7.03 Problem Set 1

Due before 5 PM on Wednesday, September 19

Hand in answers in recitation section or in the box outside of 68-120

1. You have isolated a collection of yeast mutants that form small colonies when plated on a concentration of the drug geldanamycin that does not affect the growth of wild type cells. Geldanamycin sensitive mutants **1**–**5** are *MAT**a*** and mutants **6**–**10** are *MAT*α. Your analysis begins by pairwise mating of each mutant to a wild-type strain and to the mutants of the opposite mating type. The size of the colonies of the resulting diploids are shown in the table below.

MATα Strains

	Wild-type	Mutant **6**	Mutant **7**	Mutant **8**	Mutant **9**	Mutant **10**
Wild-type	Large	Small	Large	Large	Large	Large
Mutant **1**	Large	Small	Small	Small	Large	Large
Mutant **2**	Large	Small	Large	Large	Small	Small
Mutant **3**	Large	Small	Large	Large	Large	Large
Mutant **4**	Large	Small	Large	Large	Large	Large
Mutant **5**	Large	Small	Small	Small	Large	Large

⇧

*MAT**a*** Strains

a) Which of the mutants are dominant and which are recessive?

b) Organize the 10 mutants into complementation groups (genes), indicating any ambiguities.

c) Based on these complementation data, what is the minimum number of genes that represented by this collection of geldanamycin sensitive mutants? What is the maximum number of genes?

Starting with mutant **2** (*MAT**a***) and mutant **9** (*MAT*α) you isolate a set of revertants that can form large colonies on geldanamycin. Revertants designated Revertant **1** – Revertant **5** were isolated from mutant **2** and are therefore *MAT**a***. Revertants designated Revertant **6** – Revertant **10** were isolated from mutant **9** and are therefore *MAT*α. You carry out pairwise mating of each revertant to one of the original geldanamycin mutants and to the revertants of the opposite mating type. The size of the colonies of the resulting diploids are shown in the table below.

MATα Strains

	Mutant **9**	Revertant **6**	Revertant **7**	Revertant **8**	Revertant **9**	Revertant **10**
Mutant **2**	Small	Small	Large	Small	Small	Small
Revertant **1**	Small	Small	Large	Small	Small	Large
Revertant **2**	Small	Small	Large	Large	Large	Small
Revertant **3**	Small	Small	Large	Large	Large	Small
Revertant **4**	Large	Large	Large	Large	Large	Large
Revertant **5**	Small	Small	Large	Large	Large	Small

