Student H

Written Qualifying Exam for the Graduate Program in Genetics and Molecular Biology
January 8, 2016
Woodruff Health Sciences Library Computer Lab (Room B65)

Read the entire exam before starting.

Answer one question from each of the 5 groups 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 Sara) 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 do not get 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 Sara Howard by 5:30 PM (300A Dental Building). She will collect any exams from the room at 5:30PM. If you finish sooner, you may either leave your exam in the room or deliver to Sara.

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

A continental breakfast (8:00 AM) and lunch (12-1 PM) will be available in the Calhoun room in the library. This room will also be available as a break room throughout the day

If you have questions, call Anita at 404-421-9061 (cell) or Guy 404-455-6948 (cell)/404-727-5953 (office). Do NOT be shy about asking questions, we will tell you if they are something you would be expected to know.

GOOD LUCK!

Question Group 1:
Question 1.1:

You are studying the regulation of a protein (TUM1) that you believe is important for the epithelial to mesenchymal transition (EMT) that occurs in cancer. When you treat cells with TGFβ, you see that EMT occurs, which you assess by loss of epithelial markers and the appearance of mesenchymal markers. You note that during EMT, the levels of the TUM1 protein decrease; however, the steady-state level of *TUM1* mRNA does not change. Assume you have a variety of tools available to study TUM1 regulation and function.

A. Describe how you would determine if the decrease in TUM1 protein that occurs during EMT is due to changes in the rate of translation of the protein or in the rate of turnover of the protein. Describe an experimental strategy to test each of these possibilities. Be sure to include appropriate controls and describe or illustrate what the results would look like for each outcome. (6 points)

B. You next test whether TGFβ can still induce EMT in cells that are null for *dicer* and find that TUM1 protein levels do not change in these cells. Propose a mechanism for TGFβ-induced downregulation of TUM1 protein based on your findings including the result obtained with the *dicer* null cells. (2 points)

C. Building on the hypothesis that you proposed in Part B, describe a strategy to identify *cis*-acting elements within the *TUM1* gene that could confer the regulatory mechanism described here. (4 points)

D. Assuming you identify *cis*-acting sequences within the *TUM1* gene, how would you search for candidate *trans*-acting factors that could mediate this regulation? (2 points)

E. You have found a candidate genetic locus encoding a *trans*-acting factor that you suspect contributes to the regulation of TUM1 protein levels during EMT. Design an experiment to test your hypothesis that this *trans*-acting factor is required for regulation of TUM1 expression. (4 points)

F. Finally you want to test whether the regulation of the TUM1 protein levels is actually important for EMT. So far you know that TUM1 protein levels change but you do not know whether TUM1 protein (or the decrease in the protein) in important for the process of EMT. Briefly describe an experimental approach to experimentally assess the requirement for TUM1 in EMT. (2 points)

Question 1.2:

The *Arabidopsis* RD29A protein is not detectable under normal growth conditions, but the protein accumulates to high levels in response to stresses such as low temperature. You have become interested in the mechanisms that regulate RD29A expression.

A. A GFP transgene driven by the *RD29A* promoter and 5' UTR recapitulates the expression of the endogenous *RD29A* gene in response to cold (*i.e.* GFP expression is only observed after cold exposure). You have used this transgene in a mutant screen and have isolated a mutant that no longer expresses the *RD29A:GFP* transgene in response to cold. How could you determine whether the mutation also affects expression of the endogenous *RD29A* gene? (2 points)

B. You found that the mutation discovered in Part A affect the endogenous *RD29A* gene as well as the GFP transgene that you used for your screen. Based on your results so far, propose a hypothesis to explain what aspect of gene expression cold most likely regulates to induce expression of the *RD29A* gene. Provide a rationale for your answer. (4 points)

C. Describe an experimental approach to test your hypothesis for how cold regulates the *RD29A* gene. Be sure to include appropriate controls and describe what results you would obtain if your hypothesis is correct as well as if your hypothesis is not correct (6 points)

