RNA-binding protein (Intellectual Disability RNA Binding Protein or IDRBP) associated with the family member affected by intellectual disability. Your thesis will focus on determining whether the mutation within IDRBP could contribute to intellectual disability and understanding the function of this protein.

A) Describe an assay that you could use for either mouse or fruit fly to analyze the learning and memory deficits. If you do not know of an existing assay, you can feel free to design one on your own. Be sure to describe the controls you would include to allow you to interpret the data collected. **(3 points)**

B) Design an experiment to use a model system (mouse or fruit fly) to determine whether the mutation in *IDRBP* that you discovered does cause learning and memory deficits. Describe in detail your experimental plan, and include appropriate controls. **(4 points)**

C) You have found that the mutation of *IDRBP* can cause learning and memory deficits. Propose a hypothesis for how defects in the RNA binding protein IDRB can lead to this phenotype, and describe an experiment that can directly test your hypothesis and provide insight into the molecular mechanism involved. **(6 points)**

D) In the meantime, another group has reported that IDRBP is part of the RISC complex, which is involved in miRNA-mediated control of gene expression. In the course of your experiments, you have identified an RNA target that is regulated by IDRP. How could you begin to identify candidate miRNAs that could regulate the expression of this IDRP target RNA? **(3 points)**

E) Describe an experimental approach to test whether the candidate miRNA that you identified in Part D directly regulates expression of the IDRP target RNA. **(4 points)**

STUDENT A

Question 4.1

You've joined a lab that uses *Drosophila melanogaster* to study post-transcriptional control of gene expression in neurons. The goal of your project will be use an unbiased chemical screen to identify new recessive mutations that impair climbing in a bang test. In the bang test, you place the flies in a vial, bang the vial on the table and then watch the flies climb up the sides of the vial. Wildtype flies are negatively geotaxic and rapidly climb to the top of the vial. From the collection of climbing mutants that you identify, your PI then wants you to focus on those that appear to encode factors involved in mRNA processing, translation, or stability with the idea that these proteins could be important for learning & memory.

To begin your screen, you mutagenize wildtype adult male flies by feeding them a solution of 75 mM EMS (ethylmethane sulfonate), which induces an average of one loss-of-function point mutation (*) per chromosome per sperm. You then outcross these males in bulk to Sp/CyO;TM3/TM6B virgin females, which carry a combination of dominant markers (Sp; sternal plural), a 2nd chromosome balancer (Curly-O,CyO), and two 3rd chromosome balancers (TM3, stubble and TM6B, tubby). Your undergrad minions then backcross 10,000 individual F1 flies with balanced 2nd and 3rd chromosomes to the Sp/CyO;TM3/TM6B stock to generate 10,000 different isogenized lines, each with a unique pair of 2nd and 3rd chromosome EMS-induced mutations balanced over CyO and TM6B.

For example: */CyO;*/TM6B (*=mutation)

For all your answers, be sure to include and explain the essential controls to allow you to draw conclusions from your experimental approach.

A) Ideally, you would like to select unbalanced flies to test for defects in learning & memory, but your exhausted undergrads report to you that only 3,562 lines have unbalanced flies. What happened to the other 6,438 flies of the 10,000 total? **(3 points)**

B) Of the 3,562 lines you eventually test, 176 show defects in the bang test assay. Given that each line has two mutations (*), draw out a genetic scheme that would allow you to determine which of them causes the bang test defect. **(3 points)**

C) Now with a pool of 176 mutants that show defects in the bang test, your PI would like you to clone the genes that harbor the mutation in each of these lines. Describe an approach you would take to do this using one of the mutants as an example but keeping in mind the number that you need to map. Be sure to describe the initial steps to roughly map each gene, the readouts you would use, and the final steps you might take to identify the point mutation. **(6 points)**

Once the cloning is complete, you and your PI agree to focus on one group of defective alleles that you have identified that has known links to learning & memory. These mutants comprise a complementation group that share inactivating mutations in an RNA binding protein called RBM15 that is expressed exclusively in adult brain neurons.