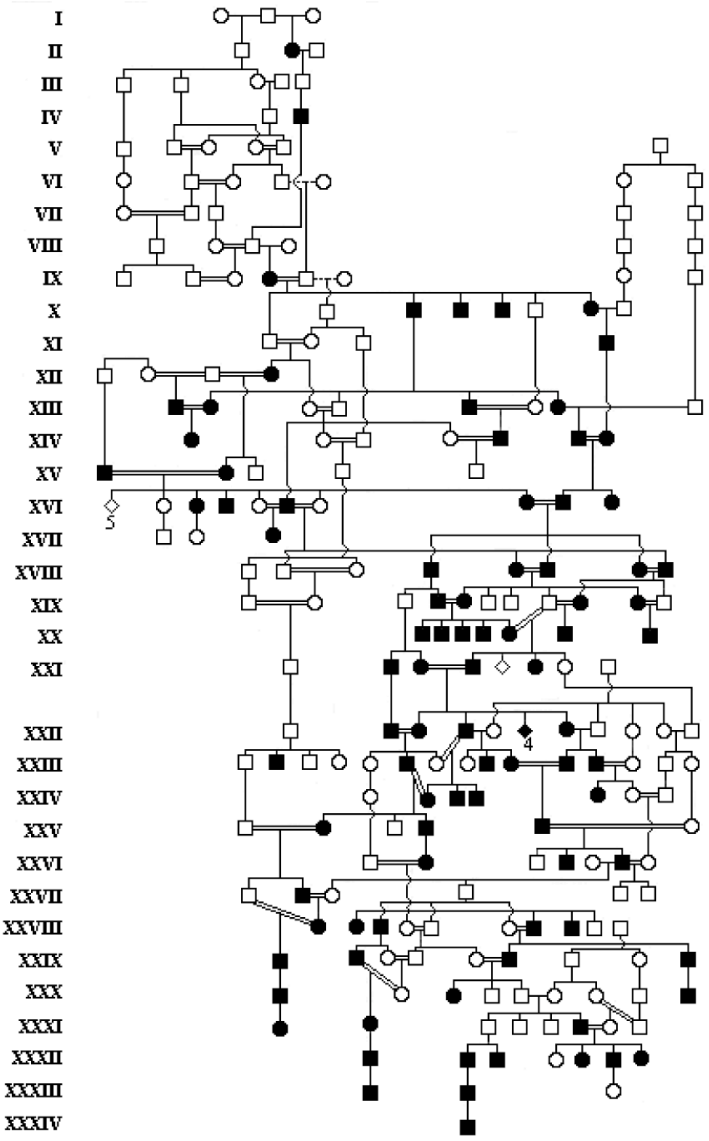
⇧

*MAT**a*** Strains

d) Which of the revertants carry recessive suppressor mutations that correct the original geldanamycin sensitivity? Which of the revertants appear to contain dominant suppressor mutations correct the original geldanamycin sensitivity? For this analysis why didn't we mate the revertants to wild type strains?

e) Based on these complementation data, what is the minimum number of genes that can suppress the geldanamycin sensitivity of mutant **2** or mutant **9**. What is the maximum number of genes?

2. Reproduced below is the pedigree of the Hapsburg royal family showing the inheritance of the trait "Hapsburg lip" (the modern medical term for this trait is *mandibular prognathism*), indicated by filled symbols. What is the most likely mode of inheritance of this trait? Justify your answer taking into account that this is a rare trait in the population at large, and that there is imperfect historical data on the characteristics of all of the individuals in the pedigree. Your answer should explain why the trait does not appear in every generation. Do you think the high frequency of consanguineous matings in this family has contributed to the incidence of the trait? Why or why not?



3. **a)** You have isolated a new mouse mutant that has silver colored fur. When this mutant is crossed to wild type about half of the offspring have silver fur suggesting that the mutant is dominant. You then cross several of the silver offspring to one another to produce 12 F2 mice, among which 8 are silver and 4 appear normal. Calculate the chi square value for this segregation with the null hypothesis that silver fur is determined by a dominant allele of a single gene. Use the table of *p* values below to determine whether this data is consistent with the expected segregation pattern using the convention that for $p < 0.05$ there is a statistically significant difference between the observed results and the results expected for a given model.

b) You are reasonably confident that you understand the inheritance of the silver fur mutation until a lab mate tells you that some coat color mutants in mammals are dominant for the color trait, but are lethal when they are homozygous (i.e. they are recessive lethal mutations). Why doesn't the data that you already have eliminate the possibility that the silver color mutation is a recessive lethal? Assuming that your silver mutation is actually not a recessive lethal, estimate how many F2 progeny you would need to analyze to statistically eliminate the possibility that the silver mutation is a recessive lethal. [For your estimate assume that exactly ¾ of the F2 progeny are silver.]

p value:	.995	.975	0.9	0.5	0.1	0.05	0.025	0.01	0.005
df = 1	.000	.000	.016	.46	2.7	3.8	5.0	6.6	7.9
df = 2	.01	.05	.21	1.4	4.6	6.0	7.4	9.2	10.6
df = 3	.07	.22	.58	2.4	6.3	7.8	9.3	11.3	12.8

4. Hemophilia A is a debilitating defect in blood clotting caused by recessive alleles of the gene for clotting factor VIII. The gene for factor VIII lies on the X chromosome and therefore hemophilia is much more frequent in males than in females. The frequency of males with hemophilia A is about 10^{-4} (the frequency of females with the disease is negligible). The frequency of new hemophilia A mutations is 3.3×10^{-5}.

a) Say you identify a male with hemophilia A, without knowing anything about his family, what is the a priori probability that he inherited the disease from his mother?

b) Now say you learn that a maternal uncle of the affected male also has hemophilia A. Using Bayes' theorem calculate the new conditional probability that the affected male inherited the disease from his mother given that one of her brothers has the disease. [Hint: you may find it useful to set this problem up by using Bayes' theorem to calculate the probability that the maternal grandmother of the affected male is a carrier.]

