7.03 Problem Set 1

Due before 5 PM on Wednesday, September 20

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

1. Imagine a natural compound called Spindlestop, which shows promising anti-tumor activity. The primary source of Spindlestop is the bark of an endangered species of tree, but suppose that the yeast *S. cerevisiae* is found to produce Spindlestop in minute, yet detectable, quantities. To increase production, you isolate 30 yeast mutants with increased levels of Spindlestop. Mutants 1-15 are mating type a (MAT a) and mutants 16-30 are mating type α (MAT α).

The analysis begins by pairwise mating of each mutant to a wild-type strain and to the mutants of the opposite mating type. The amounts of Spindlestop produced by the resulting diploids are shown in the table below ("wt" indicates wild-type quantities, "+" indicates about 10X wild-type levels).

MAT α Strains

	Wild-type	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Wild-type	wt	+	wt	wt	wt	wt	+	wt	+	+	wt	wt	wt	wt	wt	+
1	wt	+	wt	wt	wt	wt	+	wt	+	+	wt	wt	wt	+	wt	+
2	wt	+	wt	+	wt	wt	+	+	+	+	+	wt	wt	wt	+	+
3	wt	+	wt	+	wt	wt	+	+	+	+	+	wt	wt	+	+	+
4	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+
5	wt	+	wt	+	wt	wt	+	+	+	+	+	wt	wt	wt	+	+
6	wt	+	wt	+	wt	wt	+	+	+	+	+	wt	wt	wt	+	+
7	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
9	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+
10	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+
11	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+
12	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+
13	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
14	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+
15	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+

⇧

MAT a Strains

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

b) Based on the observed properties what can you conclude about the function of the gene affected by Mutant 1 with respect to Spindlestop production? What can you conclude about the gene altered by Mutant 7?

c) What is anomalous about the behavior of Mutant 3? Provide a simple genetic explanation.

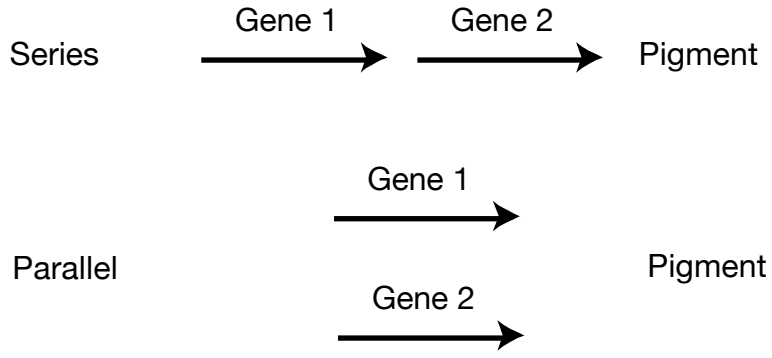
d) Organize the 30 mutations into complementation groups (genes). Please indicate any remaining ambiguities.

e) Based on these limited complementation data, what is the absolute minimum number of genes that must comprise the Spindlestop biosynthetic pathway? What is the maximum number of genes?

f) In an attempt to construct a yeast strain that produces even more Spindlestop than any of the existing mutants you decide to combine two of the recessive mutants to make a double mutant strain. Pick two of the mutants to combine and explain the reasoning for your choice.

2. Wild type mice are gray and in a large-scale breeding colony two white female mice arise from different parents. You would like to know whether the white phenotype is caused by two mutations in the same gene or in different genes. Using the concepts of dominance, recessivity, and complementation, describe a set of crosses and the interpretation of their outcomes that you would use to make this determination. Please be sure to indicate the circumstances that would prevent you from easily making this determination. (Assume you have available an unlimited number of true breeding wild-type gray mice.)

3. Genes that control coat color in mice can be thought of as steps in biochemical pathways whose products are pigmented compounds that give the fur its color. Albino mice have white fur because they lack the ability to make any pigment. Imagine mutations in two different genes that can, in certain combinations, block the production of pigment, yielding mutants with white fur. There are two different possible arrangements for two biochemical steps responsible for the formation of pigment: the two genes might act in *series* such that a loss of function of *either* gene would block the formation of pigment, or the two genes could act in *parallel* such that loss of function of *both* genes would be required to block the formation of pigment.



a) Say that you are given an albino mouse. When you cross this mouse to wild type all of the F1 progeny appear normal (i.e. like wild type). Consider the following three possibilities for the genetic basis of the albino trait: 1) recessive allele of a single gene, 2) recessive alleles of two genes acting in series 3) recessive alleles of two genes acting in parallel. For each of the three possibilities give the proportion of albino and normal looking mice among the F2 generation.

b) You cross the normal looking F1 mice among themselves producing 40 F2 mice; 15 are albino and 25 appear normal. Determine whether this data is consistent with each of the three possibilities outlined in part (a) and draw whatever conclusions you can about the inheritance of albinism. The table below gives chi square values for 1, 2 and 3 degrees of freedom. Use 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 and therefore we can reject the model on the basis of the experimental data.

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. PKU is an autosomal recessive genetic disorder resulting from a loss of the enzyme phenylalanine hydroxylase, which converts phenylalanine to tyrosine. Without this enzyme, phenylalanine and its breakdown products accumulate to toxic levels resulting in mental retardation. Fortunately, individuals homozygous for the disease allele can be spared by a phenylalanine-free diet.

Say that Sarah has a brother with phenylketonuria (PKU), but she doesn't have the disease herself.

a) What is the probability that Sarah is a carrier for PKU?

Sarah's husband was detected to have PKU at birth but he has been treated with a phenylalanine-free diet. Sarah and her husband recently had their first child and are relieved that their child shows normal phenylalanine levels and thus is not homozygous for the disease allele.

b) Based on this new information use Bayes Theorem to compute an updated probability that Sarah is a carrier for PKU.

7.03 Problem Set 1 Answer Key

1. Imagine a natural compound called Spindlestop, which shows promising anti-tumor activity. The primary source of Spindlestop is the bark of an endangered species of tree, but suppose that the yeast *S. cerevisiae* is found to produce Spindlestop in minute, yet detectable, quantities. To increase production, you isolate 30 yeast mutants with increased levels of Spindlestop. Mutants 1-15 are mating type a (MAT a) and mutants 16-30 are mating type α (MAT α).

The analysis begins by pairwise mating of each mutant to a wild-type strain and to the mutants of the opposite mating type. The amounts of Spindlestop produced by the resulting diploids are shown in the table below ("wt" indicates wild-type quantities, "+" indicates about 10X wild-type levels).

MAT α Strains

	Wild-type	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Wild-type	wt	+	wt	wt	wt	wt	+	wt	+	+	wt	wt	wt	wt	wt	+
1	wt	+	wt	wt	wt	wt	+	wt	+	+	wt	wt	wt	+	wt	+
2	wt	+	wt	+	wt	wt	+	+	+	+	+	wt	wt	wt	+	+
3	wt	+	wt	+	wt	wt	+	+	+	+	+	wt	wt	+	+	+
4	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+
5	wt	+	wt	+	wt	wt	+	+	+	+	+	wt	wt	wt	+	+
6	wt	+	wt	+	wt	wt	+	+	+	+	+	wt	wt	wt	+	+
7	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
8	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
9	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+
10	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+
11	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+
12	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+
13	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
14	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+
15	wt	+	+	wt	+	+	+	wt	+	+	wt	+	+	wt	wt	+

⇧

MAT a Strains

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

The phenotype of the diploid resulting from a mutant to wild type test cross indicates whether a particular mutation is recessive or dominant. A wild type diploid phenotype indicates that the mutation is recessive, while a mutant diploid phenotype suggests that the mutation is dominant.

Recessive: 1-6, 9-12, 14-15, 17-20, 22, 25-29
Dominant: 7-8, 13, 16, 21, 23-24, 30

b) Based on the observed properties what can you conclude about the function of the gene affected by Mutant 1 with respect to Spindlestop production? What can you conclude about the gene altered by Mutant 7?

Mutant 1 possesses a recessive mutation. Because recessive mutations tend to be loss of function mutations, we suspect that the gene affected by mutant 1 serves to suppress spindlestop biosynthesis. A recessive, loss-of-function mutation in this gene would increase spindlestop biosynthesis.

Conversely, mutant 7 possesses a dominant mutation. Because dominant mutations tend to be gain-of-function mutations, we suspect that the gene affected by mutant 7 serves to promote spindlestop biosynthesis.

c) What is anomalous about the behavior of Mutant 3? Provide a simple genetic explanation.

Mutant 3 resides in two different complementation groups. Given that a complementation group comprises a set of mutations in the same gene, mutant 3 must be a double mutant; it possesses mutations in two different genes in the spindlestop biosynthetic pathway.

d) Organize the 30 mutations into complementation groups (genes). Please indicate any remaining ambiguities.

A complementation group is a set of mutations in the same gene. Recessive mutants that fail to complement possess mutations in the same gene and thus belong to the same complementation group.

I. 1, 28, (3)
II. 2, 5, 6, 18, 22, 25, 29, (3)
III. 4, 9-12, 14-15, 17, 19-20, 26, 27
All of the dominant mutations (7-8, 13, 16, 21, 23-24, 30) are ambiguous. The complementation test can only be used to study recessive mutations.

e) Based on these limited complementation data, what is the absolute minimum number of genes that must comprise the Spindlestop biosynthetic pathway? What is the maximum number of genes?

Minimum: 3 Maximum: 11

The absolute minimum number of genes in the spindlestop biosynthetic pathway is equal to the number of identified complementation groups, each of which represents a particular gene in the pathway. The maximum number of genes is equal to the number of complementation groups plus the number of ambiguous mutations. In this example, there are 8 ambiguous mutations corresponding to the number of dominant mutations that we cannot identify with complementation analysis.

f) In an attempt to construct a yeast strain that produces even more Spindlestop than any of the existing mutants you decide to combine two of the recessive mutants to make a double mutant strain. Pick two of the mutants to combine and explain the reasoning for your choice.

Select any two mutants in different complementation groups; this will enable you to hit two different genes in the spindlestop pathway. Because the two mutations in mutant 3 (complementation groups I and II) show no additive effect on spindlestop biosynthesis, combining recessive mutants from complementation groups I and III or II and III is preferable.

2. Wild type mice are gray and in a large-scale breeding colony two white female mice arise from different parents. You would like to know whether the white phenotype is caused by two mutations in the same gene or in different genes. Using the concepts of dominance, recessivity, and complementation, describe a set of crosses and the interpretation of their outcomes that you would use to make this determination. Please be sure to indicate the circumstances that would prevent you from easily making this determination. (Assume you have available an unlimited number of true breeding wild-type gray mice.)

To determine whether two mutations occur in the same gene or in different genes, we must perform a complementation test. Before we can do this, however, we must first establish that the two mutations we're dealing with are recessive.

CROSS 1: Dominance/Recessive Test

Cross each white female with a true-breeding, wild-type gray male mouse. Denoting the wild-type allele as 'G', the white allele in Mutant 1 as 'g-1', and white allele in Mutant 2 as 'g-2', we have:

g-1/g-1 x G/G → G/g-1
g-2/g-2 x G/G → G/g-2

If 100% of the G/g-1 and G/g-2 heterozygotes from the F1 generation are phenotypically wild-type (i.e. gray), then we may conclude that the g-1 and g-2 alleles are both recessive. (Note: If we obtain additional white mice in the F1 generation from either cross, then the mutant allele from that particular cross is dominant, and we therefore cannot perform a complementation test with this mutant.)

CROSS 2: Generation of a g-1/g-1 OR g-2/g-2 white MALE mouse

Assuming that cross 1 showed the g-1 and g-2 alleles to be recessive, we would next perform a complementation test. However, we are told in the question that our two mutants are both female, and crossing these mice to one another is not possible. Therefore, we must generate a homozygous g-1/g-1 or g-2/g-2 mutant male mouse to perform a complementation test.

One could do this by crossing a G/g-1 heterozygote from the F1 generation of Cross 1 to the original g-1/g-1 mutant female. Half of the offspring from this cross will be g-1/g-1, some of which will be male. The same could be done with a G/g-2 F1 heterozygote and the original g-2/g-2 mutant female.

G/g-1 (F1 male from cross 1) x g-1/g-1 (orig. female mutant) → 50% G/g-1, 50% g-1/g-1

Also, male and female mice from the F1 generation of the g1 strain (or the g2 strain) could be mated to generate white male mice.

CROSS 3: Complementation Test

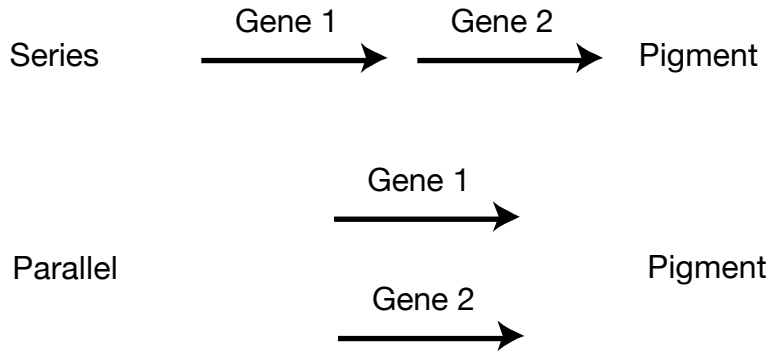
Now that we have a homozygous mutant male (g-1/g-1), we can cross this mouse to the original g-2/g-2 female mutant mouse. This is the complementation test. If the progeny from this cross are phenotypically wild-type (i.e. gray), then the mutations in g-1 and g-2 occur in different genes (complementation). If the progeny are phenotypically mutant (i.e. white), then mutations in g-1 and g-2 occur in the same gene (fail to complement).

g-1/g-1 (white male from Cross 2) x g-2/g-2 (original white female mutant)

or

g-2/g-2(white male from Cross 2) x g-1/g-1 (original white female mutant)

3. Genes that control coat color in mice can be thought of as steps in biochemical pathways whose products are pigmented compounds that give the fur its color. Albino mice have white fur because they lack the ability to make any pigment. Imagine mutations in two different genes that can, in certain combinations, block the production of pigment, yielding mutants with white fur. There are two different possible arrangements for two biochemical steps responsible for the formation of pigment: the two genes might act in *series* such that a loss of function of *either* gene would block the formation of pigment, or the two genes could act in *parallel* such that loss of function of *both* genes would be required to block the formation of pigment.



a) Say that you are given an albino mouse. When you cross this mouse to wild type all of the F1 progeny appear normal (i.e. like wild type). Consider the following three possibilities for the genetic basis of the albino trait: 1) recessive allele of a single gene, 2) recessive alleles of two genes acting in series 3) recessive alleles of two genes acting in parallel. For each of the three possibilities give the proportion of albino and normal looking mice among the F2 generation.

Hypothesis 1) Recessive allele of a single gene

F1 A/a X A/a

F2 1 A/A : 2 A/a: 1 a/a

Because the allele is recessive, only a/a mice will show the albino phenotype. We would expect a **3 normal : 1 albino** phenotypic ratio among the F2 progeny.

Hypothesis 2) Recessive alleles of two genes acting in series

F1 A/a, B/b X A/a, B/b

F2 9 A/-, B/- : 3 A/-, b/b : 3 a/a, B/- : 1 a/a, b/b

Because the genes are acting in series, possessing recessive albino alleles for either of the genes will produce the albino phenotype. Mice with an a/a OR b/b genotype will be albino. As a result, we would expect a phenotypic ratio of **9 normal : 7 albino** among the F2 progeny.