D. The mutation that causes repression of *RD29A* (from part A) is found in a gene encoding a protein that contains a domain with significant sequence similarity to DNA base excision repair glycosylases. In other proteins this domain is known to specifically remove the nitrogenous base from cytosine residues, leading to replacement with a new cytosine nucleotide. Formulate a hypothesis for how loss of such an enzymatic activity could result in gene silencing. (3 points)

E. Describe an experimental approach to test the hypothesis that you proposed in Part D. (5 points).

Question Group 2:

Question 2.1:

The Drosophila C3PO protein is a sequence-specific DNA binding transcription factor and the homolog of a human protein that is lost in a form of inherited autism in which patients lack emotions and appear 'robotic'. Two key features of C3PO are undetermined: (1) whether it activates or represses transcription of its target genes, and (2) the identity of its transcriptional targets in neurons. To uncover mechanisms by which C3PO affects neuronal function, you carry out a genetic screen for mutant alleles that modify a Drosophila phenotype produced by overexpression of C3PO in all neurons (genotype: pan-neuron>Gal4,UAS-C3PO) – these flies are hyperactive, show exaggerated social behaviors with other flies, and have shortened lifespan due to a failure to sleep. This screen yielded three dominant enhancers (e1, e2, e3) and three dominant suppressors (s1, s2, s3) of the pan-neuron>Gal4,UAS-C3PO phenotype. Assume that each of these alleles is a null loss-of-function (lof) that is organismal lethal when homozygous.

A. If C3PO protein activates the transcription of its targets, are the dominant enhancers (e1, e2, e3) or the suppressors (s1, s2, s3) that you identified more likely to be direct transcriptional targets of C3PO? Provide a justification for your answer that also explains the dominant nature of the effects. (2 points)

B. Provide one experimental approach that would test your hypothesis that the C3PO protein directly regulates expression of the genes you selected above in Part A. Assume that have successfully identified the genes affected by the e1-3 and s1-3 alleles. (4 points)

Your additional studies find that the *R2D2* mRNA is increased 5-fold in neurons that overexpress C3PO protein (neuron>Gal4,UAS-C3PO). You have a fly stock in the lab that carries a homozygous lethal lof allele of the *R2D2* gene in the lab called *R2D2 short-circuit*; this allele contains a C-to-T transition that converts a CGA codon (Arginine) to a TGA codon (stop codon).

C. Propose one molecular and one genetic test of whether any of the three enhancers (e1-3) or suppressors (s1-3) are alleles of *R2D2*? Explain each proposed approach, and how you plan to interpret the data. (3 points)

D. Control neurons extend robust axonal and dendritic arbors, but neurons that overexpress C3PO (neuron>Gal4,UAS-C3PO) lack axons and dendrites and look 'bald'. One suppressor, s1, that you name Chewie dominantly reverts this 'bald' phenotype and restores neurons to their 'hairy' appearance with normal axons and dendrites. Do you predict that the normal function of Chewie is to promote or repress axon/dendrite growth? (2 points)

E. How would you directly test your hypothesis for Chewie function using standard tools and reagents available to you in *Drosophila*? (4 points)

F. The s2 allele is your strongest dominant suppressor recovered. Upon cloning the gene, you find it encodes a nuclear localized protein named Solo that localizes to chromatin. Critically, Solo mRNA levels are unaltered in neuron>Gal4,UAS-C3PO neurons. Provide a hypothesis to explain why Solo is required for C3PO overexpression phenotypes and propose an experimental approach to test this hypothesis. (5 points)

Question Group 2:
Question 2.2:

The following Figure is from a recent paper in *Molecular Cell* in which the authors used START-seq, a method that allows for the capture of the 5'-cap and 5'-end sequence, to precisely map the transcription start sites (TSSs) of nascent transcripts in mouse macrophages. Chip-seq for RNA Polymerase II (Pol II) and TATA Binding Protein (TBP), and sequencing of micrococcal nuclease digested chromatin (MNase-seq) were also performed in parallel on the same cells. The investigators observe prevalent bidirectional transcription from these promoters of murine macrophages.