7.03 Problem Set 3

Due before 5 PM on Thursday, October 18

Hand in answers in recitation section or in the box outside of 68-120

1. Bacteria that live in an environment exposed to UV rays from sunlight often have less than 50% A•T base pairs in their genomic DNA. This is thought to result from the tendency of UV light to react with neighboring thymine residues to make a mutagenic chemical modification of DNA known as a thymine dimer. Thus bacteria that are exposed to UV avoid mutations by minimizing the occurrences of the target sequence, 5' TT 3'.

a) Would you expect a bacterial species with 35% A•T necessarily to have proteins of unusual amino acid sequence. Explain.

b) In class, we saw that the probability of a stop codon in random sequence with 50% A•T content is 3/64. Calculate the probability of a stop codon in a random DNA sequence of 35% A•T content. Calculate the probability of a stop codon in a random DNA sequence of 65% A•T content.

c) We also saw that the probability of an ORF (open reading frame) of 100 codons in random sequence with 50% A•T content is $(61/64)^{100}$. Use the values that you calculated in part **b** to calculate the probability of an ORF of 100 codons in DNA with 35% A•T content or 65% A•T content.

d) Given these calculations, describe the relationship between A•T content of the DNA of a bacterial species and the ease with which true genes can be identified in genomic sequences from that species.

2. Imagine that you have discovered two new mutagens: Compound X which only make G•C to A•T mutations and Compound Y which only makes A•T to G•C mutations (real mutagens are never this specific). You are interested in the kinds of nonsense mutations and nonsense suppressors that these mutagens can generate.

a) Catalog all of the sense codons that can be mutated to nonsense by a single base change generated by Compound X. Do the same for Compound Y.

b) Catalog all of the amber suppressor mutation that can be produced by a single base change generated by Compound X. For your answer, write out the double stranded DNA segment of the anticodon segment of the affected tRNA. Also, do the same for Compound Y.

3. You are interested in synthesis of the amino acid arginine by *E. coli.* You start by making some transposon insertion mutants that cannot grow without arginine in the medium (Arg^-).

a) You isolate three Arg^- mutants (called Arg-1^-, Arg-2^- and Arg-3^-) made by insertions of the transposon Tn5, which carries resistance to kanamycin (Kan^r). Is there a way to use P1 transduction to ascertain whether any of these insertions are in the same gene? Explain why or why not.

b) Next, you isolate three more Arg^- mutants (called Arg-4^-, Arg-5^- and Arg-6^-) by insertions of the transposon Tn10, which carries resistance to tetracycline (Tet^r). You grow P1 phage on each of the Tn5 insertion mutants and use the resulting lysates to transduce each of the mutants generated by insertion of Tn10 by selecting for Kan^r. The table below shows the percent of Kan^r transductants that are also Tet^r.

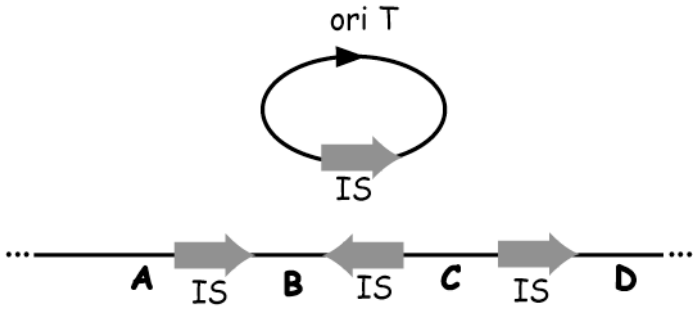
		Donor strain (Kan^r)		
		Arg-1$^-$	Arg-2$^-$	Arg-3$^-$
Recipient strain (Tet^r)	Arg-4$^-$	100%	50%	100%
	Arg-5$^-$	10%	100%	30%
	Arg-6$^-$	100%	20%	100%