Hypothesis 3) Recessive alleles of two genes acting in parallel

F1 A/a, B/b X A/a, B/b

F2 9 A/-, B/- : 3 A/-, b/b : 3 a/a, B/- : 1 a/a, b/b

Because the genes are acting in parallel, possessing recessive albino alleles for both of the genes is required to produce the albino phenotype. Only mice with an a/a, b/b genotype will be albino. As a result, we would expect a phenotypic ratio of **15 normal : 1 albino** among the F2 progeny.

b) You cross the normal looking F1 mice among themselves producing 40 F2 mice; 15 are albino and 25 appear normal. Determine whether these data are consistent with each of the three possibilities outlined in part (a) and draw whatever conclusions you can about the inheritance of albinism. The table below gives chi square values for 1, 2 and 3 degrees of freedom. Use 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 and therefore we can reject the model on the basis of the experimental data.

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

To set up the Chi Squared Test, we must first determine the expected number of normal and albino mice under each hypothetical model. Based on the expected phenotypic ratios for each model calculated in part a, we can easily determine the expected number of normal and albino mice.

	Gray Mice (wt)	Albino Mice	$\Delta = \lvert O-E \rvert$
Observed	25	15	
Exp(Hypothesis 1)	30	10	5
Exp(Hypothesis 2)	22.5	17.5	2.5
Exp(Hypothesis 3)	37.5	2.5	12.5

The degrees of freedom equals the number of phenotypic classes minus 1. In this example, there are 2 phenotypic classes, normal and albino.

Degrees of freedom = # of classes – 1
Degrees of freedom = 2 – 1 = 1

We can now calculate the χ^2 value for each hypothesis with $\chi^2 = \Sigma (O-E)^2/E$ by plugging in the values we obtained above for each hypothesis.

Hypothesis 1:
$\chi^2 = \Sigma (O-E)^2/E = 5^2/10 + 5^2/30 = 3.33$

For 1 degree of freedom and a χ^2 value of 3.33, $0.05 < p < 0.1$. Given the convention that we only reject hypotheses for $p < 0.05$, we cannot reject the hypothesis.

Hypothesis 2:
$\chi^2 = \Sigma\ (O-E)^2/E = 2.5^2/22.5 + 2.5^2/17.5 = 0.63$

For 1 degree of freedom and a χ^2 value of 0.63, $0.1 < p < 0.5$. Given the convention that we only reject hypotheses for $p < 0.05$, we cannot reject the hypothesis.

Hypothesis 3:
$\chi^2 = \Sigma\ (O-E)^2/E = 12.5^2/2.5 + 12.5^2/37.5 = 66.6$

For 1 degree of freedom and a χ^2 value of 66.6, $p < 0.05$. Given the convention that we reject hypotheses for $p < 0.05$, we can reject hypothesis 3. The statistically significant difference between the observed data and the expected results of the model warrants rejection of the hypothesis.

Given that the p value for hypothesis 2 is larger than that for hypothesis 1, the model of two genes acting in series is more statistically consistent with the observed results. While this doesn't provide complete certainty that hypothesis 2 is accurate, we can reasonably say that this model represents the most likely mode of inheritance of the three models tested.

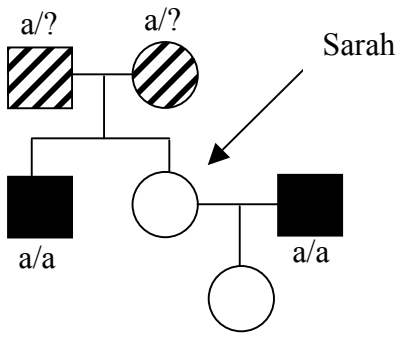
4. PKU is an autosomal recessive genetic disorder resulting from a loss of the enzyme phenylalanine hydroxylase, which converts phenylalanine to tyrosine. Without this enzyme, phenylalanine and its breakdown products accumulate to toxic levels resulting in mental retardation. Fortunately, individuals homozygous for the disease allele can be spared by a phenylalanine-free diet.

Say that Sarah has a brother with phenylketonuria (PKU), but she doesn't have the disease herself.

a) What is the probability that Sarah is a carrier for PKU?

Sarah's husband was detected to have PKU at birth but he has been treated with a phenylalanine-free diet. Sarah and her husband recently had their first child and are relieved that their child shows normal phenylalanine levels and thus is not homozygous for the disease allele.

b) Based on this new information use Bayes Theorem to compute an updated probability that Sarah is a carrier for PKU.



a) Despite our incomplete knowledge of the genotypes of Sarah's parents, we may calculate the probability that Sarah is a carrier for PKU as follows:

p(Sarah is carrier) = p(Parents are both Aa)*p(Sarah gets 1 A and 1 a) +
2*p(1 parent is aa and one parent is Aa)*p(Aa parent gives A to Sarah)

Notes:
(i) The combination of parental genotypes must allow Sarah to be 'Aa'.
(ii) Because Sarah's brother is affected, we know that each parent must have at least one 'a' allele.
(iii) The remaining possible combinations for parental genotypes are (1) Aa and Aa, (2) Aa and aa, and (3) aa and Aa. Combinations (2) and (3) are equally likely, and we may therefore condense the algebraic expression above by simply doubling the second term.
(iv) With three equally likely combinations of parental genotypes, the probability of any one combination is 1/3.
(v) Because Sarah is unaffected, her genotype cannot be 'aa'. If her parents are both Aa, this leaves 3 possibilities – AA, Aa, and aA. Her probability of being a carrier is then 2/3.
Taking these notes into account, we can insert values into the above equation to calculate the probability that Sarah is a carrier:

Taking these notes into account, we can insert values into the above equation to calculate the probability that Sarah is a carrier:

p(Sarah is carrier) = (1/3)*(2/3) + 2*(1/3)*(1) = 8/9

From the information given, there is an 88.9% chance that Sarah is a carrier for PKU.

****ASSUMING THAT BOTH OF SARAH'S PARENTS ARE CARRIERS****

a) p(Sarah is carrier) = 2/3 (See note (v)

****ASSUMING UNKNOWN GENOTYPES FOR SARAH'S PARENTS****

b) Because Sarah's child is unaffected, this new information should lead us to predict that the new probability that Sarah is a carrier (as opposed to AA) is less than 0.89. We will apply Bayes' Theorem to confirm:

X = Sarah is a carrier for PKU
Y = Sarah's child is unaffected by PKU

p(X) = 8/9
p(not X) = 1 – 8/9 = 1/9
p(Y | X) = ½
p(Y | not X) = 1

Bayes' Theorem:
p(X | Y) = p(Y | X)*p(X) / [p(Y | X)*p(X) + p(Y | not X)*p(not X)]
p(X | Y) = (1/2)*(8/9) / [(1/2)*(8/9) + (1)*(1/9)] = 0.80

Accounting for the new information that Sarah's child is unaffected by PKU, we find that, as expected, the probability that Sarah is a carrier decreases from 0.89 to 0.80.

****ASSUMING THAT BOTH OF SARAH'S PARENTS ARE CARRIERS****

X = Sarah is a carrier for PKU
Y = Sarah's child is unaffected by PKU

p(X) = 2/3
p(not X) = 1 – 2/3 = 1/3
p(Y | X) = ½
p(Y | not X) = 1

Bayes' Theorem:
p(X | Y) = p(Y | X)*p(X) / [p(Y | X)*p(X) + p(Y | not X)*p(not X)]
p(X | Y) = (1/2)*(2/3) / [(1/2)*(2/3) + (1)*(1/3)] = 0.50

7.03 Problem Set 2

Due before 5 PM on Friday, September 29

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

1. Hemophilia A is a X-linked recessive disorder characterized by dysfunctional blood clotting, due to a mutation in the gene for the clotting component, Factor VIII.

Jennifer's brother has hemophilia A, but neither Jennifer nor anyone else in her family show symptoms of the disorder.

a) If Jennifer has a son, what is the probability that he will have hemophilia?

b) Would this probability be different if Jennifer's husband had hemophilia? Explain.

Imagine that there is a DNA-based marker on the X chromosome that is 5 cM away from the gene for Factor VIII. This marker has two alleles, designated x1 and x2, that can be distinguished by a simple analysis of the DNA in a blood sample. The genotype of Jennifer for this marker is x1/x2, the genotype of her mother is x1/x2, the genotype of her father is x2, and the genotype of her brother is unknown.

c) Draw pictures to illustrate all possible arrangements of the given alleles for both of Jennifer's X chromosomes. Be sure to account for both of the DNA marker alleles and both alleles of the gene associated with hemophilia A.

d) What is the probability of each of the above arrangements occurring?

e) Given that Jennifer is a carrier, which of the above arrangements must be correct?

f) Given that Jennifer is a carrier, what is the probability that she will have a son that carries the x2 allele and has hemophilia?

2. The traits controlled by three autosomal *Drosophila* genes are easily distinguishable, since one locus determines body color, one eye color, and the other wing size. Mutants homozygous for recessive alleles of these genes exhibit black body (b/b), purple eyes (pr/pr), and vestigial wings (vg/vg). [Wild-type flies have brown bodies, red eyes, and large wings.]

The gene order is Vg—Pr—B and the distance between Vg and Pr is 13 cM and the distance between Pr and B is 7 cM. You cross a fly from a true-breeding line with brown body, red eyes, and large wings to a fly from a true-breeding line with black body, purple eyes, and vestigial wings. The F1 progeny (which have brown bodies, red eyes, and large wings) are then crossed to flies with black body, purple eyes, and vestigial wings. List all of the phenotypic classes that you would expect from this cross and calculate the number of each class expected out of a total of 1000 progeny.

3. Consider two unlinked yeast genes, His3 and His4. Each of these genes encodes an enzyme in the histidine biosynthesis pathway. Therefore, the His3$^-$ and His4$^-$ mutants require histidine for growth (i.e. the mutants are phenotypically His$^-$).

a) You mate a MAT**a** His3$^-$ strain to a MATα His4$^-$ strain to produce heterozygous diploids. If the resulting diploids are His$^+$ what does this tell you about the His3$^-$ and His4$^-$ mutations?

b) Next you sporulate the diploids to produce 60 tetrads. Describe the types of tetrads you should get (in terms of the ratio His$^+$: His$^-$) and the expected number among the 60 tetrads.

c) Would you expect the result from part (b) to be different if both the His3$^-$ and His4$^-$ mutants were dominant? Explain.

d) Now you cross a wild type MAT**a** strain to a MATα His3$^-$ His4$^-$ double mutant. Out of 60 tetrads from this cross, how many of each tetrad type would you expect to get?

e) The His4 gene is relatively large. Say that you have a His4-1$^-$ mutation that is at one end of the gene and a His4-2$^-$ mutation at the other end of the gene. If the His4-1$^-$ and His4-2$^-$ mutations are exactly 1 cM apart, how many of each tetrad type would you expect from a cross of a MAT**a** His4-1$^-$ strain with a MATα His4-2$^-$ strain (you analyze a total of 50 tetrads)?

f) How many of each tetrad type would you expect from a cross of a MAT**a** His4-1$^-$ strain with a MATα His4-1$^-$ His4-2$^-$ double mutant (you analyze a total of 50 tetrads)?

7.03 Problem Set 2

Due before 5 PM on Friday, September 29

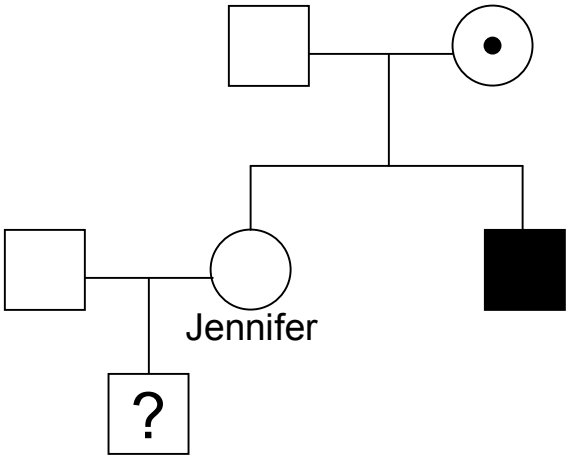
Hand in answers in recitation section or in the box outside of 68-120

1. Hemophilia A is a X-linked recessive disorder characterized by dysfunctional blood clotting, due to a mutation in the gene for the clotting component, Factor VIII.

Jennifer's brother has hemophilia A, but neither Jennifer nor anyone else in her family show symptoms of the disorder.

a) If Jennifer has a son, what is the probability that he will have hemophilia?

Based on the above information we can draw the following pedigree:



Since we know that Jennifer's brother has hemophilia, we are certain that Jennifer's mother must be a carrier. For Jennifer to have an affected son she would have to first be a carrier and subsequently pass the mutant X chromosome on to her son.

Probability Jennifer is a carrier = 1/2
Probability Jennifer would pass her mutant X chromosome to her son = 1/2
If Jennifer had a son, the probability that he would have hemophilia is therefore (1/2)(1/2) = 1/4.

b) Would this probability be different if Jennifer's husband had hemophilia? Explain.

The probability would NOT be different if Jennifer's husband had hemophilia. Since hemophilia is inherited as an X-linked recessive trait, the son will not inherit an X chromosome from his father. Only the mother's genotype is relevant.

Imagine that there is a DNA-based marker on the X chromosome that is 5 cM away from the gene for Factor VIII. This marker has two alleles, designated x1 and x2, that can be distinguished by a simple analysis of the DNA in a blood sample. The genotype of Jennifer for this marker is x1/x2, the genotype of her mother is x1/x2, the genotype of her father is x2, and the genotype of her brother is unknown.

c) Draw pictures to illustrate all possible arrangements of the given alleles for both of Jennifer's X chromosomes. Be sure to account for both of the DNA marker alleles and both alleles of the gene associated with hemophilia A.

a) x1 / + ; x2 / +

b) x1 / m ; x2 / +

d) What is the probability of each of the above arrangements occurring?

Each arrangement is equally likely. a) 50% b) 50%

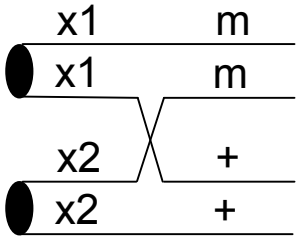
e) Given that Jennifer is a carrier, which of the above arrangements must be correct?

Arrangement b) must be correct.

f) Given that Jennifer is a carrier, what is the probability that she will have a son that carries the x2 allele and has hemophilia?

For Jennifer to have a son that carries the x2 allele and has hemophilia, two conditions must be met.
Condition 1. Jennifer must pass on an X chromosome carrying the x2 allele of the DNA marker and the mutant allele of the gene coding for Factor VIII.
Condition 2. Jennifer's husband must pass on a Y chromosome.

To satisfy Condition 1, there must be a recombination event between the DNA marker and the gene coding for Factor VIII during Meiosis 1 in Jennifer.



Since we know the DNA marker and the gene coding for Factor VIII are 5 cM apart, we know Jennifer has a 5% chance of passing on a recombinant chromosome. Therefore, there is a 2.5% chance she will pass on chromosome A and a 2.5% chance she will pass on chromosome B.

A x2 / m **B** x1 / +

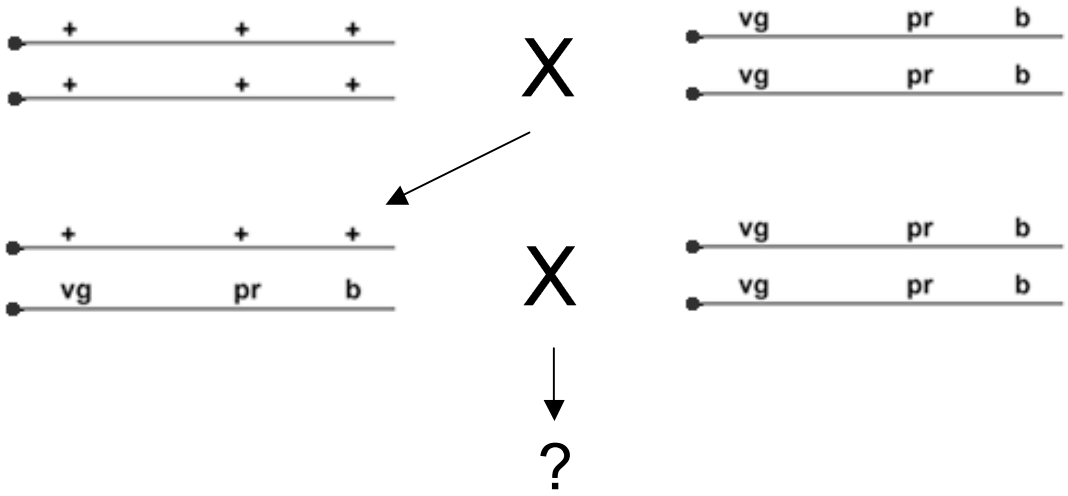
There is a 50% chance that Jennifer's husband will pass on a Y chromosome.