D) Describe a technical approach you could use to test the hypothesis that selective loss of RBM15 *only* in neurons is sufficient to impair learning & memory. **(4 points)**

E) As a follow-up to this, propose an approach to identify neuronal RNAs that physically associate with RBM15 in adult brain neurons. **(4 points)**

STUDENT A

Question 4.2

You recently read a study from the Netherlands that reported an association between heart disease and hormonal contraceptive use. You know that heart disease has multiple SNPs that have been identified through GWAS, and you want to evaluate whether there is a gene-environment interaction between SNPs and hormonal contraceptive use that increases risk for heart disease.

A) Describe the conditions that would have to be satisfied to test this interaction? Include the rationale behind each condition and the approach you would use to test each. **(4 points)**

B) What considerations would you need to think about (controls, sample size, etc.)? **(4 points)**

C) You next test the interaction between each of 1,000,000 common SNPs across the genome and hormonal contraceptive use as a predictor of heart disease. **Figure 1** describes your results. What is the purpose of this plot? What do the results displayed in this figure suggest about your analysis? **(4 points)**

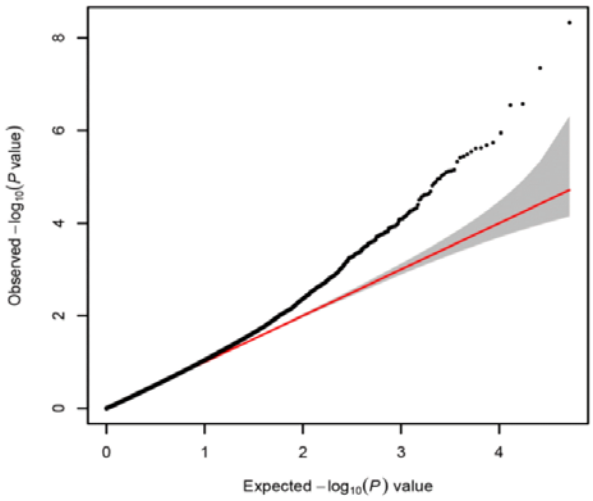


Figure 1. Quantile-quantile (QQ) plot for p-values. The shading indicates the 95% confidence

D) Figure 2 also describes your results. What is the purpose of this plot? What does it suggest about your analysis? Note that you do not need to consider any information from Part C for this question. **(4 points)**

E) Describe an approach to validate your findings. **(4 points)**

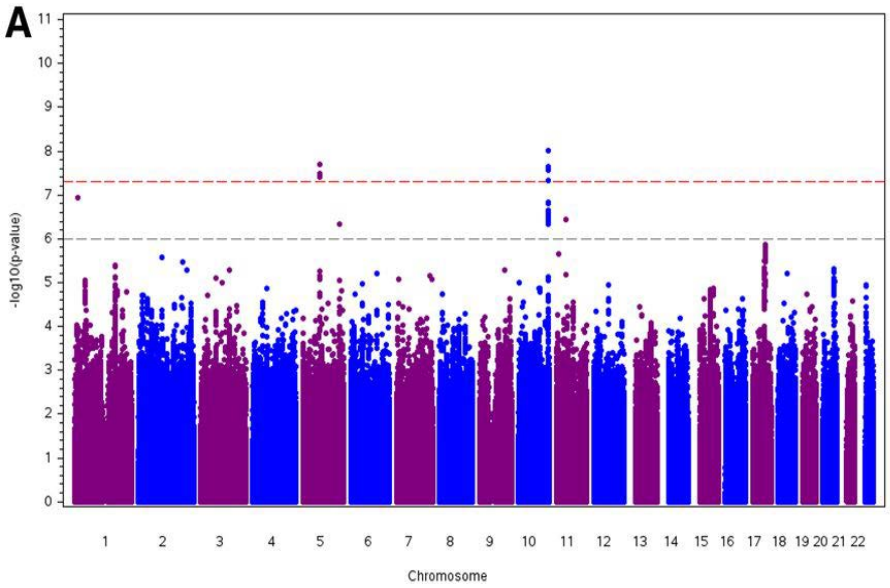


Figure 2. Manhattan plot of results. Dashed lines indicate threshold for significance after multiple test correction.