Draw as detailed a map as you can showing the relationships between the different transposon insertions. Give map distances in cotransduction frequencies and indicate any ambiguities.

c) What is the minimum and maximum number of Arg genes represented by these transposon insertion mutations?

d) Next, you discover that a gene (Gal2) required for growth on galactose is located near to the Tn5 insertion in Arg-1. You have two point mutations in the Gal2 gene, called *Gal2-1*$^-$and *Gal2-2*$^-$. To map these mutations relative to the Tn5 insertion you set up two reciprocal crosses. In the first cross you grow P1 on a strain that carries the Tn5 insertion and the *Gal2-1*$^-$ mutation. You then use this lysate to infect a *Gal2-2*$^-$mutant and select for Kan^r. From 100 Kan^r transductants examined, 95 are Gal$^-$ and 5 are Gal$^+$. In the second cross you grow P1 on a strain that carries the Tn5 insertion and the *Gal2-2*$^-$ mutation. You then use this lysate to infect a *Gal2-1*$^-$ mutant, and select for Kan^r. From 100 Kan^r transductants examined, all are Gal$^-$. Draw a genetic map showing the relative positions of the Tn5 insertion and the *Gal2-1*$^-$ and *Gal2-2*$^-$ mutations. Express any distances that as cotransduction frequencies (points will be deducted for giving distances that can't be obtained from this data).

4. The diagram below shows the F factor and a portion of the *E. coli* chromosome that has three different insertion sequences (IS) of the same type as is carried on F.



a) Starting with the F$^+$ strain diagramed you isolate an Hfr strains that can transfer the D marker early, but transfers C very late. Show the recombination event that would produce this Hfr and the final structure of the Hfr, showing all of the markers as well as the position and orientation of each IS sequence and the origin of transfer (ori T).

b) Using the Hfr described in part **a**, you wish to isolate an F' factor that carries the C gene. Show the recombination event that would produce the desired F' and the final structure of the F'. Describe in general terms how you could select for the desired F'.

7.03 Problem Set 4

Due before 5 PM on Thursday, November 1

Hand in answers in recitation section or in the box outside of 68-120

1. You have isolated a set of five different mutations in the Lac operon designated Lac-1 through Lac-5. For each mutant you apply a set of genetic tests that include i) testing the phenotype of the mutant itself, ii) the phenotype of the mutant in a merodiploid carrying an F' with the wild type Lac operon (F'Lac^+), iii) the phenotype of the mutant in a merodiploid carrying an F' with the Lac operon with a LacZ– mutant (F'$LacZ^-$), and iv) the phenotype of a merodiploid in a wild type strain carrying an F' factor with the mutant and a $LacZ^-$ mutant (F' Lac-X, $LacZ^-$). For each set of experimental results describe, as specifically as you can, the type of each Lac mutant. Also indicate whether that type of mutant would occur relatively frequently or relatively rarely and explain why.

a)
Lac-1 : constitutive
Lac-1/F'Lac^+ : constitutive
Lac-1/F' $LacZ^-$: constitutive
Lac^+/ F' Lac-1 $LacZ^-$: constitutive

b)
Lac-2 : constitutive
Lac-2/F'Lac^+ : constitutive
Lac-2/F' $LacZ^-$: constitutive
Lac^+/ F' Lac-2 $LacZ^-$: inducible

c)
Lac-3 : constitutive
Lac-3/F'Lac^+ : inducible
Lac-3/F' $LacZ^-$: inducible
Lac^+/ F' Lac-3 $LacZ^-$: inducible

d)
Lac-4 : uninducible
Lac-4/F'Lac^+ : inducible
Lac-4/F' $LacZ^-$: uninducible
Lac^+/ F' Lac-4 $LacZ^-$: inducible

e)

Lac-5 :	uninducible
Lac-5/F'Lac$^+$:	uninducible
Lac-5/F' LacZ$^-$:	uninducible
Lac$^+$/ F' Lac-5 LacZ$^-$:	uninducible