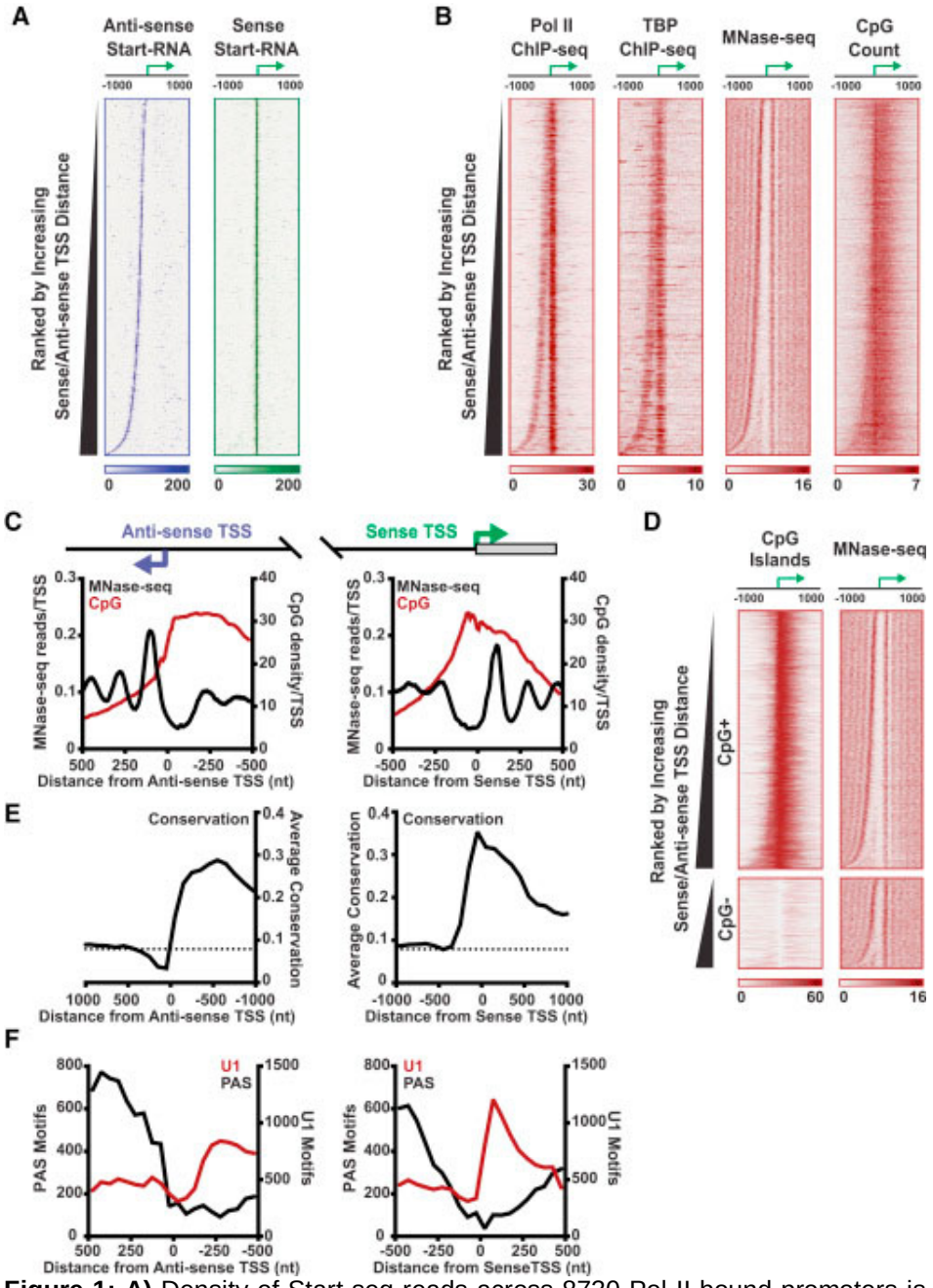


Figure 1: A) Density of Start-seq reads across 8730 Pol II bound promoters is shown. TSS = transcription start site. **B)** Density of Pol II, TBP ChIP-seq reads and of MNase-seq reads across the same promoters as in A. **C)** Average MNase-seq read density (black) and CpG density (red) across the same 8730 promoters shown in A. **D)** CpG density (left) and MNase-seq read density (right) across CpG island (CpG+) and non-CpG island (CpG-) promoters. **E)** Conservation score across mammals. Dashed line is average score across the mouse genome. F) Average number of PAS (poly-A sequences) in black, and U1 motifs in red in 50-bp bins.