Therefore, there is a (.50)(.025)= .0125 or 1.25% chance that Jennifer will have a son that carries the x2 allele and has hemophilia.

2. The traits controlled by three autosomal *Drosophila* genes are easily distinguishable, since one locus determines body color, one eye color, and the other wing size. Mutants homozygous for recessive alleles of these genes exhibit black body (b/b), purple eyes (pr/pr), and vestigial wings (vg/vg). [Wild-type flies have brown bodies, red eyes, and large wings.]

The gene order is Vg—Pr—B and the distance between Vg and Pr is 13 cM and the distance between Pr and B is 7 cM. You cross a fly from a true-breeding line with brown body, red eyes, and large wings to a fly from a true-breeding line with black body, purple eyes, and vestigial wings. The F1 progeny (which have brown bodies, red eyes, and large wings) are then crossed to flies with black body, purple eyes, and vestigial wings. List all of the phenotypic classes that you would expect from this cross and calculate the number of each class expected out of a total of 1000 progeny.

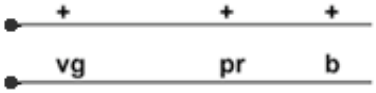
Cross Scheme



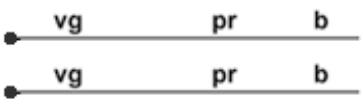
To determine the phenotypic classes that would result from this cross, we must consider the possible products of meiosis in the female.

If no recombination occurs, we would generate the following two classes:

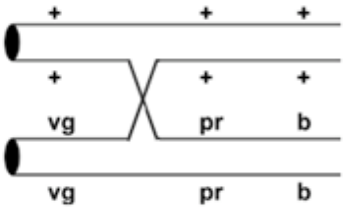
Class A would be phenotypically Wild-type



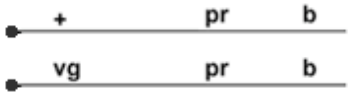
Class B would have a black body, purple eyes and vestigial wings.



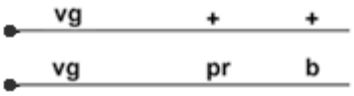
If a single recombination event occurs between vg and pr, we would generate the following two classes:



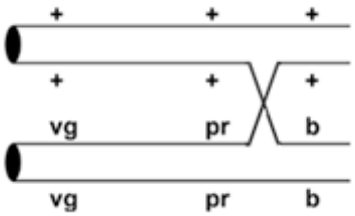
Class C would have a black body and purple eyes



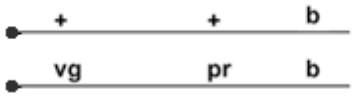
Class D would have vestigial wings



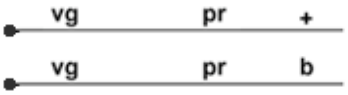
If a single recombination event occurs between pr and b, we would generate the following two classes:



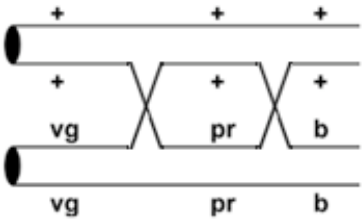
Class E would have a black body



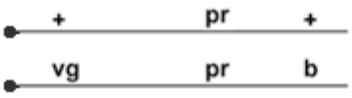
Class F would have purple eyes and vestigial wings



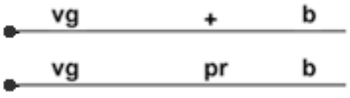
If we had a recombination event between vg and pr AND between pr and b, we would generate the following two classes:



Class G would have purple eyes



Class H would have a black body and vestigial wings



Since we know that vg and pr are 13 cM apart, we would predict that out of 1000 progeny, 130 flies would have resulted from a crossover between vg and pr.

Looking at the classes above, we know that Classes C, D, G and H all result from a recombination event between vg and pr.
Therefore, Class C + Class D + Class G + Class H = 130

Since we know that pr and b are 7 cM apart, we would predict that out of 1000 progeny, 70 flies would have resulted from a crossover between pr and b.

Looking at the classes above, we know that Classes E, F, G and H all result from a recombination event between pr and b.
Therefore, Class E + Class F + Class G + Class H = 70

The only classes for which you could not directly predict the expected number of progeny are Class G and Class H. Class G and Class H are the progeny resulting from double crossover events.

The probability of having a double crossover event is the product of the probability of either single crossover event. Therefore, the probability of having a double crossover is

$$P(DC) = (.13)(.07) = .009$$

So out of 1000 progeny, 9 would be either Class G or Class H. Since Class G and Class H will occur with equal frequency, we will say that we would expect to see

4.5 Class G flies out of 1000 and 4.5 Class H flies out of 1000

We can now go back to our previous equations and solve for the predicted number of progeny in each of the remaining classes.

We know that Class C progeny will occur with equal frequency as Class D progeny. Therefore, Class C = Class D. Our equation now becomes

$$2 \times (\text{Class C}) + 9 = 130$$ **Note: "9" is the sum of Classes G and H**

$$2C = 121$$

$$C = 60.5$$

60.5 Class C flies out of 1000 and 60.5 Class D flies out of 1000

We know that Class E progeny will occur with equal frequency as Class F progeny. Therefore, Class E = Class F. Our equation now becomes

$$2 \times (\text{Class E}) + 9 = 70$$ **Note: "9" is the sum of Classes G and H**

$$2E = 61$$

$$E = 30.5$$

30.5 Class E flies out of 1000 and 30.5 Class F flies out of 1000

To determine the number of expected progeny in Class A and Class B, we simply subtract the total of the previous classes from 1000.
Class A + Class B = 1000 - (Class C + Class D + Class E + Class F + Class G + Class H)
Class A + Class B = 1000 - 191 = 809

Since we know that Class A and Class B progeny will occur with equal frequency, we know that the number of Class A progeny and the number of Class B progeny will be equal to 809/2 or 404.5.

404.5 Class A flies out of 1000 and 404.5 Class B flies out of 1000

Therefore, out of 1000 progeny you would expect to see:
404.5 Class A Wildtype
404.5 Class B black body, purple eyes, and vestigial wings
60.5 Class C black body, purple eyes
60.5 Class D vestigial wings
30.5 Class E black body
30.5 Class F purple eyes and vestigial wings
4.5 Glass G purple eyes
4.5 Class H black body and vestigial wings

3. Consider two unlinked yeast genes, His3 and His4. Each of these genes codes for an enzyme in the histidine biosynthesis pathway and therefore $His3^-$ and $His4^-$ mutants require histidine for growth (i.e. the mutants are phenotypically His^-).

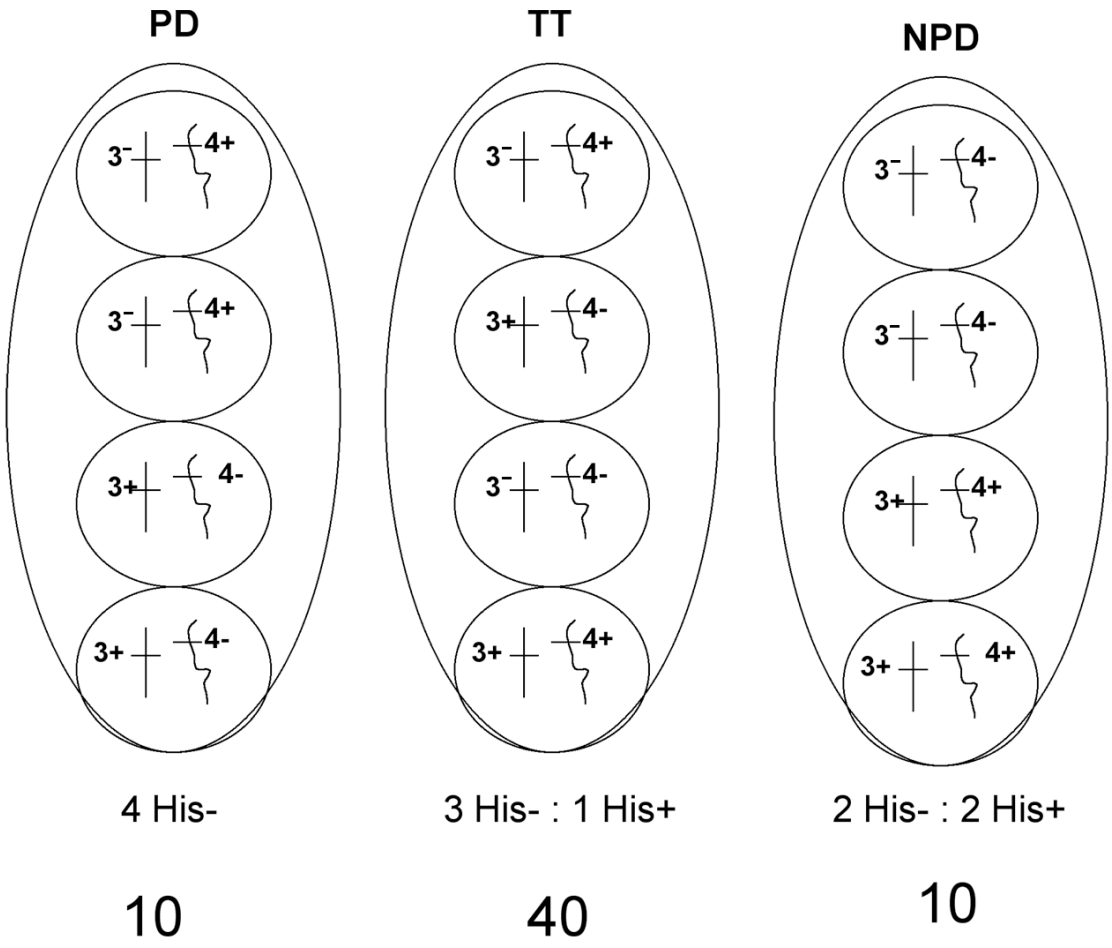
a) You mate a MAT**a** $His3^-$ strain to a MATα $His4^-$ strain to produce heterozygous diploids. If the resulting diploids are His^+ what does this tell you about the $His3^-$ and $His4^-$ mutations?

The His3- and His4- mutations are both recessive.
In the heterozygous diploid, you have His3+/His3- and His4+/His4-. The overall His+ phenotype shows the His3+ and His4+ alleles are dominant to their respective mutant alleles.

b) Next you sporulate the diploids to produce 60 tetrads. Describe the types of tetrads you should get (in terms of the ratio His^+: His^-) and the expected number among the 60 tetrads.

For two unlinked genes, the ratio expected for all of the different tetrad types is as follows:
1 PD: 4 TT: 1 NPD.

Please refer to the diagram below for exact genotypes of all of the spores within the asci and the resulting His phenotype.



c) Would you expect the result from part (b) to be different if both the His3⁻ and His4⁻ mutants were dominant? Explain.

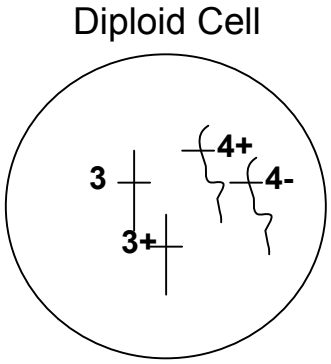
No, there would be no difference. The results we obtain from a tetrad analysis are directly from the spores, which are haploid. With only one copy of each gene in the spores, dominant AND recessive alleles can be unmasked and therefore observed in the experiment.

d) Now you cross a wild type MAT**a** strain to a MATα His3⁻ His4⁻ double mutant. Out of 60 tetrads from this cross, how many of each tetrad type would you expect to get?

As in b.), there will still be the following numbers, *with relation to His phenotype*:
10 All His- 40 3His-:1His+ 10 2His-:2His+

The only difference is that the PDs are now 2 His-: 2 His+ and the NPDs are the 4 His-. This is due to the change in the genotypes of the parents from that in part b.)

In both b.) and d.), the heterozygous diploids resulting from mating the parents will look the same:



Therefore once meiosis is initiated the outcome (in terms of His phenotypes of spores) will be the same.

e) The His4 gene is relatively large. Say that you have a His4-1$^-$ mutation that is at one end of the gene and a His4-2$^-$ mutation at the other end of the gene. If the His4-1$^-$ and His4-2$^-$ mutations are exactly 1 cM apart, how many of each tetrad type would you expect from a cross of a MAT**a** His4-1$^-$ strain with a MATα His4-2$^-$ strain (you analyze a total of 50 tetrads)?

We are dealing with two mutations in the same gene, which means they must be linked. For linked loci, the equation we use is:

cM = 100 x [(TT + 6NPD)/2 Σ]

We know the mutations are 1 cM apart, and we look at 50 tetrads, so

1 = 100 x [(TT + 6 NPD)/100],

1 = TT + 6NPD

Recall that NPDs are designators of a particular double cross-over that can occur in meiosis, so in order to score a tetrad as an NPD, there must have been that specific cross-over within the 50 tetrads we analyze.
The probability of getting a double-cross over event between loci 1cM apart is

probability of 1 cross-over (.01) x probability of another cross-over (.01) = 1/10000.

Another way to say this is that, on average, 1 out of every 10000 tetrads results from a double cross-over. Since only 1 in 4 double cross-overs result in tetrad we can recognize (NPD), we would only see a quarter of the total double-crossovers, which means that on average, we would have to score 40,000 tetrads to observe an NPD. Since it is HIGHLY UNLIKELY that within the first 50 tetrads we come across an NPD, the number of NPDs is zero.

Returning to our equation:

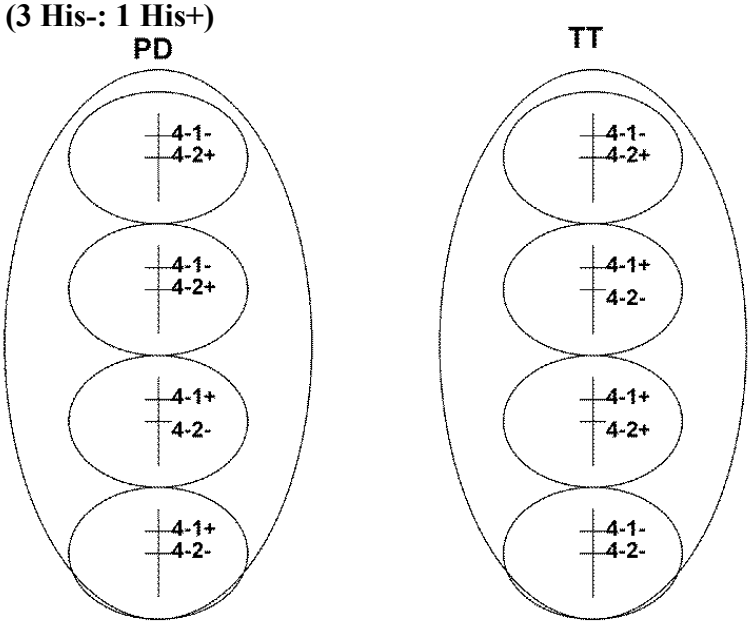
1 = TT + 6NPD

1 = TT + 6 x 0

1 = TT

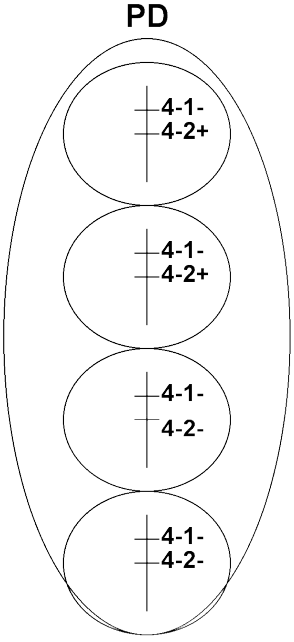
The remaining tetrads must be PDs, so there are 49 PDs.