STUDENT A

Question 5.1

The discovery of the (Clustered Regularly Interspaced Short Palindromic Repeats) CRISPR/Cas9 system has revolutionized biology both by facilitating genome editing in the laboratory and by opening new possibilities for therapeutic approaches. As shown in **Figure A**, the natural *Streptococcus thermophiles* Cas9 nuclease, which was the first Cas9 characterized, functions with the aid of two RNAs (crRNA and tracrRNA). The CRISPR/Cas9 system generates double-stranded breaks in the DNA backbone by creating two independent single-stranded nicks in the DNA (as indicated by the two scissors). The resulting double-stranded break can be repaired from a template that is also provided into the system.

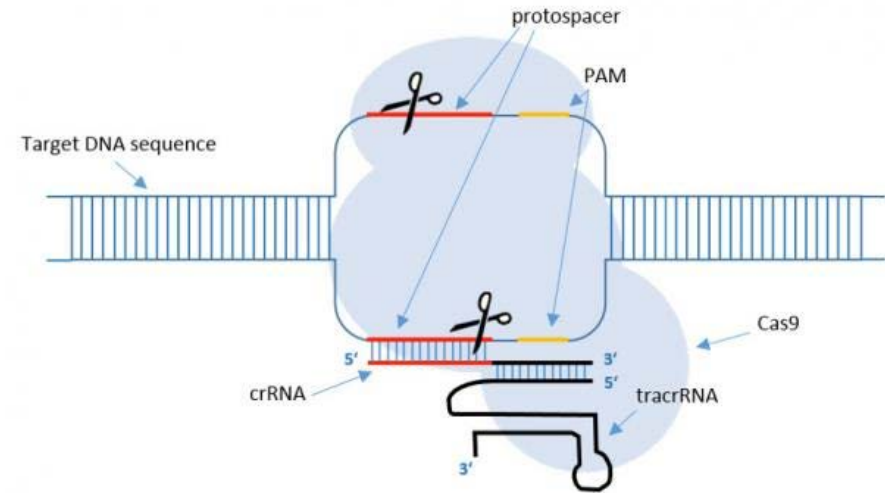


Figure A. Schematic of CRISPR/Cas9 Machinery.

The CRISPR/Cas9 system has been studied *in vitro* by reconstituting the components. Investigators can incubate a circular plasmid containing a target site with the appropriate components and monitor cleavage of the non-cut super-coiled (nSC) plasmid to once cut, nicked circular DNA (OC) or fully cleaved (with two close nicks) linear (L) DNA product.

A) The Cas9 nuclease contains two putative nuclease domains, one found in RuvC nucleases and one termed the HNH domain. For Figure D, the investigators used different active site mutants of Cas9. The (D9A) amino acid substitution occurs in the RuvC-like domain while the H599A amino acid substitution occurs in the HNH domain. What do the results of the experiments shown in **Figure D** tell you about the mechanism of cleavage by Cas9? (**3 points**)