Using the data from the figure shown, answer the following questions.

A. Based on panels B –D, what can you conclude regarding the relationship between nucleosome positioning, sense and anti-sense transcription, and CpG density at promoters? (4 points)
B. What can you infer from panel E regarding selective pressure to conserve sequences in and around promoters? (4 points)
C. What is the significance of poly-A sequences (PAS) and U1 sites with regard to transcriptional regulation? (2 points)
D. Provide a hypothesis to explain the functional relevance of the PAS and U1 data in panel F. (4 points)
E. How would you experimentally test the hypothesis that you proposed in Part C? (6

points)

Question Group 3:

Question 3.1:

A. Name two people who have won the Nobel Prize for different bodies of scientific work (i.e., you cannot name two people who both got the prize for the same contributions). (1 point each name)

B. For each recipient, briefly describe the work that led to this prestigious award and explain how this work changed modern science and/or medicine. (5 points each)

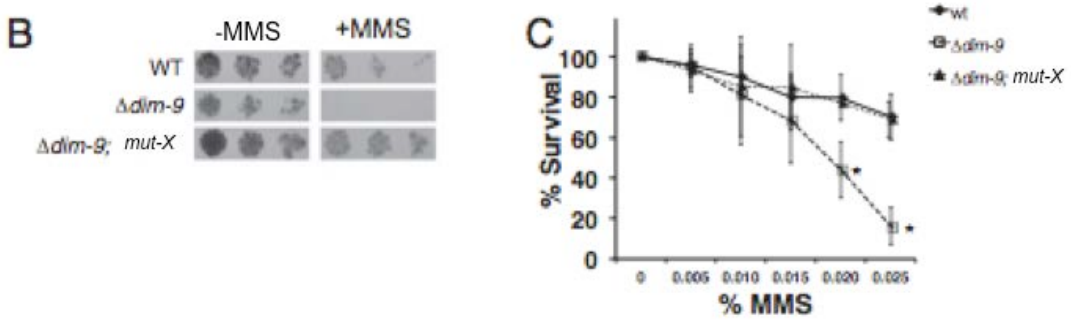
C. Name one recent scientific discovery that you believe is worthy of the Nobel Prize. Describe why you think this discovery is worthy of this recognition. In your description, note how the discovery could impact both scientific research and also consider any potential clinical or applied implications. (8 points)

Question 3.2:

The following questions pertain to the recent publication by Zack Lewis and Bob Schmitz, both of whom participated in IBS515.

Basenko, E., Sasaki, T., Ji, L., Prybol, C.J., Burckhardt, R.M., Schmitz, R.J. and Z.A. Lewis (2015) Genome-wide redistribution of H3K27me3 is linked to genotoxic stress and defective growth. *PNAS* E6339-36348.

In Neurospora, the DCDC Complex initiates heterochromatin formation by trimethylating H3K9 (DIM5 is the catalytic subunit) to create binding sites for multiple heterochromatin protein 1 (HP1)-containing complexes, which in turn direct methylation of cytosine bases in DNA and deacetylation of histones in heterochromatic regions. Strains with mutations in the genes that encode components of the DCDC complex are hypersensitive to genotoxic stress and fail to complete sexual development. This paper isolated mut-X by screening for mutants that are better able to survive genotoxic stresses in a DCDC-deficient (*Δdim-9*) background.