49 PDs (4His-); 1 TT (3 His-: 1 His+)



f) How many of each tetrad type would you expect from a cross of a MAT**a** His4-1$^-$ strain with a MATα His4-1$^-$ His4-2$^-$ double mutant (you analyze a total of 50 tetrads)?

You will get all PDs, which are tetrads with all His- spores. See diagram below...



7.03 Problem Set 3

Due before 5 PM on Wednesday, October 18

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

1. The following DNA sequence fragment comes from the middle of a bacterial gene. To start the analysis of this coding sequence you will first need to find the open reading frame (note that you do not know the orientation of the gene).

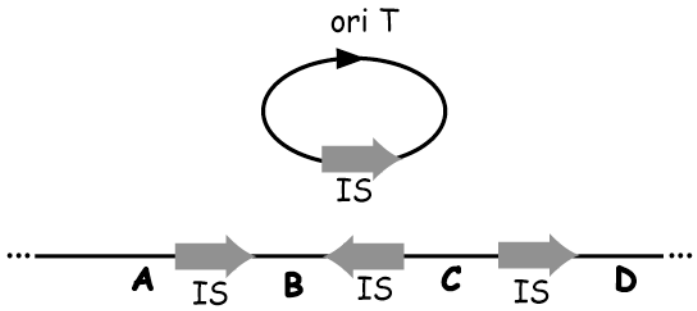
5' CTCGGCTAATATCGATCGCTAGTGTCATAGCTCTCGGGTAATGACGATCACGA 3'

a) Within this segment of DNA note all of the possible nonsense mutations that can be produced by a single base change by a mutagen that causes only transition mutations (G•C to A•T *or* A•T to G•C).

b) Within this segment of DNA note all of the possible single-base change nonsense mutations that can be produced by a single-base transversion mutation.

c) Consider the gene for $tRNA^{trp}$. Write out the double-stranded DNA segment of this gene that codes for the anticodon of $tRNA^{trp}$ (be sure to label 5' and 3' ends). Write out all of the possible mutations that can convert $tRNA^{trp}$ to a nonsense suppressing tRNA. For each mutation indicate whether it is a transition or transversion and which kind of nonsense mutation will be suppressed.

2. 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. Assume that you have available a variety of strains with mutations in the genetic markers A, B, C, D.



a) Describe with as much detail as you can how you would use this F^+ strain to isolate an F' factor that carries the B marker. For your answer diagram any relevant intermediate strains as well as the final F' factor. For your answer please show all of the markers as well as the position and orientation of each IS sequence and the origin of transfer (ori T).

3. Wild type *E. coli* can utilize the sugar galactose and is therefore phenotypically Gal$^+$. You have isolated a mutant that you call *gal1*$^-$, which cannot grow on galactose (Gal$^-$).

a) You have a wild type (Gal$^+$) strain carrying a Tn5 insertion. You grow P1 phage on this strain and use the resulting phage lysate to infect the *gal1*$^-$ strain, selecting for kanamycin resistance (Kan^r). Among 100 Kan^r transductants, you find that 75 are Gal$^+$ and 25 are Gal$^-$. What does this result tell you about the relationship between the *gal1*$^-$ mutation and the Tn5 insertion?

b) You grew P1 phage on one of the Gal$^-$ Kan^r transductants isolated in part (**a**) and then used these phage to transduce a wild-type strain. What fraction of the Kan^r transductants would be Gal$^+$?

c) You isolate a second Gal$^-$ mutation, which you designate *gal2*$^-$. Using the same P1 lysate as in part (**a**) you infect the *gal2*$^-$ strain, selecting for Kan^r transductants. In this case, none of the 100 Kan^r transductants are Gal$^+$. What does this result tell you about the relationship between the *gal1*$^-$ and *gal2*$^-$ mutations?

d) Next, you isolate a third Gal$^-$ strain, called *gal3*$^-$. Preliminary P1 transduction experiments indicate that *gal3*$^-$ is linked to the Tn5 insertion described in part (**a**). To map *gal3*$^-$ relative to *gal1*$^-$ you set up two reciprocal crosses. In the first cross you grow P1 on a strain that carries the Tn5 insertion and the *gal1*$^-$ mutation. You then use this lysate to infect a *gal3*$^-$ mutant and select for Kan^r. From 100 Kan^r transductants examined, 85 are Gal$^-$ and 15 are Gal$^+$. In the second cross you grow P1 on a strain that carries the Tn5 insertion and the *gal3*$^-$ mutation. You then use this lysate to infect a *gal1*$^-$ mutant, and select for Kan^r. From 100 Kan^r transductants examined, 98 are Gal$^-$and 2 are Gal$^+$. Draw a genetic map showing the relative positions of the Tn5 insertion and the *gal1*$^-$ and *gal3*$^-$ mutations. Express any measured distances as co-transduction frequencies.

e) Explain why it is necessary to carry out two reciprocal three-factor crosses in part (**d**) in order to determine the relative positions of the *gal1*$^-$ and *gal3*$^-$ mutations.

4. An F^- $HisA^-$ *E. coli* strain can be converted to His^+ by a variety of different genetic manipulations including: transduction with a P1 phage lysate grown on a $HisA^+$ strain, mating to an Hfr strain that carries $HisA^+$ on the chromosome, or mating to a strain with $HisA^+$ on an F' factor. You are given a variety of $HisA^-$ strains with unknown genetic properties. You subject each strain to a variety of genetic tests to diagnose how it may have been altered. Based on the outcome of these tests, deduce which genetic capabilities have been altered then propose a specific type of mutation or genetic alteration that might give rise to these properties. (This question is intended to stretch your thinking about bacterial genetics somewhat beyond what has been explicitly covered in lecture. Possibilities you should consider include acquisition of various kinds of mutations or extra chromosomal elements. For some strains more than one mechanism is possible.)

a) Strain 1 can be converted to His^+ by conjugation with either a $HisA^+$ Hfr or an F' $HisA^+$ strain, but cannot be converted to His^+ by P1 transduction.

b) Strain 2 can be converted to His^+ by conjugation with an F' $HisA^+$ strain, but cannot be converted to His^+ by P1 transduction or by conjugation with a $HisA^+$ Hfr.

c) Strain 3 can be converted to His^+ by P1 transduction, but cannot be converted to His^+ by conjugation with either a $HisA^+$ Hfr or an F' $HisA^+$ strain.

d) Strain 4 can be converted to His^+ by P1 transduction or by conjugation with a $HisA^+$ Hfr, but cannot be converted to His^+ by congugation with an F' $HisA^+$ strain.

e) Strain 5 cannot be converted to His^+ by P1 transduction or by conjugation with either a $HisA^+$ Hfr or an F' $HisA^+$ strain.

7.03 Problem Set 3

Due before 5 PM on Wednesday, October 18

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

1. The following DNA sequence fragment comes from the middle of a bacterial gene. To start the analysis of this sequence you will first need to find the open reading frame (note that you do not know the orientation of the gene).

5' CTCGGCTAATATCGATCGCTAGTGTCATAGCTCTCGGGTAATGACGATCACGA 3'

a) Within this segment of DNA note all of the possible nonsense mutations that can be produced by a single base change by a mutagen that causes only transition mutations (G•C to A•T *or* A•T to G•C).

The first step is to determine the open reading frame (ORF). There are 6 possibilities: 2 orientations (left to right or right to left) and 3 reading frames (start from the first, second, or third base pair). The correct ORF is the one without any stop codons, since this DNA is in the middle of a bacterial gene. There is only one ORF that fits the requirement, and it reads from right to left, starting from the second nucleotide.

So the coding DNA (that will end up resembling the mRNA) strand reads:

5' T CGT GAT CGT CAT TAC CCG AGA GCT ATG ACA CTA GCG ATC GAT ATT AGC CGA G 3'

Transition mutations are mutations that change one purine·pyrimidine pair to the other purine·pyrimidine pair. There is only one single base transition nonsense mutation and it is highlighted by ().

5' T CGT GAT CGT CAT TAC CCG AGA GCT ATG ACA CTA GCG ATC GAT ATT AGC (C)GA G 3'

You can find the potential nonsense mutations by mutating each stop codon's base pair, as follows. Then search for the mutated stop codons in the sequence (Highlighted in yellow)

Stop Codon	Mutate 1st nt	Mutate 2nd nt	Mutate 3rd nt
TAA	TGA	TAG	CAA
TAG	CAG	TGG	TAA
TGA	CGA	TAA	TGG

1b) Within this segment of DNA note all of the possible single-base change nonsense mutations that can be produced by a single-base transversion mutation.

Transversion mutations change a purine to pyrimidine and vice versa. There are two nonsense transversion mutations, and they are highlighted by [].

The first mutation can change from a C → A or G. The second mutation is a mutation of A → T.

5' T CGT GAT CGT CAT TA[C] CCG [A]GA GCT ATG ACA CTA GCG ATC GAT ATT AGC CGA G 3'

You can also find these potential nonsense mutations by mutating each stop codon's base pair. (The ones crossed out are degenerative of a previous mutation.)

Stop Codon	Mutate 1st nt	Mutate 2nd nt	Mutate 3rd nt
TAA	AAA	TTA	TAT
	GAA	TCA	TAC
TAG	AAG	TTG	~~TAT~~
	GAG	TCG	~~TAC~~
TGA	AGA	~~TTA~~	TGT
	GGA	~~TCA~~	TGC

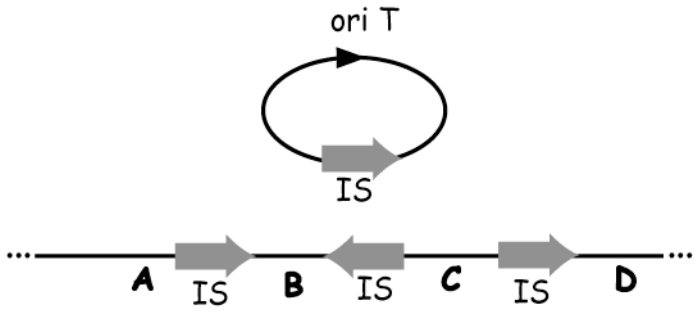
1c) Consider the gene for $tRNA^{trp}$. Write out the double-stranded DNA segment of this gene that codes for the anticodon of $tRNA^{trp}$ (be sure to label 5' and 3' ends). Write out all of the possible mutations that can convert $tRNA^{trp}$ to a nonsense suppressing tRNA. For each mutation indicate whether it is a transition or transversion and which kind of nonsense mutation will be suppressed.

Trp codon: 5' UGG 3'
Trp anticodon: 3' ACC 5'
Gene for the anticodon of $tRNA^{trp}$: 5' TGG 3' → template strand
3' ACC 5' → coding strand

Possible mutations:

1. Mutate first G·C -> A·T (transition) get a UAG (amber) nonsense mutation suppressed
2. Mutate second G·C -> A·T (transition) get a UGA (opal) nonsense mutation suppressed
3. Mutate both G·C --> A·T (two transitions, double mutations rare) get a UAA (ochre) nonsense mutation suppressed

2. 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. Assume that you have available a variety of strains with mutations in the genetic markers A, B, C, D.



a) Describe with as much detail as you can how you would use this F^+ strain to isolate an F' factor that carries the B marker. For your answer diagram any relevant intermediate strains as well as the final F' factor. For your answer please show all of the markers as well as the position and orientation of each IS sequence and the origin of transfer (ori T).

(The numbers refer to the diagram on the next page)

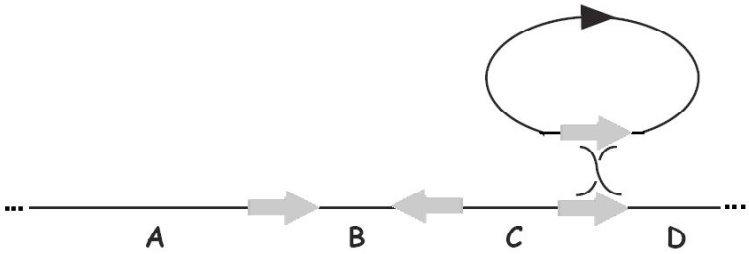
First, a crossover must happen between the IS on the F factor and the IS between markers C and D on the *E. coli* chromosome (1.)

This results in an Hfr strain that can transfer marker D early and efficiently, but transfers marker B very late (2.). Note that this is the only crossover event that results in an Hfr strain with these properties. These properties can be screened for by conjugating the Hfr strain to D^- and B^- strains, respectively.

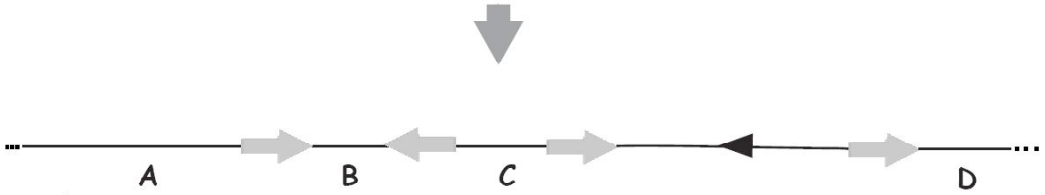
Within the identified Hfr strain, the *E. coli* chromosome can "loop around" such that a crossover event occurs between the "first" and "last" IS sequences of interest (3.).

If this event occurs, the resulting strain will contain an F' factor carrying the B marker (as well as the C marker) (4.). Note that this F' factor can now transfer B and C early and efficiently. Thus, identification of a strain from a population of the Hfr strain in (2.) that can now transfer B efficiently will allow us to isolate the desired F' factor.

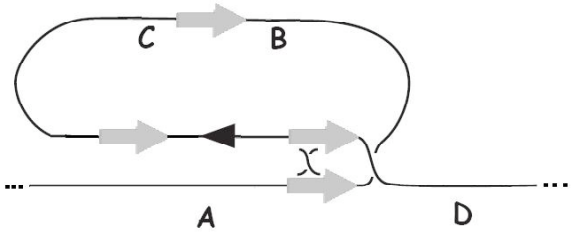
1.)



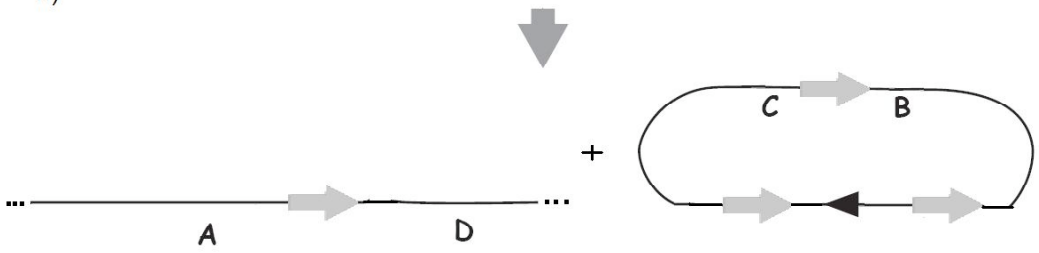
2.)



3.)



4.)