2. You are studying a new strain of *E. coli* that can grow on the sugar raffinose. Utilization depends on the enzyme raffinase, which is encoded by the gene Raf1. The enzyme raffinase is not expressed unless glucose is absent from the medium and raffinose is present in the growth medium.

	raffinase activity		
	– raffinose–glucose	+ raffinose +glucose	+ raffinose –glucose
wild type (Raf1$^+$)	–	–	+
Raf1$^-$	–	–	–

To dissect the regulatory circuit for Raf1 expression, you isolate a set of regulatory mutants by transposon mutagenesis.

a) Mutant Raf2 was isolated as a Tn5 insertion and gives constitutive expression of raffinase. P1 transduction experiments of the Tn5 insertion that causes the Raf2 mutation show that Raf2(Tn5) is not linked to Raf1.

	raffinase activity		
	– raffinose–glucose	+ raffinose +glucose	+ raffinose –glucose
Raf2(Tn5)	+	–	+

Based on this information propose the type of regulatory functions affected by the Raf2 mutation. Draw a diagram showing your proposed regulatory pathway(s) showing the normal function of the Raf2 gene product, the inducer raffinose, and the anti-inducer glucose.

b) Mutant Raf3 was isolated as a Tn10 insertion and gives uninducible expression of raffinase. P1 transduction experiments of the Tn10 insertion that causes the Raf3 mutation show that Raf3(Tn10) is not linked either to Raf1 or to Raf2(Tn5). The table below shows the phenotype of Raf3(Tn10) and a Raf2(Tn5) Raf3(Tn10) double mutant.

	raffinase activity		
	– raffinose–glucose	+ raffinose +glucose	+ raffinose –glucose
Raf3(Tn10)	–	–	–
Raf2(Tn5) Raf3(Tn10)	+	–	+

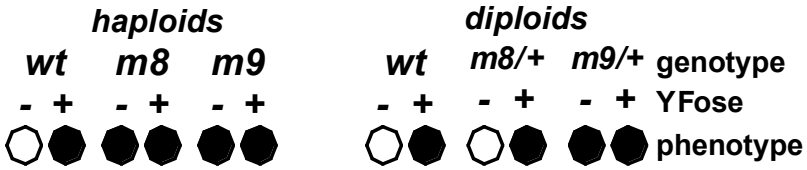
Based on this information propose the type of regulatory functions affected by the Raf3 mutation. Draw a diagram showing your proposed regulatory pathway(s) showing the normal function of both the Raf2 and Raf3 gene products, the inducer raffinose, and the anti-inducer glucose.

c) Mutant Raf4 was also isolated as a Tn10 insertion and gives uninducible expression of raffinase. P1 transduction experiments of the Tn10 insertion that causes the Raf4 mutation show that Raf4(Tn10) is not linked to Raf1 , to Raf2(Tn5), or to Raf3(Tn10). The table below shows the phenotype of Raf4(Tn10) and a Raf2(Tn5) Raf4(Tn10) double mutant.

	raffinase activity		
	– raffinose–glucose	+ raffinose +glucose	+ raffinose –glucose
Raf4(Tn10)	–	–	–
Raf2(Tn5) Raf4(Tn10)	–	–	–

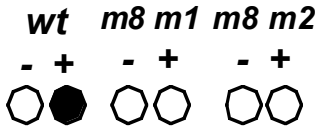
Based on this information propose the type of regulatory functions affected by the Raf4 mutation. Draw a diagram showing your proposed regulatory pathway(s) showing the normal function of both the Raf2 Raf3, and Raf4 gene products, the inducer raffinose, and the anti-inducer glucose. (There are at least two models that will fit the data).

3. You are interested in the regulation of YFose utilization in *Saccharomyces cerevisiae.* Wild type (wt) yeast induce expression of the YFase enzyme in the presence of YFose. You have an assay for YFase expression (YFase expression is depicted as dark colonies and lack of YFase depicted as white). After mutagenesis, you find two mutant strains (with mutations m8 and m9) that display the following characteristics as haploids and as heterozygotes in diploids from a cross with the wild type:



a) For each mutation, specify whether it confers a constitutive or uninducible YFase phenotype and whether it is dominant or recessive.