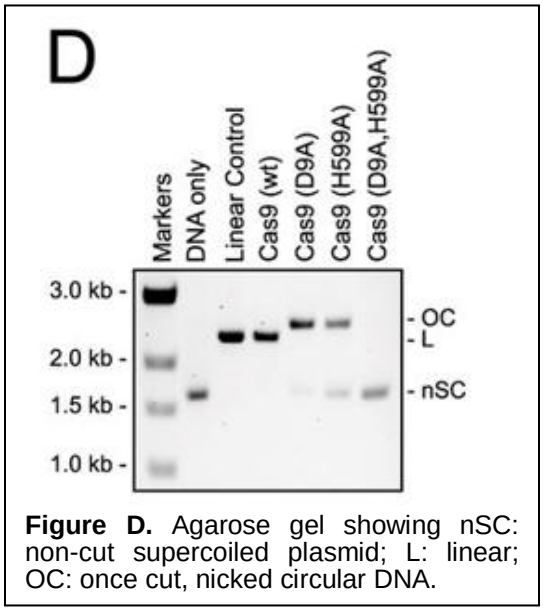
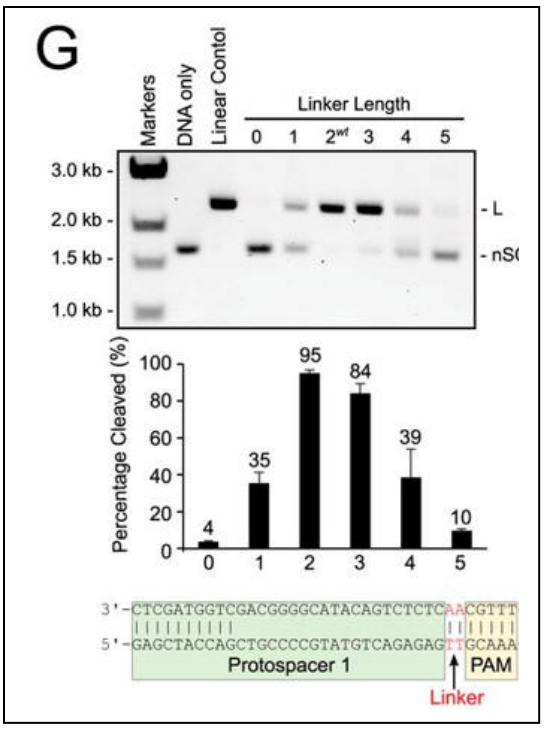


Figure D. Agarose gel showing nSC: non-cut supercoiled plasmid; L: linear; OC: once cut, nicked circular DNA.



B) As shown in **Figure A** above, Cas9-mediated cleavage occurs between the protospacer sequence (red) and a target site termed the PAM sequence (yellow). To investigate the impact of the length of the linker between these sequences, the authors took advantage of the *in vitro* cleavage assay as shown in **Figure G**. Based on the data presented, what do you conclude is the optimal spacer length and why? (**3 points**)

STUDENT A

C) How would you test whether the sequence of the spacer region matters? Describe the experimental approach and how you would interpret the data you obtain. (**4 points**)

D) Explain how the crRNA and tracrRNA sequences can be combined by molecular biologists to create a single guide RNA (sgRNA) that specifies the Cas9 target sequence (**2 points**).

E) The human genome is large, so the wild-type Cas9 protein (in complex with a single guide RNA) sometimes cuts at sites other than the intended target, which is termed an off-target effect. Explain how a singly mutated Cas9 (D9A or H599A) could be used to eliminate such off-target double-strand breaks (**3 points**).

F) Explain how the double mutant D9A/H599A Cas9 could be used to prevent the transcription of a target gene (CRISPR interference) (**3 points**).

G) Explain how the introduction of a double-strand break in DNA by Cas9 can lead to inactivating mutations in the target gene (**2 points**).

STUDENT A

Question 5.2

You are interested in a mutation that causes excessive hair growth in mice, which you have mapped as a small deletion on Chromosome 3. A single transcript is annotated to that region in the genome browser. Upon further examination of the primary sequence of that transcript, it appears that there are only a few, fairly short predicted open reading frames (ORFs). You suspect that the transcript might represent a non-coding RNA.

A) Starting with the phenotype, describe the genetic and molecular approaches you used to map the deletion. **(5 points)**

B) While the one transcript is the only one currently annotated to this region, this does not rule out the potential existence of other (unknown) transcripts. Describe an approach to determine whether there are additional transcripts arising from within this same region and/or affected by the deletion and any verification approach you would employ for any of your findings. **(5 points)**

C) Design an approach to determine whether the 'hairy' phenotype results from the loss of function of the presumptive ncRNA itself versus the loss of a small peptide encoded by the RNA. Assume you have all the necessary resources (time, $$, personnel). **(5 points)**

D) The excessive hair growth is later found to be due to the overproduction of a hormone "yeti' Given your answer to Part C above, propose a model for how the deletion of the transcript identified might be causally involved in the phenotype observed. **(5 points)**