3. Wild type *E. coli* can utilize the sugar galactose and is therefore phenotypically Gal$^+$. You have isolated a mutant that you call *gal1*$^-$, which cannot grow on galactose (Gal$^-$).

a) You have a wild type (Gal$^+$) strain carrying a Tn5 insertion. You grow P1 phage on this strain and use the resulting phage lysate to infect the *gal1*$^-$ strain, selecting for kanamycin resistance (Kan^r). Among 100 Kan^r transductants, you find that 75 are Gal$^+$ and 25 are Gal$^-$. What does this result tell you about the relationship between the *gal1*$^-$ mutation and the Tn5 insertion?

The gal1- mutation and the Tn5 insertion can be considered linked due to the 75% co-transduction frequency.

The donor strain is Gal+ (with Tn5 insertion) and the recipient strain is Gal-. Therefore, in the Gal+ Kan^r transductants, *gal1+* was co-transduced with Tn5. This indicates that *gal1* and Tn5 are within 10^5 base pairs apart and are linked. The distance between Tn5 and the gal1- mutation is:

(75/100)* 100% = 75%.

b) You grew P1 phage on one of the Gal$^-$ Kan^r transductants isolated in part (**a**) and then used these phage to transduce a wild-type strain. What fraction of the Kan^r transductants would be Gal$^+$?

0.25 - The donor strain is Gal- (with Tn5 insertion) and the recipient strain is Gal+ (WT). Thus, Gal+ Kan^r transductants result when Tn5 and *gal1-* are not co-transduced. Since the probability of co-transduction of *gal1-* and Tn5 is 0.75 (**from part a**), then:

P (*gal1-* and Tn5 not co-transduced) = 1 – P(*gal1-* and Tn5 co-transduced)
= 1- 0.75
= 0.25

c) You isolate a second Gal$^-$ mutation, which you designate *gal2*$^-$. Using the same P1 lysate as in part (**a**) you infect the *gal2*$^-$ strain, selecting for Kan^r transductants. In this case, none of the 100 Kan^r transductants are Gal$^+$. What does this result tell you about the relationship between the *gal1*$^-$ and *gal2*$^-$ mutations?

gal1- and *gal2-* are unlinked and in different genes (0% co-transduction frequency, more than 10^5 bp apart)

The donor strain is Gal+ and the recipient strain is Gal-. Since we do not see any Kan^r Gal+ transductants, we can conclude that Tn5 and *gal2* were never co-transduced. This indicates that the distance between *gal2-* and Tn5 is at least one phage head apart (10^5 bp). We also know from **part (a)** that Tn5 and *gal1* are linked because they have a co-transduction frequency of 75%. Thus, we can conclude that gal1 and gal2 are greater than 10^5 bp from each other, unlinked, and in two different genes.

With regard to the relative order, if we knew that Tn5 was between *gal1* and *gal2* we could conclusively state that *gal1* and *gal2* are also greater than one phage head apart. However, since Tn5 and *gal1* are 75% co-transduced (meaning closely linked), we can probably still conclude that *gal1* and *gal2* are greater than one phage head apart.

d) Next, you isolate a third Gal$^-$ strain, called *gal3*$^-$. Preliminary P1 transduction experiments indicate that *gal3*$^-$ is linked to the Tn5 insertion described in part (**a**). To map *gal3*$^-$ relative to *gal1*$^-$ you set up two reciprocal crosses. In the first cross you grow P1 on a strain that carries the Tn5 insertion and the *gal1*$^-$ mutation. You then use this lysate to infect a *gal3*$^-$ mutant and select for Kan^r. From 100 Kan^r transductants examined, 85 are Gal$^-$ and 15 are Gal$^+$. In the second cross you grow P1 on a strain that carries the Tn5 insertion and the *gal3*$^-$ mutation. You then use this lysate to infect a *gal1*$^-$ mutant, and select for Kan^r. From 100 Kan^r transductants examined, 98 are Gal$^-$and 2 are Gal$^+$. Draw a genetic map showing the relative positions of the Tn5 insertion and the *gal1*$^-$ and *gal3*$^-$ mutations. Express any measured distances as co-transduction frequencies.

Order: Tn5----*gal3*----*gal1*. The only distance determinable is Tn5 to *gal1*, which is 75% (**from part a**). The best way to solve this type of problem is to draw out the two crosses. We will consider only two possible orders, instead of three, because the order where Tn5 is in the middle is impractical for three-factor co-transduction experiments.

There are two reciprocal crosses and two possible orders, thus there are four diagrams to draw: (1+ and 1- represent WT *gal1+* and mutant *gal1-*; 3+ and 3- represent WT *gal3+* and mutant *gal3-*)

	Order #1 (Tn5, *gal1*, *gal3*)	Order #2 (Tn5, *gal3*, *gal1*)
Cross #1 (Tn5, 1- infect 3-) 15 Gal+	----Tn5--------1^-----------3^+--- X X X X ----------1^+----------3^----	----Tn5--------3^+----------1^---- X X ----------3^-----------1^+--
Cross #2 (Tn5, 3- infect 1-) 2 Gal+	----Tn5--------1^+----------3^---- X X ----------1^-----------3^+--	----Tn5--------3^-----------1^+--- X X X X ----------3^+----------1^----

The data given allow us to determine which of the two possible orders is correct. As in any three-factor cross, we determine the order by looking for the rarest class. In this case, the rarest class is Gal+, and the genotype is Tn5, 1+, 3+.

From these data, we see that cross #1 produced 15 Gal+, while cross #2 produced only 2 Gal+. This is the key observation that allows us to determine order.

Looking back at the diagram, each X represents a crossover that is needed to make a Tn5, 1+ 3+ genotype.

If we assume that order #1 is correct, then we would get more Gal+ transductants in cross #2 than cross #1 because only two crossovers are required as opposed to four. If we assume that order #2 is correct, then we would get more Gal+ transductants from cross #1. The data show that cross #1 produces more Gal+ transductants, thus order #2 is correct:

-----Tn5-------*gal3*-------*gal1*-----
|----------75%----------|

The distance between Tn5 and *gal3* cannot be determined. Though we see that we get 15 Gal+ from a Tn5 and *gal3*+ co-transduction, the 15 does not represent all possible Tn5 and *gal3* co-transductions. The 15 only represents a crossing over somewhere to the left of Tn5 and the other crossing over event happening between *gal3* and *gal1* to generate the Gal+ phenotype. The co-transduction frequency of Tn5 and *gal3* must represent the crossing over events that happen to the left of Tn5 and anywhere to the right of *gal3*.

The distance between *gal3* and *gal1* also is not determinable, even if we could solve the distance between Tn5 and *gal3*. This is because we selected for kanamycin resistant transductants, therefore all data are relative to the Tn5 marker. Co-transduction frequencies are not additive.

e) Explain why it is necessary to carry out two reciprocal three-factor crosses in part (**d**) in order to determine the relative positions of the *gal1*$^-$ and *gal3*$^-$ mutations.

We need both crosses to determine which cross is less frequent, since both a double crossover and a quadruple crossover can result in the same Gal+ phenotype.

4. An F^- $HisA^-$ *E. coli* strain can be converted to His^+ by a variety of different genetic manipulations including: transduction with a P1 phage lysate grown on a $HisA^+$ strain, mating to an Hfr strain that carries $HisA^+$ on the chromosome, or mating to a strain with $HisA^+$ on an F' factor. You are given a variety of $HisA^-$ strains with unknown genetic properties. You subject each strain to a variety of genetic tests to diagnose how it may have been altered. Based on the outcome of these tests, deduce which genetic capabilities have been altered then propose a specific type of mutation or genetic alteration that might give rise to these properties. (This question is intended to stretch your thinking about bacterial genetics somewhat beyond what has been explicitly covered in lecture. Possibilities you should consider include acquisition of various kinds of mutations or extra chromosomal elements. For some strains more than one mechanism is possible.)

4a) Strain 1 can be converted to His^+ by conjugation with either a $HisA^+$ Hfr or an F' $HisA^+$ strain, but cannot be converted to His^+ by P1 transduction.

A large deletion mutation may not be complemented by the 100 kb DNA fragment carried by the phage

The mutations could be in genes that are greater than 100kb apart.

The mutant *E. coli* strain could be resistant to phage P1.

b) Strain 2 can be converted to His^+ by conjugation with an F' $HisA^+$ strain, but cannot be converted to His^+ by P1 transduction or by conjugation with a $HisA^+$ Hfr.

This may result from a large deletion mutation, so there is no recombination due to the absence of homologous DNA in the mutant chromosome.

This may result from a mutation in the cell's homologous recombination system, e.g. in gene RecA.

c) Strain 3 can be converted to His^+ by P1 transduction, but cannot be converted to His^+ by conjugation with either a $HisA^+$ Hfr or an F' $HisA^+$ strain.

This may result from mutations that disrupt histidine biosynthesis *and* render the mutant *E. coli* strain conjugation^{--} (i.e. unable to recognize the mating pilus).

d) Strain 4 can be converted to His^+ by P1 transduction or by conjugation with a $HisA^+$ Hfr, but cannot be converted to His^+ by congugation with an F' $HisA^+$ strain.

This may result from a mutation that:

Results in a dominant his^- allele (intragenic)

Cannot re-circularize or stably maintain the F plasmid, rapidly degrades the F plasmid, or covalently modifies the plasmid such that gene expression is prevented.

e) Strain 5 cannot be converted to His^+ by P1 transduction or by conjugation with either a $HisA^+$ Hfr or an F' $HisA^+$ strain.

This may result from a mutation resulting in a constitutive trans-acting (extragenic) repressor (or super-repressor) of histidine biosynthesis.

7.03 Problem Set 4

Due before 5 PM on Monday, October 30

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

1. In lecture we have seen in a qualitative way how different of Lac mutants behave. In this problem we will use some simple assumptions to develop a more quantitative description of Lac gene expression. Say that a wild type Lac+ *E. coli* strain produces <1 unit of ß-galactosidase when no inducer is present and 100 units of enzyme when an inducer such as IPTG is present. In addition assume that for merodiploids that carry two copies of the Lac operon that the total ß-galactosidase is the sum of the enzyme expressed from each operon:

	ß-galactosidase activity	
	– IPTG	+ IPTG
Lac I^+ O^+ Z^+ Y^+ A^+	<1 unit	100 units
Lac I^+ O^+ Z^+ Y^+ A^+ / F' Lac I^+ O^+ Z^+ Y^+ A^+	<1 unit	200 units

Further assume that the amount of ß-galactosidase expressed is inversely proportional to the activity of the Lac repressor protein. Thus a mutant in the promoter for the Lac I gene that expresses half of the amount of repressor protein (call this allele Lac I-↓) will only give half the level of repression:

	ß-galactosidase activity	
	– IPTG	+ IPTG
Lac I-↓ O^+ Z^+ Y^+ A^+	50 unit	100 units
Lac I-↓ O^+ Z^+ Y^+ A^+ / F' Lac I^+ O^+ Z^+ Y^+ A^+	<1 unit	200 units

a) Consider a Lac I^{-d} allele that interferes with the repressor headpiece binding to DNA but can still oligomerize. Assume that the subunits in a repressor tetramer mix at random and that a tetramer with one $LacI^{-d}$ subunit has half the activity as a wild type tetramer and that tetramers with two or more $LacI^{-d}$ subunits have no activity. Given these assumptions fill in the table below with the expected levels of ß-galactosidase activity.

	ß-galactosidase activity	
	– IPTG	+ IPTG
Lac I^{-d} O^+ Z^+ Y^+ A^+		
Lac I^{-d} O^+ Z^+ Y^+ A^+ / F' Lac I^+ O^+ Z^+ Y^+ A^+		

b) Now consider what would happen if you combined a Lac I^s allele with a $LacI^{-d}$ allele. Remember that a Lac I^s mutation locks the repressor in a conformation where it binds tightly to the operator site regardless of whether inducer is present. Fill in the table below for this double mutant designated Lac I^{s-d}.

	ß-galactosidase activity	
	– IPTG	+ IPTG
Lac I^{s-d} O^+ Z^+ Y^+ A^+		
Lac I^{s-d} O^+ Z^+ Y^+ A^+ / F' Lac I^+ O^+ Z^+ Y^+ A^+		

c) Next consider what would happen if you combined a Lac I-↓ allele with a $LacI^{-d}$ allele. This double mutant, designated $LacI^{-d}$ -↓ should express $LacI^{-d}$ protein at half the level as wild type. Fill in the table below.

	ß-galactosidase activity	
	– IPTG	+ IPTG
$LacI^{-d}$ -↓ O^+ Z^+ Y^+ A^+		
$LacI^{-d}$ -↓ O^+ Z^+ Y^+ A^+ / F' Lac I^+ O^+ Z^+ Y^+ A^+		

d) Now you isolate a mutant in the promoter for the LacI gene that increases the amount of repressor protein ten-fold. For the purpose of this problem we will designate this allele is Lac I-↑, although in real life such alleles are called $LacI^q$. Consider what would happen if you combined a Lac I-↑ allele with a $LacI^{-d}$ allele. This double mutant, designated $LacI^{-d}$ -↑ should express $LacI^{-d}$ protein at ten times the level as wild type. By filling in the table below and comparing the results with **part c)** you should see a good example of how the degree to which an allele is dominant depends on the level of expression.

	ß-galactosidase activity	
	– IPTG	+ IPTG
$LacI^{-d}$ -↑ O^+ Z^+ Y^+ A^+		
$LacI^{-d}$ -↑ O^+ Z^+ Y^+ A^+ / F' Lac I^+ O^+ Z^+ Y^+ A^+		

e) Plasmid cloning vectors derived from R-factors usually are present at ten or more copies per cell. Imagine that you have cloned the Lac operon (without the LacI gene) into the vector pBR322. When this plasmid (pBR322 Lac O^+ Z^+ Y^+ A^+) is in a wild type strain, you find to your surprise that although the operon contains an intact promoter and operator, the LacZ, LacY and LacA are not repressed properly. However when this plasmid is in a strain that carries a Lac I-↑ allele in the chromosome nearly normal regulation is restored.

	ß-galactosidase activity	
	– IPTG	+ IPTG
Lac I^+ O^+ Z^+ Y^+ A^+ / pBR322 Lac O^+ Z^+ Y^+ A^+	900 units	1000 units
LacI-↑ O^+ Z^+ Y^+ A^+ / pBR322 Lac O^+ Z^+ Y^+ A^+	2 units	1000 units

Explain in simple qualitative terms why Lac I-↑ restores normal regulation to the Lac operon on a plasmid.