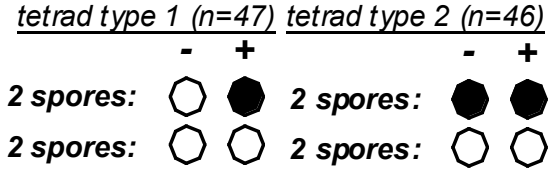
YFase-expressing yeast die in the presence of compound X. The YFase assay is expensive and you realize you can use compound X, your mutations, and mutagenesis to identify more components of YFase regulation. You mutagenize *m8* haploids, plate them on compound X in the absence of YFose and select for survivors. (Denote the gene defined by *m8* as *yfg8*.) You isolate two surviving strains, and test them for YFase regulation (you designate these strains as containing the mutations *m1* and *m2)*:



b) Which of the following statements about the mutation m1 or m2 cannot be excluded by existing data (indicate all that apply):
(i) The mutation confers an uninducible phenotype and is dominant.
(ii) The mutation is recessive phenotype and is in a gene that directly negatively regulates *yfg8.*
(iii) The mutation is recessive and is in a gene that is directly negatively regulated by *yfg8.*
(iv) The mutation is a back mutation of the original *yfg8* mutation.
(v) The mutation is linked to *yfg8.*
(vi) The mutation is unlinked to *yfg8*

Next, you cross your two new strains (note that these strains still possess the original *m8* mutation) to wild type; resultant diploids display normal YFase regulation. You then sporulate and observe only two types of tetrads. Below is shown sample data from a cross with *m8 m1* haploids and wild type, but similar results are obtained with *m8 m2* haploids crossed with wild type.

m8 m1 X wt; sporulate, test spore phenotype with replica plating to plates + YFose



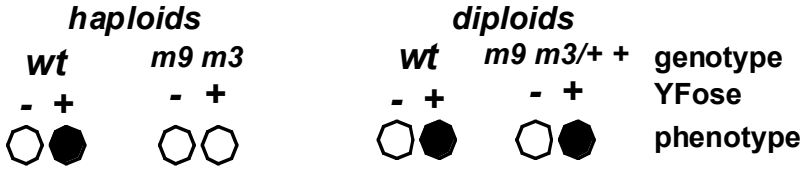
example of what data shows: in tetrad type 1 all 4 spores were YFase(-) in the absence of YFose, and 2 spores were YFase (+) in presence of YFose

c) Which of the possible statements (i through vi) can still apply?

You find that the *m1* mutation is a mutation in the *yfg1* gene, which encodes YFase itself. *m2* is not a mutation in *yfg1* and therefore defines a new gene you call *yfg2*.

d) Based upon these and prior observations, draw a simple, linear regulatory pathway including YFose, *yfg1, yfg2*, and *yfg8*.

Encouraged by your results, you aim to understand the *m9* mutation. You mutagenize the *m9* strain and find a new mutant (designated *m3*) that shows the following phenotype:



e) Which of the following statements are true (state all that apply):
(i) *m3* is recessive and confers an uninducible phenotype

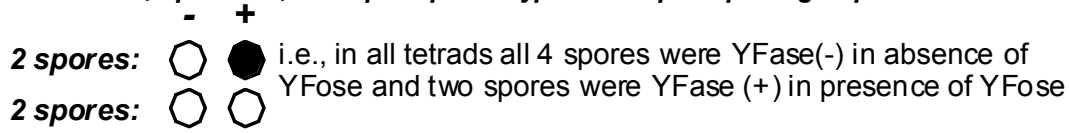
(ii) *m3* is recessive confers a constitutively active phenotype
(iii) *m3* is a dominant suppressor of the phenotype (YFose off) caused by *m9*.
(iv) *m3* is a recessive suppressor of the phenotype (YFose off) caused by *m9*.

f) Which of the following models are possible explanations for *m3* based upon existing data (indicate all that apply):

(i) *m3* is a loss-of-function mutation in the gene carrying the *m9* mutation (e.g., *m9* and *m3* are both mutations in the *yfg-2* gene and cause opposite affects on *yfg-2* function).
(ii) *m3* is a mutation in a different gene than that carrying the *m9* mutation, but the two genes are linked.
(iii) *m3* is a mutation in a different gene than that carrying the *m9* mutation, and the two genes are unlinked.