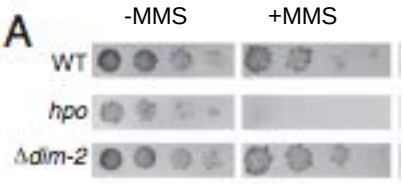


This newly identified mut-X is an allele that renders the PRC2 complex defective. PRC2 (Polycomb Repressive Complex-2) methylates H3K27 and directs transcriptional repression of PRC2-target domains in Neurospora and other fungi.

A. From data in the accompanying figure (shown on the next page), what can we infer about H3K9me and H3K27me localization in wildtype and *Δdim5* (DCDC-deficient) strains? (4 pts)

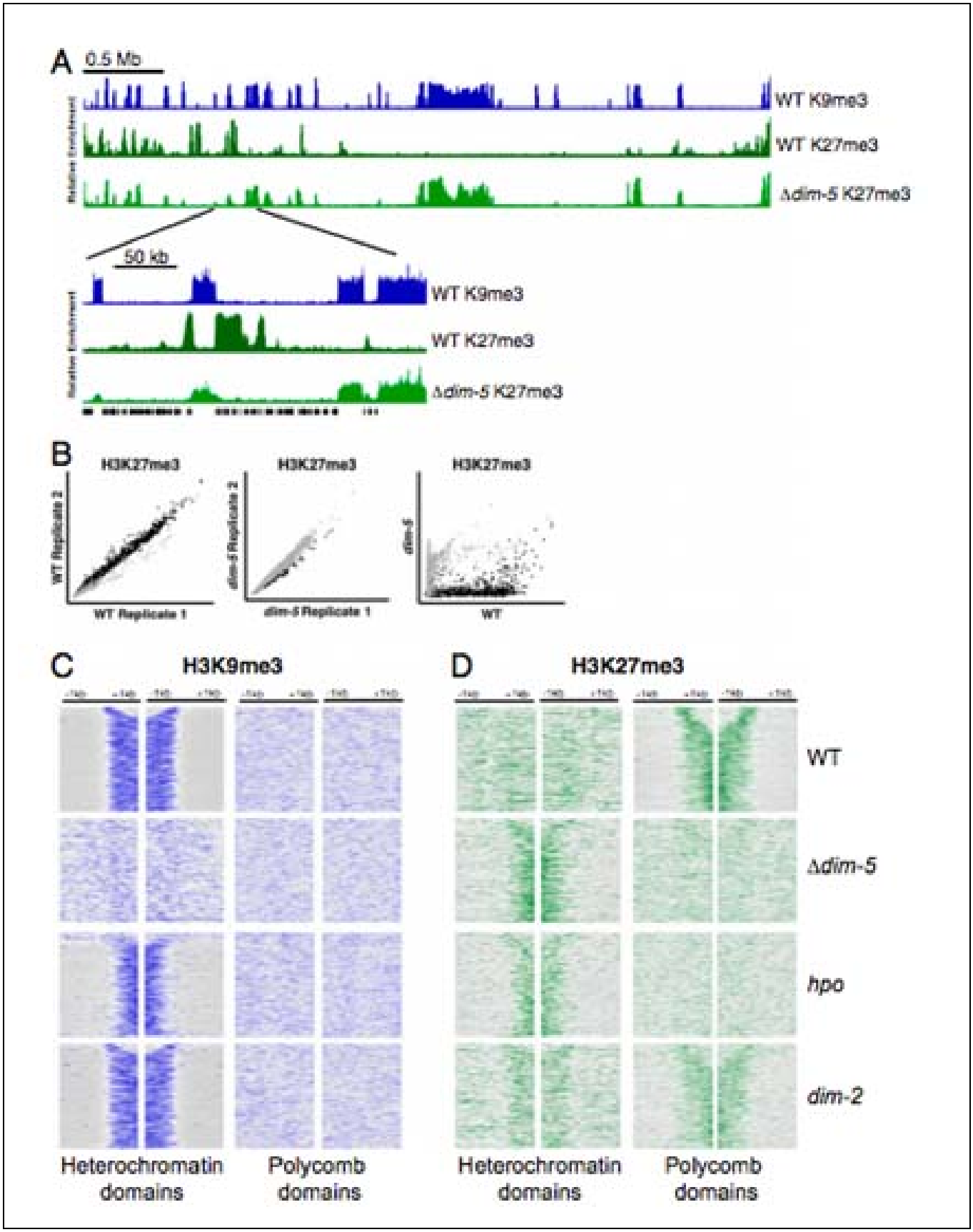
B. What is one hypothesis that could explain the relationship between H3K27me and genotoxic stress and how could you experimentally test this hypothesis? (8 pts)