2. You are studying a new strain of *E. coli* that can utilize the disaccharide sucrose efficiently. You find that utilization depends on the enzyme sucrase, which is encoded by the gene Suc1. Suc1 is not expressed unless sucrose is present in the growth medium.

a) You have isolated two mutations that prevent expression of sucrase, which you designate SucA$^-$and SucB$^-$. P1 phage mapping experiments using a Tn5 insertion linked to SucA$^-$ shows that the insertion is also linked to SucB$^-$, but is not linked to Suc1. You construct an F' factor that carries the SucA SucB region of the chromosome and use this F' factor to perform a variety of tests shown below:

	sucrase activity	
	– sucrose	+ sucrose
wild type (Suc1$^+$)	–	+
Suc1$^-$	–	–
SucA$^-$	–	–
SucA$^-$ / F' SucA$^+$ SucB$^+$	–	+
SucB$^-$	–	–
SucB$^-$ / F' SucA$^+$ SucB$^+$	–	+
SucA$^-$ SucB$^+$/ F' SucA$^+$ SucB$^-$	–	+

Describe the basic genetic properties of the SucA$^-$ and SucB$^-$ mutations, explaining the rationale for your conclusions, and make a proposal for the type of regulatory functions affected by the SucA$^-$ and SucB$^-$mutations.

b) Diagram two possible models for regulatory pathways for Suc1 that can explain the behavior of the SucA$^-$ and SucB$^-$ mutations. For each model include a role for the inducer sucrose. Explain why or why not double mutant analysis could be used to distinguish between the two models.

c) Next, you isolate a third mutant, SucC$^-$, which gives constitutive sucrase expression even in the absence of sucrose. The SucC$^-$ mutation is linked to the same Tn5 insertion described in **part a)** indicating that it is carried on the F' now designated F' SucA$^+$ SucB$^+$ SucC$^+$. (although you should note that we do not know the order of the SucA$^+$ SucB$^+$ and SucC$^+$ alleles). Genetic tests of the SucC$^-$ mutation yield the following:

	sucrase activity	
	– sucrose	+ sucrose
SucC$^-$	+	+
SucC$^-$ / F' SucA$^+$ SucB$^+$ SucC$^+$	+	+
SucC$^-$ / F' SucA$^-$ SucB$^+$ SucC$^+$	+	+
SucC$^-$ / F' SucA$^+$ SucB$^-$ SucC$^+$	+	+

As above, classify the SucC$^-$ mutation in terms of its basic genetic properties and explain how you arrived at your conclusions.

d) Using the linked Tn5 you carry out two different P1 transduction experiments. You grow P1 on a Tn5 SucC$^-$ host and infect a SucA$^-$ recipient, selecting for Kan^r. Among the Kan^r transductants, about 10% show normally regulated sucrase expression, while the rest show either uninducible expression or half are constitutive for sucrase expression. When you use the same P1 lysate to infect a SucB$^-$ recipient, among 1000 Kan^r transductants, about half are uninducible and half are constitutive for sucrase expression, but none show normally regulated sucrase expression. What do these linkage experiments tell you about the SucC$^-$ mutation. Be as specific as possible.

e) Finally, you construct a SucA$^-$ SucC$^-$ double mutant by P1 transduction (in real life this would not be trivial since there is no way to know a priori what the phenotype of this double mutant would be and you may want to think about how you might screen transductants for the double mutant). You find that the SucA$^-$ SucC$^-$ double mutant gives constitutive sucrase expression. Now using all of the information you have diagram the entire pathway for Suc1 regulation indicating the function of each of the elements affected by the SucA$^-$, SucB$^-$, and SucC$^-$ mutations and the inducer sucrose.

7.03 Problem Set 4

Due before 5 PM on Monday, October 30

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

1. In lecture we have seen in a qualitative way how different Lac mutants behave. In this problem we will use some simple assumptions to develop a more quantitative description of Lac gene expression. Say that a wild type Lac+ *E. coli* strain produces <1 unit of ß-galactosidase when no inducer is present and 100 units of enzyme when an inducer such as IPTG is present. In addition assume that for merodiploids that carry two copies of the Lac operon that the total ß-galactosidase is the sum of the enzyme expressed from each operon:

	ß-galactosidase activity	
	– IPTG	+ IPTG
Lac $I^+ O^+ Z^+ Y^+ A^+$	<1 unit	100 units
Lac $I^+ O^+ Z^+ Y^+ A^+$ / F' Lac $I^+ O^+ Z^+ Y^+ A^+$	<1 unit	200 units

Further assume that the amount of ß-galactosidase expressed is inversely proportional to the activity of the Lac repressor protein. Thus a mutant in the promoter for the Lac I gene that expresses half of the amount of repressor protein (call this allele Lac I-$) will only give half the level of repression:

	ß-galactosidase activity	
	– IPTG	+ IPTG
Lac I-$ $O^+ Z^+ Y^+ A^+$	50 unit	100 units
Lac I-$ $O^+ Z^+ Y^+ A^+$ / F' Lac $I^+ O^+ Z^+ Y^+ A^+$	<1 unit	200 units

a) Consider a Lac I^{-d} allele that interferes with the repressor headpiece binding to DNA but can still oligomerize. Assume that the subunits in a repressor tetramer mix at random and that a tetramer with one $LacI^{-d}$ subunit has half the activity as a wild type tetramer and that tetramers with two or more $LacI^{-d}$ subunits have no activity. Given these assumptions fill in the table below with the expected levels of ß-galactosidase activity.

	ß-galactosidase activity	
	– IPTG	+ IPTG
Lac $I^{-d} O^+ Z^+ Y^+ A^+$	**100 units**	**100 units**
Lac $I^{-d} O^+ Z^+ Y^+ A^+$ / F' Lac $I^+ O^+ Z^+ Y^+ A^+$	**~163 units**	**200 units**

Wildtype strains will always fully express ß-gal when an inducer (like IPTG) is present. When the inducer is absent, the lac repressor will (normally) bind to the operator of the lac operon and prevent expression. Dominant negative mutations in the lac repressor prevent the repressor from binding to the operator even in the absence of the inducer, due to a change in the operator binding site in the lac repressor gene product. Thus, the $LacI^{-d}$ gene product cannot repress ß-gal expression and a constitutive phenotype is observed.

For a strain containing both Lac I^{-d} and Lac I^+ in the absence of the inducer:

In this particular case, the ratio of wild type Lac I gene product to mutant Lac I is 1:1, so each subunit in a given tetrameric repressor has a 50% chance of being normal and a 50% chance of being mutant.

The probability of a tetramer having four normal subunits is $(1/2)^4 = (1/16)$

The probability of a tetramer having three normal subunits and one mutant subunit is $4*(1/2)^1*(1/2)^3 = (4/16)$ (by binomial expansion)

The problem says that tetramers with 4 normal subunits have full repression activity, tetramers with 3 normal subunits have ½ of normal repression activity, and tetramers with 2 or more mutant subunits have no repression ability. The total probability of wildtype repression is, then, (1)*(1/16) + (1/2)*(4/16) = (3/16) (the sum of each probability x the amount of wildtype activity)

So, the probability of ß-galactosidase expression is 1 – 3/16 = 13/16, and the total amount of ß-galactosidase expression is (13/16)*200 units = 162.5 units.

b) Now consider what would happen if you combined a Lac I^s allele with a $LacI^{-d}$ allele. Remember that a Lac I^s mutation locks the repressor in a conformation where it binds tightly to the operator site regardless of whether inducer is present. Fill in the table below for this double mutant designated Lac I^{s-d}.

	ß-galactosidase activity	
	– IPTG	+ IPTG
Lac I^{s-d} O^+ Z^+ Y^+ A^+	**100 units**	**100 units**
Lac I^{s-d} O^+ Z^+ Y^+ A^+ / F' Lac I^+ O^+ Z^+ Y^+ A^+	**~163 units**	**200 units**

If a Lac I^s allele is combined with a $LacI^{-d}$ allele, then the Lac repressor protein expressed from this gene will be unable to bind both the inducer and the operator. However, if the strain is unable to bind to the operator (the effect of the $LacI^{-d}$ allele), then the inducer has no real effect on the repressor. As such, a Lac I^s allele has no additional effect on a $LacI^{-d}$ allele, and expression patterns will be similar to those in part 1a).

c) Next consider what would happen if you combined a Lac I-$ allele with a $LacI^{-d}$ allele. This double mutant, designated $LacI^{-d}$ -$ should express $LacI^{-d}$ protein at half the level as wild type. Fill in the table below.

	ß-galactosidase activity	
	– IPTG	+ IPTG
$LacI^{-d}$ -$ O^+ Z^+ Y^+ A^+	**100 units**	**100 units**
$LacI^{-d}$ -$ O^+ Z^+ Y^+ A^+ / F' Lac I^+ O^+ Z^+ Y^+ A^+	**~121 units**	**200 units**

Again, the method and reasoning for this problem are equivalent to part 1a). The Lac I-$ allele just alters the mutant protein levels, which changes the probabilities.

For a strain containing both Lac I^{-d}-$ and Lac I^{+} in the absence of the inducer:

The mutant to wild-type subunit ratio is now 1:2

The probability of a tetramer having four normal subunits is now $(2/3)^4 = (16/81)$

The probability of a tetramer having three normal subunits and one mutant subunit is now
$4*(1/3)^1*(2/3)^3 = (32/81)$ (by binomial expansion)

The total probability of wildtype repression is, then, (1)*(16/81) + (1/2)*(32/81) = (32/81) (the sum of each probability x the amount of wildtype activity)

So, the probability of non-repression, or ß-galactosidase expression, is 1 – 32/81 = 49/81, and
the total amount of ß-galactosidase expression is (49/81)*200 units = 120.987654321 units.

d) Now you isolate a mutant in the promoter for the LacI gene that increases the amount of repressor protein ten-fold. For the purpose of this problem we will designate this allele is Lac I-#, although in real life such alleles are called $LacI^{q}$. Consider what would happen if you combined a Lac I-# allele with a $LacI^{-d}$ allele. This double mutant, designated $LacI^{-d}$ -# should express $LacI^{-d}$ protein at ten times the level as wild type. By filling in the table below and comparing the results with **part c)** you should see a good example of how the degree to which an allele is dominant depends on the level of expression.

	ß-galactosidase activity	
	– IPTG	+ IPTG
$LacI^{-d}$ -# O^{+} Z^{+} Y^{+} A^{+}	**100 units**	**100 units**
$LacI^{-d}$ -# O^{+} Z^{+} Y^{+} A^{+} / F' Lac I^{+} O^{+} Z^{+} Y^{+} A^{+}	**~200 units**	**200 units**

The method and reasoning for this problem are equivalent to part 1a). The Lac I-# allele also just alters the mutant protein levels, which changes the probabilities.

For a strain containing both Lac I^{-d}-# and Lac I^{+} in the absence of the inducer:

The mutant to wild-type subunit ratio is now 10:1

The probability of a tetramer having four normal subunits is $(1/11)^4 = \sim 0$

The probability of a tetramer having three normal subunits and one mutant subunit is $4*(10/11)^1*(1/11)^3$, which is also approx. 0.

Thus, there is almost no wildtype-like regulation, and even in the presence of a Lac I^{+} allele, a
$LacI^{-d}$ -# allele results in constitutive expression.

e) Plasmid cloning vectors derived from R-factors usually are present at ten or more copies per cell. Imagine that you have cloned the Lac operon (without the LacI gene) into the vector pBR322. When this plasmid (pBR322 Lac O^+ Z^+ Y^+ A^+) is in a wild type strain, you find to your surprise that although the operon contains an intact promoter and operator, the LacZ, LacY and LacA are not repressed properly. However when this plasmid is in a strain that carries a Lac I-# allele in the chromosome nearly normal regulation is restored.

	ß-galactosidase activity	
	– IPTG	+ IPTG
Lac I^+ O^+ Z^+ Y^+ A^+ / pBR322 Lac O^+ Z^+ Y^+ A^+	900 units	1000 units
LacI-# O^+ Z^+ Y^+ A^+ / pBR322 Lac O^+ Z^+ Y^+ A^+	2 units	1000 units

Explain in simple qualitative terms why Lac I-# restores normal regulation to the Lac operon on a plasmid.

Essentially, the normal levels of the LacI gene product (the Lac Repressor) are insufficient to repress every copy of the Lac operon present in the cell, because there are approximately ten copies of the plasmid in each cell. However, the Lac-# allele, which results in ten times as much of the repressor gene product, provides sufficient capability to repress the approximately ten copies of the Lac operon present in the cell.

2. You are studying a new strain of *E. coli* that can utilize the disaccharide sucrose efficiently. You find that utilization depends on the enzyme sucrase, which is encoded by the gene Suc1. Suc1 is not expressed unless sucrose is present in the growth medium.

a) You have isolated two mutations that prevent expression of sucrase, which you designate SucA$^-$and SucB$^-$. P1 phage mapping experiments using a Tn5 insertion linked to SucA$^-$ shows that the insertion is also linked to SucB$^-$, but is not linked to Suc1. You construct an F' factor that carries the SucA SucB region of the chromosome and use this F' factor to perform a variety of tests shown below:

	sucrase activity		
	– sucrose	+ sucrose	**interpretation**
wild type (Suc1$^+$)	–	+	**sucrose is an inducer of the system**
Suc1$^-$	–	–	**Suc1- is uninducible**
SucA$^-$	–	–	**SucA- is uninducible**
SucA$^-$ / F' SucA$^+$ SucB$^+$	–	+	**SucA- is recessive**
SucB$^-$	–	–	**SucB- is uninducible**
SucB$^-$ / F' SucA$^+$ SucB$^+$	–	+	**SucB- is recessive**
SucA$^-$ SucB$^+$/ F' SucA$^+$ SucB$^-$	–	+	**SucA- and SucB- complement**

Describe the basic genetic properties of the SucA$^-$ and SucB$^-$ mutations, explaining the rationale for your conclusions, and make a proposal for the type of regulatory functions affected by the SucA$^-$ and SucB$^-$mutations.

SucA-: recessive, uninducible, trans-acting (positive regulator)
Because there is no sucrase activity in the presence or absence of sucrose in the SucA- mutant, we can conclude that the SucA- mutation produces an uninducible phenotype. That is, the system cannot be turned on in the presence or absence of inducer.

We know that SucA- is recessive based on the merodiploid, SucA$^-$ / F' SucA$^+$ SucB$^+$, which shows a phenotype of normal regulation. In this merodiploid, a wild-type copy of SucA is sufficient to restore normal regulation of Suc1 expression.

Finally, we can reason that SucA is trans-acting based on linkage analysis. Because the Tn5 insertion is linked to SucA but is unlinked to Suc1, we can conclude that SucA is unlinked to Suc1 as well. As a result, SucA must be a trans-acting factor.

Based on the conclusions above, SucA must encode for a net activator (positive regulator). The SucA- mutation must be a LOF mutation in the activator, resulting in an uninducible phenotype as is observed experimentally.

SucB-: recessive, uninducible, trans-acting (positive regulator)
Based on the same analysis as above, we can conclude that SucB- is recessive, uninducible, and trans-acting. As a result, SucB must also encode for a positive regulator.

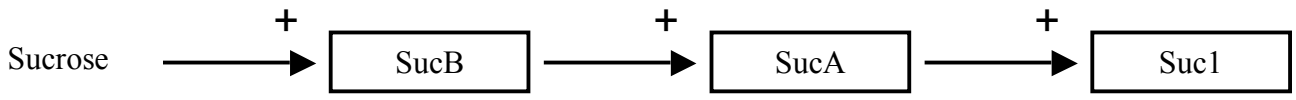
SucA- and SucB- are mutations in different genes:
Because SucA- and SucB- are recessive mutations, we can perform complementation analysis on the merodiploid, SucA$^-$ SucB$^+$/ F' SucA$^+$ SucB$^-$. Because this merodiploid strain displays a wild-type phenotype, SucA- and SucB- complement and are thus mutations in different genes.

b) Diagram two possible models for regulatory pathways for Suc1 that can explain the behavior of the SucA$^-$ and SucB$^-$ mutations. For each model include a role for the inducer sucrose. Explain why or why not double mutant analysis could be used to distinguish between the two models.

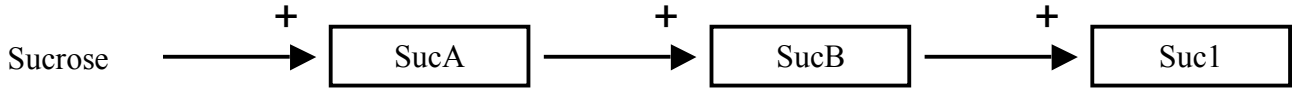
Based on the information provided in part (a), we know that SucA and SucB are both positive regulators. In addition, we know that sucrose is an inducer of the system.

It is important to note that we do not know the relative order of SucA and SucB. As a result, we can draw two possible models for the Suc1 regulatory pathway.