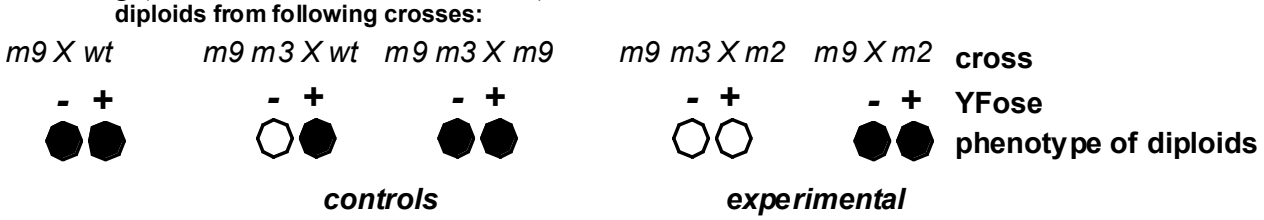
Following a cross between your strain *(m9 m3)* and wild type, you sporulate diploids and observe that all tetrads have the following distribution of spores:

m9 m3 X wt; sporulate, test spore phenotype with replica plating to plates + YFose



g) Which of the statements from **f)** are shown to be false by this new data?

Finally, you cross *m9 m3* haploids with an an *m2* singly mutant haploid strain, and observe the following (control crosses are also shown):



h) Which of the statements from **f)** remains as the most consistent with shown data?

7.03 Problem Set 5

Due before 5 PM on Wednesday, November 21

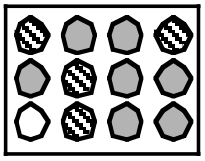
Hand in answers in recitation section or in the box outside of 68-120

1. You are studying the regulation of gene expression in yeast induced by signalX using microarrays. Shown are hybridization data from a panel of 12 genes; DNA for each gene is fixed on the microarray in known positions (labeled with numbers). For each experiment, RNA is isolated from two conditions and labeled with a fluorophore that fluoresces either red or green. Then a competitive hybridization is performed with the added RNA and the DNA fixed on the array; for example if there is more RNA for gene1 in an RNA pool labeled red than in the RNA pool labeled green then, following hybridization, the spot will glow red.

a) You isolate RNA from cells in the absence of signalX and label the RNA red. You isolate RNA from cells in the presence of signalX and label RNA green. Which genes are upregulated by factor X, which genes are down-regulated, and which genes are not affected transcriptionally by factor X.

+signalX; red

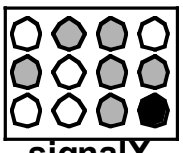
-signalX; green



b) Propose an explanation for the above data for array gene #9.

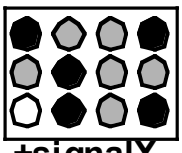
c) You study a deletion mutant of *sgx-1* (*sgx1-)* that completely lacks the *sgx-1* gene, by performing microarray analyses in the absence of factor X and in the presence of factor X. In both cases, competitive hybridizations are done with RNA from *sgx1-* cells (labeled red) and from wt cells (labeled green).

***sgx1-*; red**
wt; green



-signalX

***sgx1-*; red**
wt; green



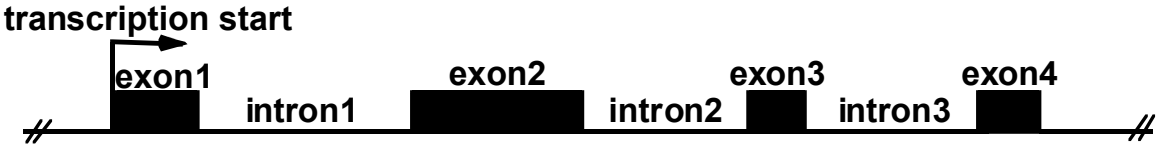
+signalX

(i) It is known that one of the spots on the microarray is DNA from the *SGX1* gene. Which spot?