C. How do genotoxic sensitivity (Panel A below) and the lysine methylation patterns (Figure on next page) in the *hpo* (lacks HP1) and *Δdim-2* (lacks DNA methylation) mutants inform your hypothesis from B? (5 pts)



D. Given the limited information you have, suggest one interesting follow-up question that these data would lead you to explore and explain why this is the next logical or the most important follow-up question? (3 pts)

Question 3.2 (continued)



Question Group 4:
Question 4.1:

PART I. Sandra's father, Jim, who is 82, was recently diagnosed with Turnage's Disease, a very rare (1/100,000) autosomal dominant adult onset disease. Sandra's mother is deceased. In addition to concern for her dad, Sandra is also worried about the risk that she or her siblings or children might have inherited the disease, which has age-dependent penetrance. The specific gene and causal mutations responsible for Turnage's Disease are not yet known, but the genomic region was recently identified by a linkage study.

A. Sandra is the oldest of four children; she has two younger brothers and one younger sister. Sandra is married to Hank, and they have two young daughters. Draw a pedigree illustrating this extended family. (3 points)

Empiric data suggest that Turnage's Disease has age-dependent penetrance such that by age thirty 10% of people carrying the disease gene will show symptoms, by age sixty 50% will show symptoms, and by age eighty 90% will show symptoms.

B. Sandra is currently 37 and not showing symptoms of Turnage's Disease. From the family history presented, estimate the risk Sandra will show symptoms of Turnage's Disease by the time she is 60? Show your work. (3 points)

C. From the family history presented, what is the risk Sandra's older daughter will show symptoms of Turnage's Disease by the time she is 30? Show your work. (3 points)

D. Jim has genetic testing to define his genotype at a marker within the genomic region linked with Turnage's Disease in another pedigree; his genotype at this marker is determined to be (1,3). Sandra's genotype at the same locus is determined to be (2, 3). Does this information alter your estimate of Sandra's risk of having inherited a Turnage's Disease mutation from her father? Explain. (2 points)

PART II. Now you have turned your attention to a different project where you seek to understand how genetic burden might influence outcomes of children with congenital heart defects (CHD). A colleague at the Children's Health Care of Atlanta has a biorepository containing DNA samples from 4000 children who underwent surgery for CHD and their resulting outcome information. The outcome information includes a large collection of patient information such as: survival of the surgery and educational attainment in school. While you do not have matched cases, you do have access to the Exome Aggregation Consortium collection of 60,000 exomes of adults (who did not have a congenital heart defect at birth), and can match ancestry with your cases in the biorepository.

E. If you were interested in determining if *common variants* influence measured outcomes, what type of experiment would you propose? Justify the advantages and disadvantages of such an experiment. (3 points)

F. If you were interested in assessing the impact of both rare and common variants, what is the most efficient experiment you could perform? Provide rationale for your choice (3 points)

G. Draw a graph illustrating the relationship between allele frequency and effect size for a disease phenotype. (3 points)

Question 4.2:

Heart failure is a disabling, costly human condition. Although the injured adult mammalian heart has a very poor regenerative capacity, an example of successful heart regeneration exist in vertebrates. Zebrafish (adults) are particularly well-studied for their ability to fully regenerate cardiac muscle (cardiomyocytes) after surgical resection of about 20% of the ventricle. Thus, you have decided to molecularly study heart regeneration in zebrafish with the goal of understanding why this ability may have been lost in mammals. (Hint: Wound healing after surgery takes a few days, regeneration is in progress 1 week after surgery and complete 1 month after surgery.)