Model #1:



Model #2:



We cannot use double mutant analysis to distinguish between these two models because the SucA- and SucB- mutations have the same phenotype. In order to do epistasis analysis, the two mutations you are examining must have different phenotypes.

c) Next, you isolate a third mutant, SucC$^-$, which gives constitutive sucrase expression even in the absence of sucrose. The SucC$^-$ mutation is linked to the same Tn5 insertion described in **part a)** indicating that it is carried on the F' now designated F' SucA$^+$ SucB$^+$ SucC$^+$. (although you should note that we do not know the order of the SucA$^+$ SucB$^+$ and SucC$^+$ alleles). Genetic tests of the SucC$^-$ mutation yield the following:

	sucrase activity		
	– sucrose	+ sucrose	**interpretation**
SucC$^-$	+	+	**SucC- is constitutive**
SucC$^-$ / F' SucA$^+$ SucB$^+$ SucC$^+$	+	+	**SucC- is dominant**
SucC$^-$ / F' SucA$^-$ SucB$^+$ SucC$^+$	+	+	
SucC$^-$ / F' SucA$^+$ SucB$^-$ SucC$^+$	+	+	

As above, classify the SucC$^-$ mutation in terms of its basic genetic properties and explain how you arrived at your conclusions.

As indicated above, SucC- is a dominant mutation that produces a constitutive phenotype. We can conclude that SucC is a trans-acting factor because it is unlinked to Suc1. Based on this information, SucC- may be a mutation that results in a super-activator or a dominant-negative repressor. SucC may encode for a positive regulator or a negative regulator.

Complementation analysis cannot be employed here because SucC- is dominant.

d) Using the linked Tn5 you carry out two different P1 transduction experiments. You grow P1 on a Tn5 SucC$^-$ host and infect a SucA$^-$ recipient, selecting for Kan^r. Among the Kan^r transductants, about 10% show normally regulated sucrase expression, while the rest show either uninducible expression or half are constitutive for sucrase expression. When you use the same P1 lysate to infect a SucB$^-$ recipient, among 1000 Kan^r transductants, about half are uninducible and half are constitutive for sucrase expression, but none show normally regulated sucrase expression. What do these linkage experiments tell you about the SucC$^-$ mutation. Be as specific as possible.

SucC- encodes for a super-activator.

Cross #1:
In order to have a normally regulated Kan^r transductant, a crossover event must occur between SucA and SucC during homologous recombination between the host and recipient DNA. This occurs 10% of the time, indicating that SucA- and SucB- are not tightly linked. Based on the complementation analysis in part (a), we know that SucA and SucB are different genes.

Cross #2:

In order to have a normally-regulated Kan^r transductant, a crossover event must occur between SucB and SucC. No normally regulated transductants are observed. As a result, SucB- and SucC- must be tightly linked such that a crossover event between the two is extremely unlikely. The most likely scenario is that SucB- and SucC- are mutations in the same gene.

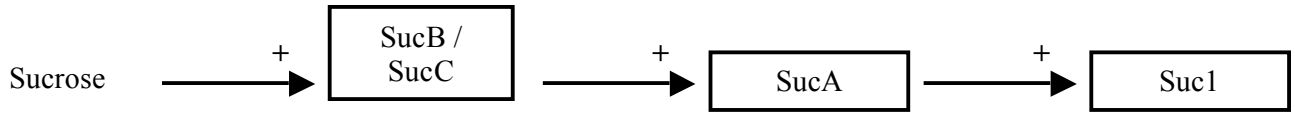
In part (a), we determined that SucB encodes for a net activator (positive regulator). Because the SucC- mutation is in the same gene as SucB-, we can conclude that SucC- encodes for a super-activator.

e) Finally, you construct a SucA$^-$ SucC$^-$ double mutant by P1 transduction (in real life this would not be trivial since there is no way to know a priori what the phenotype of this double mutant would be and you may want to think about how you might screen transductants for the double mutant). You find that the SucA$^-$ SucC$^-$ double mutant gives constitutive sucrase expression. Now using all of the information you have diagram the entire pathway for Suc1 regulation indicating the function of each of the elements affected by the SucA$^-$, SucB$^-$, and SucC$^-$ mutations and the inducer sucrose.

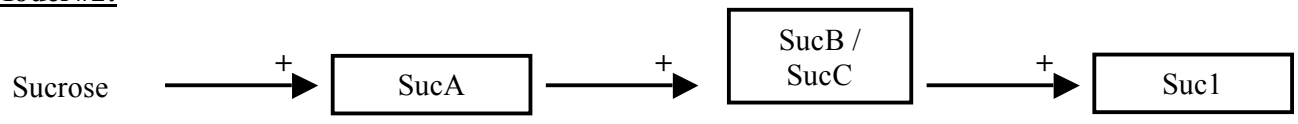
There are two possible models to consider:

The double mutant, SucA- SucC-, gives constitutive sucrase expression. In epistasis analysis, the phenotype of the double mutant matches the phenotype of the downstream mutation. In this case, because SucC- gives constitutive sucrase expression, SucC must be downstream of SucA. Model #2 is correct.

Model #1:



Model #2:



7.03 PROBLEM SET 5

BASED ON LECTURES 20-25
DUE BEFORE 5PM ON WEDNESDAY, NOVEMBER 15
SUBMIT ANSWERS DURING RECITATION OR PLACE IN BOX OUTSIDE OF THE BIOLOGY EDUCATION OFFICE

1) Genetic pathways in eukaryotes often are investigated using gene fusions. This approach could be used in yeast to study regulation of a detoxification gene, DTX1. This gene encodes an enzyme that neutralizes benzene, which is a known carcinogen. To investigate DTX1 regulation, you made a gene fusion (P_{DTX1} - LacZ) that consists of the cis regulatory region of DTX1 and the coding region of LacZ. When integrated into the yeast genome, this gene fusion shows wild-type expression. In addition, you isolated two recessive loss-of-function mutations, dtx2- and dtx3-, which show un-inducible gene fusion expression. Your analysis shows that the dtx2- and dtx3- mutations reside in different genes, are not linked to each other, and are not linked to the gene fusion.

a) Diagram three possible models that illustrate the wild-type regulatory relationships among DTX2 and DTX3 and P_{DTX1} - LacZ.

b) You have access to a dominant allele of the DTX2 gene, dtx2-^{D}, which causes constitutive expression of P_{DTX1} - LacZ. Describe the experiment you would perform to distinguish between two of the three models diagrammed in **part (a)**. Show all crosses, the resulting tetrads, and how each result should be interpreted.

c) Now you want to investigate the mechanism by which DTX2 and DTX3 act in the pathway. To determine if DTX2 and DTX3 can bind to each other, you decide to perform a yeast-two-hybrid assay. Which cis-regulatory regions would you put upstream of LacZ?

Suppose you engineer the protein fusion genes DTX2 : AD and DTX3 : DB on a selectable plasmid. What results from the yeast-two-hybrid assay would show that DTX2 and DTX3 interact/bind directly to each other? Complete the chart below and include the necessary controls.

Assay	Yeast strain	Reporter gene expression (expected)
Control 1		None
Control 2	DTX2 : AD /	
Control 3	DTX3 : DB /	
Experiment 1 (assume no direct interaction)		
Experiment 2 (assume direct interaction)		

d) What if one of the controls listed above shows reporter gene expression? For each control, suggest a possibility as to what is occurring in the cell if reporter gene expression is observed.

Control 1 ⟶

Control 2 ⟶

Control 3 ⟶

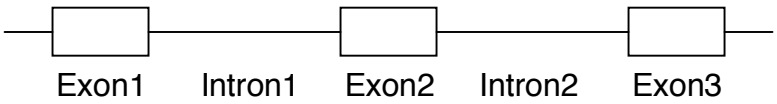
e) Based on the results from the yeast-two-hybrid assay, what can be concluded about your models in **part a**?

2) Genes in the body that suppress the development of cancer are known as tumor suppressor genes. Due to its role in the repair of double-strand breaks in DNA, REP1 is regarded as a tumor suppressor gene. Individuals with loss-of-function mutations in both REP1 alleles show a much greater risk for developing some cancers. After brainstorming with colleagues, it is decided that the role of REP1 in cancer development could be investigated effectively using genetically modified mice.

a) In order to obtain an effective mouse model to study the role of REP1 in cancer development, what genotype would you generate? Explain.

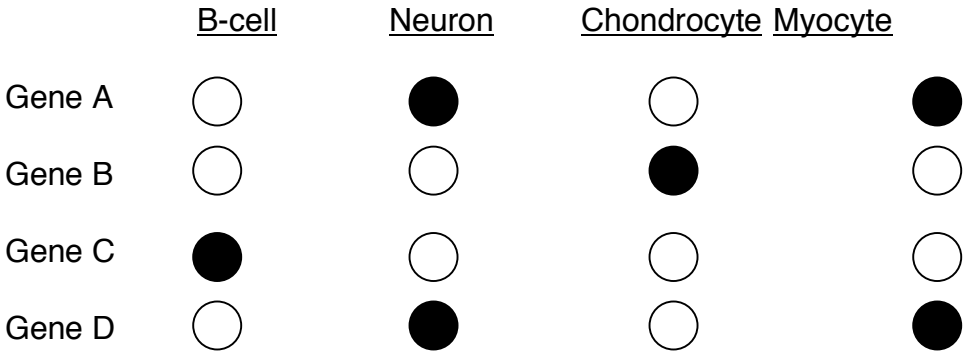
b) Would pronuclear injection or gene targeting techniques be required to construct the desired mouse model? Explain.

c) Exon 2 is essential for the tumor-suppressor activity of the REP1 gene (illustrated below). Draw the DNA construct you would use to modify the mouse genome. How would it integrate into the genome? How would you detect its integration?



A number of genes, such as gene **X**, are involved in the normal regulation of cell growth. Oncogenes are mutated versions (alleles) of these genes that promote unregulated cell growth and lead to cancer. In one type of cancer, B-cell lymphoma, gene **X** is highly expressed. You hypothesize that the over-expression of gene **X** is the causative mutation of B-cell lymphoma.

The expression data illustrated below were generated using microarray analysis (dark circles signify unusually high expression and clear circles signify no expression).



d) Given these results, propose a strategy to generate a mouse model to investigate the role of gene **X** in B-cell lymphoma. Be sure to include the genotype you would create, whether pronuclear injection or gene targeting methods would be used, and how gene **X** would be expressed in B-cells.

3) PhiP and IQ are heterocyclic amines that are mammary gland carcinogens in mice. Both of these chemicals are present in certain food products such as cooked meats. To better understand the biology behind the carcinogenic properties of PhiP and IQ, we would like to identify genes that protect cells from their toxicity.

Wildtype yeast grow at a reduced rate in the presence of 50 mM PhiP but arrest completely in the presence of 100 mM PhiP.

a) Using the yeast *Saccharomyces cerevisiae*, design a screen to isolate mutants that are hypersensitive to PhiP. Be as specific as possible.

From this screen you identify two mutants that are hypersensitive to PhiP. You name these mutants Mut1 and Mut2. You find that Mut1 and Mut2 both confer a recessive mutant phenotype.

b) Design an experiment to determine if Mut1 and Mut2 are alleles of the same gene. Be as specific as possible.

You determine that Mut1 and Mut2 define two different genes. Furthermore, you have mapped and cloned Mut1. Now that you know the sequence of the MUT1 gene you decide to make a gene fusion consisting of the cis regulatory region of MUT1 ligated to the LacZ coding sequence. A single copy of this gene fusion (P_{MUT1}- LacZ) then is incorporated into Chromosome III of the yeast genome. You perform the following experiments and test for ß-galactosidase activity:

Genotype	no PhiP	25 mM PhiP
P_{MUT1}- LacZ	-	+
Mut1; P_{MUT1}- LacZ	-	+
Mut2; P_{MUT1}- LacZ	-	-

c) What do these results tell you about Mut1 and Mut2? Why is this experiment performed with 25mM PhiP and not 50 mM PhiP?

You want to identify a mutant that constitutively expresses P_{MUT1}- LacZ. You mutagenize your P_{MUT1}- LacZ haploid strain and look for ß-galactosidase activity in the absence of PhiP. You identify a mutant that you call Mut3.

Genotype	no PhiP	50 mM PhiP
P_{MUT1}- LacZ	-	+
Mut3; P_{MUT1}- LacZ	+	+

You cross this mutant to a wildtype strain (no P_{MUT1}- LacZ) and do the following experiments:

Genotype	no PhiP	50 mM PhiP
P_{MUT1}- LacZ/ +	-	+
Mut3/+; P_{MUT1}- LacZ/+	-	+

d) What conclusions can you draw from the results of the above experiments? Be as specific as possible.

e) Given the available data, draw the two most likely pathways that illustrate the involvement of PhiP, Mut2, and Mut3 in the regulation of MUT1. Assume the regulatory genes operate in series.

You cross the Mut2 strain to the Mut3 strain (both containing the P_{MUT1}- LacZ reporter), sporulate the diploid, and then measure the ß-galactosidase activity of the resulting spores grown on three different types of media. You examine 50 tetrads and observe the following:

# of tetrads	no PhiP	25 mM PhiP	50 mM PhiP
49	+	+	+
	+	+	+
	-	-	Arrest
	-	-	Arrest
1	+	+	+
	+	+	+
	-	-	Arrest
	-	+	+

f) Based on these data, what can you conclude about the relative positions of MUT2 and MUT3 in the pathway? Be as specific as possible.

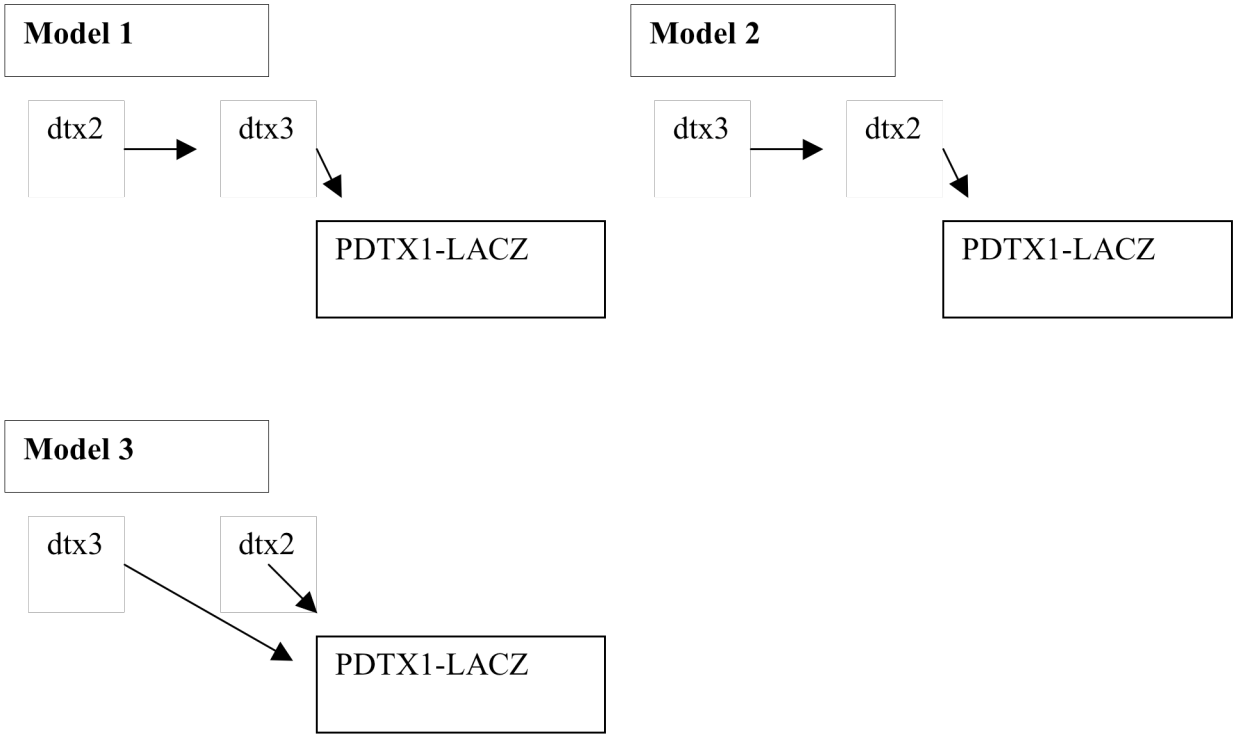
g) Which of the two models drawn in **part e** must be correct?