(ii) Why are spots 1, 4, 6, and 10 white in the absence of signalX?

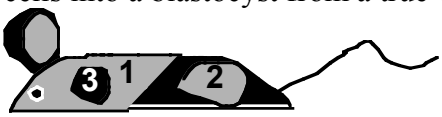
(iii) Does the *sgx1-* mutation cause an uninducible or a constitutively active phenotype for the genes regulated by factor X?

2. You are studying a gene associated with human cancer, blastoma1 (B1), by creating a mouse knock out lacking the gene. Loss-of-function mutations in human B1 cause cancer. You use a homologous recombination approach in embryonic stem cells derived from a true-breeding, black-coated mouse strain. Note: the black coat phenotype is dominant and the gray coat color phenotype is recessive.



a) Your aim is to delete exon 2 entirely, while leaving exons 1, 3, and 4 intact. Diagram a homologous recombination strategy that would accomplish this. Depict the features of your targeting construct that allow selection for integration of your construct and selection for homologous rather than non-homologous integration.

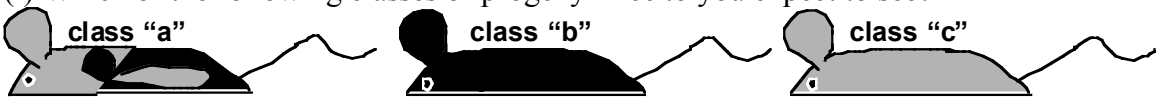
b) After confirming your approach in part (A) worked by PCR, you inject your clonal, recombinant ES cells into a blastocyst from a true-breeding, gray-coated mother and obtain a mouse "M" shown below.



(i) Which, if any, of regions 1, 2, and 3 are derived from your ES cells?
(ii) Assuming the mouse B1 gene acts as does its human ortholog, do you anticipate that this mouse will have cancer?

c) You cross mouse "M" to a gray-coat mouse.

(i) Which of the following classes of progeny mice to you expect to see?



(ii) If you observe some class "b" mice, what must have been true about the nature of the chimerism in mouse "M"?

(iii) Do you anticipate that any of these mice will have cancer?

d) You do obtain some class "b" mice and cross them to one another. Do you expect any of the progeny to have cancer? If so, what percentage of the mice?

3. a) A hypothetical island population has a high frequency of loss-of–function alleles for a gene that cause dominant resistance to a tropical parasite (heterozygous advantage = .05). However, the same allele causes a recessive circulatory system disease (homozygotes have fitness = 0.2). Calculate the expected frequency for the allele in the island population that should have been established over many generations of selection for heterozygotes and against homozygotes.

b) The neighboring mainland has a large population that has not been exposed to the parasite, however new mutations to produce new loss-of–function alleles for the gene occur at a rate $\mu = 10^{-6}$ and homozygotes for the allele have fitness = 0.2 in the mainland population as well. Calculate the expected frequency for the allele in the mainland population that should have been established over many generations of new mutations counterbalanced by selection against homozygotes.

c) Now imagine that the island becomes part of the same country as the mainland such that 10% of the total population comes from the original island population. What will the frequency of the circulatory disease be for the entire population of the country, assuming there is *no* genetic mixing of people from the island with the population at large?

d) Assuming that there continues to be very little genetic mixing of the island population with the rest of the country. Calculate the frequency of the circulatory disease (for the entire population) after 10 generations, assuming that parasitic disease has been eradicated (i.e. there is no longer a heterozygote advantage) but that the individuals that are homozygous for the loss-of-function allele still have fitness = 0.2 because of the circulatory disease.

e) Now recalculate the frequency of the circulatory disease for the entire population assuming that there is complete mixing of the populations (i.e. mating between individuals is completely random with respect to whether an individual came from the island or mainland population).