A. To begin this project, you want to identify genes in adult zebrafish that are activated (or repressed) during heart regeneration. Describe ONE approach to identify gene expression changes during regeneration, and describe TWO approaches to validate/confirm these expression changes. (4 points)

In Part A you identified 10 genes that are of particular interest to you because their expression is significantly changed in cardiomyocytes at 7 days post-surgery, but expression returns to normal level by 14 days. Five of these genes are strongly upregulated (up) and five are strongly downregulated (down).

B. Describe and justify one experimental approach you could use to test whether the <u>up</u> genes actually play a role in zebrafish heart regeneration. (3 points)

C. Describe and justify one experimental approach you could use to test whether the <u>down</u> genes actually play a role in zebrafish heart regeneration. (3 points)

D. Two of the 5 up genes you examined above lead to lethality in zebrafish embryos when homozygous for loss-of-function alleles because they play an important role in heart development. Describe TWO approaches you could take to get around this problem and determine their function during regeneration. (4 points)

You have finally identified 4 genes (3 up, 1 down) that appear to be crucial for the ability of zebrafish cardiomyocytes to promote heart regeneration. You now want to proceed and test whether any of these 4 genes display any characteristics in a mouse model that may explain the difference in the regenerative capacity of mammals versus zebrafish. (You can assume that you have a mouse model for Myocardial Infarction (MI) and cardiomyocyte cell cultures).

E. Describe an initial molecular approach and expectations to determine if any of these 4 candidate genes are worth further investment in mammalian regeneration research. (2 points)

F. You identify 1 candidate (an up gene) that meets your criteria from Part E. What approach could you take to most efficiently assess if this candidate gene could help restore the regenerative capacity in mammalian cardiomyocytes? (4 points)

Question Group 5:

Question 5.1

Although the tumor suppressor gene (TSG) protein that you are studying is present in normal cells and readily detectable by immunoblotting, TSG protein is not detectable by immunoblotting in any of the six breast cancer cell lines that you test. The mRNA transcript for TSG also cannot be detected in the cancer cell lines. Your sequence the *TSG* locus and fund no deletions or mutations. In fact the entire chromosome where *TSG* is located shows no changes. Your goal is to understand how TSG is regulated with a focus on how expression is lost in the breast cancer cell lines. Assume all reagents that you require are readily available and that (unlike in real life) they all work well.

A. Design your experimental approach to determine if decreased TSG expression in the tumor cell lines is due to transcriptional or post-transcriptional (i.e., mRNA stability) regulation. (4 points)

B. Your experiment in A provides evidence that TSG expression is regulated at the level of transcription. Factor Y is the only known trans-activator for TSG. You show by immunoblotting that there is plenty of Factor Y present in the tumor cells. You hypothesize that Factor Y is unable to bind to the sole, well-characterized *TSG* promoter in the tumor cells. Design an experimental approach to test your hypothesis that Factor Y does not bind to the TSG promoter in the cancer cell lines. Describe the approach in detail including the results that you would obtain if your result is correct or if it is not correct. Be sure to include appropriate controls (6 points)

You show in experiment B that Factor Y does NOT bind the TSG promoter in your cancer cell lines. You sequence the Factor Y locus in several of the cancer cell lines.

C. In one cancer cell line, you sequence the TSG gene and find that within the well-characterized DNA binding pocket a highly conserved lysine residue is changed to glutamic acid. Suggest a hypothesis for how this amino acid substitution could interfere with Factor Y binding to DNA. (2 points)

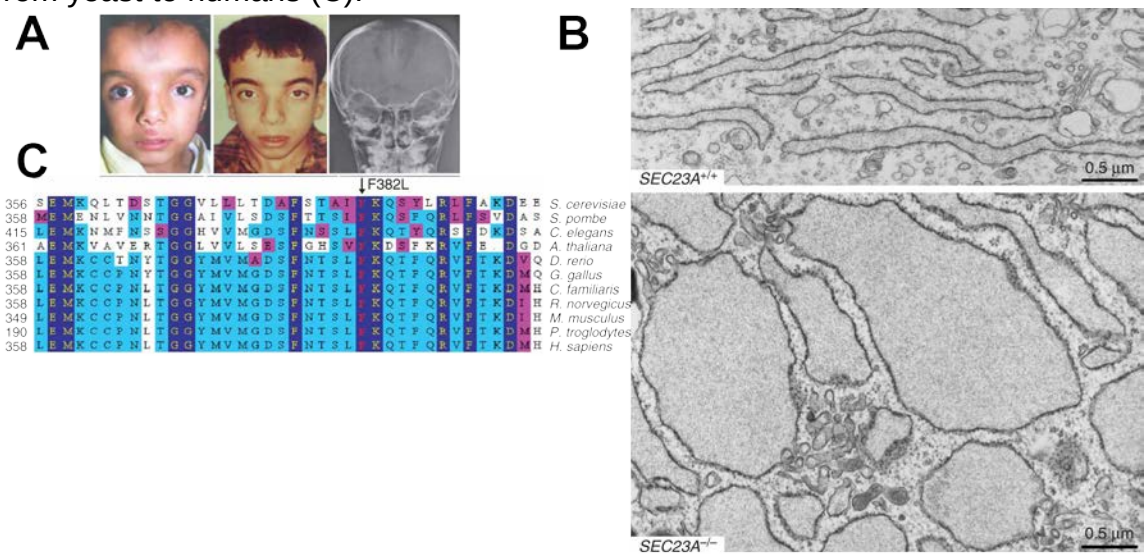
D. Design an experimental approach to test your hypothesis in Part C and describe the results that would support your hypothesis. (3 points)

E. In the remaining cell lines, you sequence the *Factor Y* gene and find no mutations. Someone suggests that you test whether Trichostatin A, a histone deacetylase (HDAC) inhibitor, affects TSG expression in these cancer cell lines, and you observe a significant increase in TSG expression upon treating cells with the Trichostatin A. Suggest a hypothesis for why Trichostatin A treatment could lead to an increase in *TSG* transcription. (2 points)

F. Design an experimental approach to test your explanation for how Trichostatin A could increase TSG expression and describe the results that would support this hypothesis. (3 points)

Question 5.2:

Cranio-lenticulo-sutural dysplasia (CLSD) is a **non-lethal** autosomal recessive syndrome characterized by late-closing fontanels, cataracts, facial dysmorphisms and skeletal defects (A). Electron micrographs from patient fibroblasts reveal an extensive enlargement of endoplasmic reticulum cisternea (B). The genetic defect corresponds to an amino acid substitution (F382L) in sec23A, a subunit of the COPII complex. Sequence alignment reveals that this residue is conserved from yeast to humans (C).



The COPII cytosolic complex is critical in the formation of vesicles that deliver secretory proteins from the endoplasmic reticulum to the Golgi complex.

Saccharomyces cerevisiae mutants lacking *SEC23* (Δ*sec23*) are not viable. Similarly, yeast Sec23 protein carrying a Phenylalanine 382 to Leucine (F382L) amino acid substitution fails to rescue the lethal phenotype of the *sec23* deletion. However, a Phenylalanine 382 to Tyrosine (F382Y) amino acid substitution confers a temperature sensitive phenotype when expressed as the only copy of Sec23 in yeast cells.

Based on this information:

A. Propose two distinct hypotheses that could explain why patients with the F382L amino acid substitution in sec23A are viable whereas the corresponding amino acid substitution in Sec23 is lethal in yeast. (4 points)

B. Suggest an explanation for why the F382L amino acid change yields a non-functional protein in yeast whereas the F382Y amino acid change yields a protein that still retains some function. (3 points)

C. Choose one of your hypotheses (from Part A) and propose an experimental strategy (with controls) to test the hypothesis. Feel free to use yeast and/or human cells in your strategy if needed. Be sure to describe what results you would expect to obtain if your hypothesis is correct and what results you would obtain if your hypothesis is not correct (8 points)

D. How would you use the F382Y temperature sensitive yeast mutant to better understand the genetic pathway where *SEC23* participates? Describe an experimental approach you could employ (5 points)