7.03 PROBLEM SET 5 KEY

BASED ON LECTURES 20-25
DUE BEFORE 5PM ON WEDNESDAY, NOVEMBER 15
SUBMIT ANSWERS DURING RECITATION OR PLACE IN BOX OUTSIDE OF THE BIOLOGY EDUCATION OFFICE

1) Genetic pathways in eukaryotes often are investigated using gene fusions. This approach could be used in yeast to study regulation of a detoxification gene, DTX1. This gene encodes an enzyme that neutralizes benzene, which is a known carcinogen. To investigate DTX1 regulation, you made a gene fusion (P_{DTX1} - LacZ) that consists of the cis regulatory region of DTX1 and the coding region of LacZ. When integrated into the yeast genome, this gene fusion shows wild-type expression. In addition, you isolated two recessive loss-of-function mutations, dtx2- and dtx3-, which show un-inducible gene fusion expression. Your analysis shows that the dtx2- and dtx3- mutations reside in different genes, are not linked to each other, and are not linked to the gene fusion.

a) Diagram three possible models that illustrate the wild-type regulatory relationships among DTX2 and DTX3 and P_{DTX1} - LacZ.



These are the three simplest models.

b) You have access to a dominant allele of the DTX2 gene, dtx2-^{D}, which causes constitutive expression of P_{DTX1} - LacZ. Describe the experiment you would perform to distinguish between two of the three models diagrammed in **part (a)**. Show all crosses, the resulting tetrads, and how each result should be interpreted.

Cross dtx2^{D} to dtx3-. The resulting diploid will be heterozygous for both dtx2-D and dtx3-. Sporulate this diploid to get the following tetrads listed below.

	Genotypes
PD	**2 uninducible, 2 constitutive**
NPD	**2 wild-type, 2 double mutant (either constitutive or uninducible)**
TT	**1 uninducible, 1 constitutive, 1 wild-type, 1 double mutant (either constitutive or uninducible)**

Take the double mutant spore from the NPD tetrad.

- **If PDTX1-LacZ is constitutively expressed in the double mutant, then discard model 1.**
- **If PDTX2-LacZ is uninducible, then discard model 2.**
- **This technique does not address parallel pathway models.**

c) Now you want to investigate the mechanism by which DTX2 and DTX3 act in the pathway. To determine if DTX2 and DTX3 can bind to each other, you decide to perform a yeast-two-hybrid assay. Which cis-regulatory regions would you put upstream of LacZ?

You would place the promoter from the Gal1 sequence in front of LacZ. The UAS sequence is essential as it is the sequence in the Gal1 promoter to which the DNA binding domain of Gal4 binds.

Suppose you engineer the protein fusion genes DTX2 : AD and DTX3 : DB on a selectable plasmid. What results from the yeast-two-hybrid assay would show that DTX2 and DTX3 interact/bind directly to each other? Complete the chart below and include the necessary controls.

Assay	Yeast strain	Reporter gene expression (expected)
Control 1	**Pgal1::LacZ**	None
Control 2	DTX2 : AD / **Pgal1::LacZ**	**None**
Control 3	DTX3 : DB / **Pgal1::LacZ**	**None**
Experiment 1 (assume no direct interaction)	**Pgal1::LacZ + DTX2:AD + DTX3:DB**	**None**
Experiment 2 (assume direct interaction)	**Pgal1::LacZ + DTX2:AD + DTX3:DB**	**++++**

The controls outlined in the table above are essential to do this experiment. There are other possible controls. For example, switch AD and DB so that DTX2 and DTX3 are attached to different domains and use this to confirm the interaction again.

d) What if one of the controls listed above shows reporter gene expression? For each control, suggest a possibility as to what is occurring in the cell if reporter gene expression is observed.

There are several possible answers for "part d". The purpose of the question was to get you to think about how the assay works and why controls are essential.

Control 1 ⟶

There is something activating expression in a strain with only the fusion gene. This strain will not be useful; it is necessary to integrate the reporter gene somewhere else, or use a different plasmid.

Control 2 ⟶

DTX2:AD is activating reporter gene expression without the DNA binding domain of Gal4. DTX2 may bind directly to DNA close enough to the reporter fusion gene to activate its expression.

Control 3 ⟶

DTX3:BD is recruiting RNA Polymerase or is recruiting other factors that promote LacZ expression without the activation domain.

e) Based on the results from the yeast-two-hybrid assay, what can be concluded about your models in **part a**?

If the proteins do interact with each other in the yeast two hybrid assay, it suggests that the proteins may interact with each other when they promote DTX1 expression. If the proteins interact, it is strong evidence that they may act in a pathway in series instead of in parallel.

If you find they do not interact, new information will be necessary.

2) Genes in the body that suppress the development of cancer are known as tumor suppressor genes. Due to its role in the repair of double-strand breaks in DNA, REP1 is regarded as a tumor suppressor gene. Individuals with loss-of-function mutations in both REP1 alleles show a much greater risk for developing some cancers. After brainstorming with colleagues, it is decided that the role of REP1 in cancer development could be investigated effectively using genetically modified mice.

a) In order to obtain an effective mouse model to study the role of REP1 in cancer development, what genotype would you generate? Explain.

REP1-/-

To model this loss-of-function mutation, we must generate a mouse that lacks two functional copies of REP1. Based on REP1's known role as a tumor suppressor, we would expect REP1-/- mice to show a greater risk of developing certain cancers.

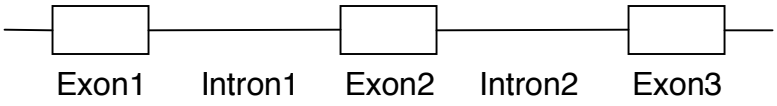
(NOTE: In certain cases, generating a homozygous mutant as we've done here is not possible. If a gene is serving an essential role, then losing the function of both of its alleles will produce unviable offspring. Here, since we are told that REP1-/- individuals are at a greater risk of developing cancers, we know that homozygous REP1 mutants are viable.)

b) Would pronuclear injection or gene targeting techniques be required to construct the desired mouse model? Explain.

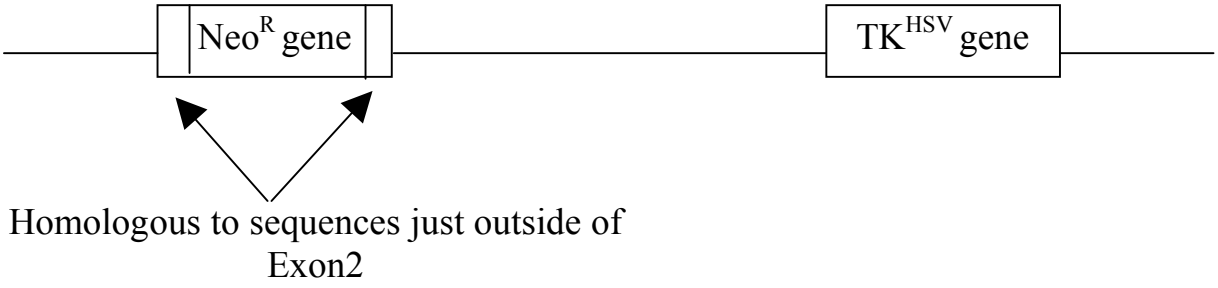
Gene targeting.

You hypothesize that the increased risk of cancer is due to two inactive versions of REP1. To model this situation in a mouse, you need to knock-out both functional copies of the REP1 gene. Gene targeting is the only method to do this.

c) Exon 2 is essential for the tumor-suppressor activity of the REP1 gene (illustrated below). Draw the DNA construct you would use to modify the mouse genome. How would it integrate into the genome? How would you detect its integration?



DNA construct:

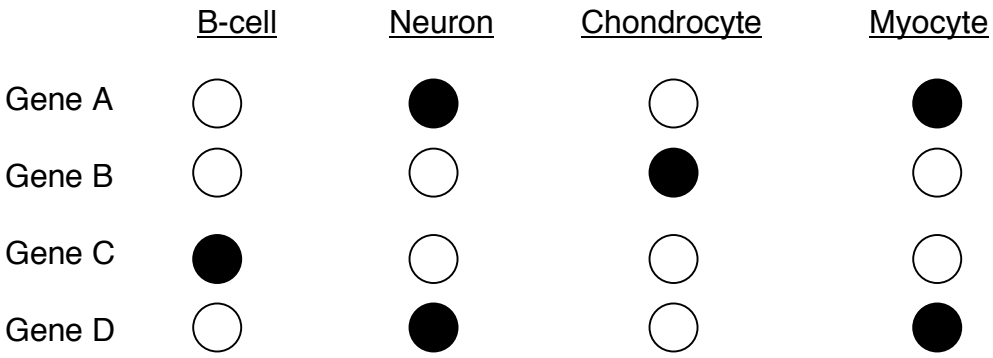


The DNA construct contains two regions that are homologous to the mouse DNA surrounding Exon 2, separated by a gene that confers antibiotic resistance (e.g., Neomycin). The construct will homologously recombine with the mouse genomic DNA at the homologous sequences and replace Exon 2 with the Neo^R gene. We can use this newly acquired drug resistance gene to positively select for ES cells that have undergone homologous recombination.

We additionally need to include a TK gene outside of the homologous sequences. This is used as a negative selection marker against integration events that occurred via *non*-homologous recombination. If the DNA construct integrates randomly using non-homologous end-joining, then the TK gene also will be integrated and the cells will be selected against with a drug (review lecture notes).

A number of genes, such as gene **X**, are involved in the normal regulation of cell growth. Oncogenes are mutated versions (alleles) of these genes that promote unregulated cell growth and lead to cancer. In one type of cancer, B-cell lymphoma, gene **X** is highly expressed. You hypothesize that the over-expression of gene **X** is the causative mutation of B-cell lymphoma.

The expression data illustrated below were generated using microarray analysis (dark circles signify unusually high expression and clear circles signify no expression).



d) Given these results, propose a strategy to generate a mouse model to investigate the role of gene **X** in B-cell lymphoma. Be sure to include the genotype you would create, whether pronuclear injection or gene targeting methods would be used, and how gene **X** would be expressed in B-cells.

We need to create a translational fusion protein consisting of the promoter region of Gene C and the coding region of Gene X (P_{GeneC}- GeneX). Gene C was chosen as it shows high levels of expression in B-cells. To add this DNA construct into the mice genome, the simplest method is pronuclear injection. After injecting into the male pronucleus of a fertilized egg, the construct will insert randomly into the mouse genome.

Every cell in the developed mouse will now possess this fusion gene, but Gene X will be expressed in a cell-type dependent manner because of its upstream promoter from Gene C. Namely, X will be highly expressed in B-cells, but not expressed in neurons, chondrocytes, and myocytes (See 3rd row of microarray analysis above).

We have created a transgenic mouse that over-expresses Gene X selectively in B cells. If over-expression of Gene X is a causative mutation of B-cell lymphoma as we have hypothesized, we would expect these transgenic mice to develop B-cell lymphoma.

3) PhiP and IQ are heterocyclic amines that are mammary gland carcinogens in mice. Both of these chemicals are present in certain food products such as cooked meats. To better understand the biology behind the carcinogenic properties of PhiP and IQ, we would like to identify genes that protect cells from their toxicity.

Wildtype yeast grow at a reduced rate in the presence of 50 mM PhiP but arrest completely in the presence of 100 mM PhiP.

a) Using the yeast *Saccharomyces cerevisiae*, design a screen to isolate mutants that are hypersensitive to PhiP. Be as specific as possible.

Mutate *S. cerevisiae* using a mutagen (example: EMS or UV light) and culture them on plates containing 0 mM PhiP. Replica plate these colonies onto plates with 50mM PhiP. Look for colonies that grow on 0, but not on 50mM PhiP. These colonies are hypersensitive to PhiP (Wild-type yeast still grow at 50mM, but not at 100mM).

From this screen you identify two mutants that are hypersensitive to PhiP. You name these mutants Mut1 and Mut2. You find that Mut1 and Mut2 both confer a recessive mutant phenotype.

b) Design an experiment to determine if Mut1 and Mut2 are alleles of the same gene. Be as specific as possible.

Cross the Mut1 and Mut2 haploid strains to produce a diploid. Plate the diploid on 50mM PhiP. If it does not grow on 50mM PhiP, then Mut1 and Mut2 are alleles of the same gene. If it does grow, then it has the wild-type phenotype, complementation occurred, and the mutations reside in different genes.

You determine that Mut1 and Mut2 define two different genes. Furthermore, you have mapped and cloned Mut1. Now that you know the sequence of the MUT1 gene you decide to make a gene fusion consisting of the cis regulatory region of MUT1 ligated to the LacZ coding sequence. A single copy of this gene fusion (P_{MUT1}- LacZ) then is incorporated into Chromosome III of the yeast genome. You perform the following experiments and test for ß-galactosidase activity:

Genotype	no PhiP	25 mM PhiP
P_{MUT1}- LacZ	-	+
Mut1; P_{MUT1}- LacZ	-	+
Mut2; P_{MUT1}- LacZ	-	-

c) What do these results tell you about Mut1 and Mut2? Why is this experiment performed with 25mM PhiP and not 50 mM PhiP?

Mut2 is a positive regulator of Mut1 expression. The fusion gene is used to monitor Mut1 expression, which seems to be induced by PhiP. These results show that Mut2 causes uninducible expression of the fusion gene, it acts in trans (it's in a different gene), and it is recessive.

25 mM of PhiP is required because at 50 mM PhiP the colonies from Mut1 and Mut2 strains would not grow.

You want to identify a mutant that constitutively expresses P_{MUT1}- LacZ. You mutagenize your P_{MUT1}- LacZ haploid strain and look for ß-galactosidase activity in the absence of PhiP. You identify a mutant that you call Mut3.

Genotype	no PhiP	50 mM PhiP
P_{MUT1}- LacZ	-	+
Mut3; P_{MUT1}- LacZ	+	+

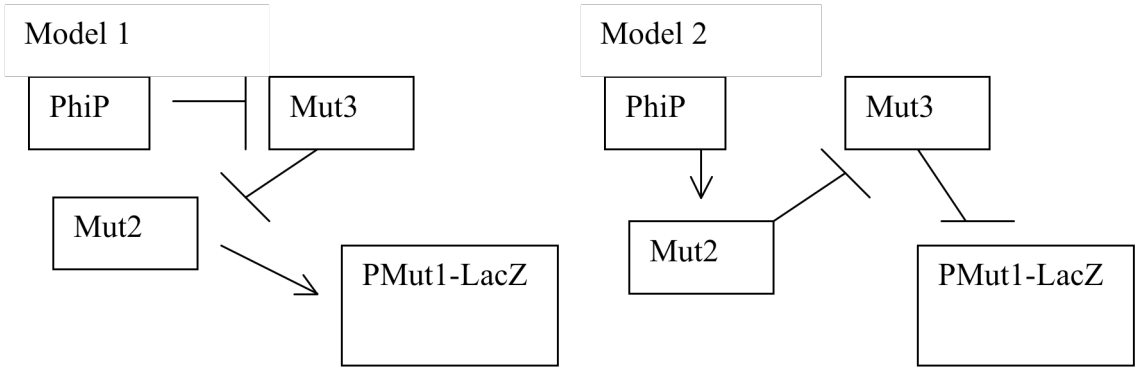
You cross this mutant to a wildtype strain (no P_{MUT1}- LacZ) and do the following experiments:

Genotype	no PhiP	50 mM PhiP
P_{MUT1}- LacZ/ +	-	+
Mut3/+; P_{MUT1}- LacZ/+	-	+

d) What conclusions can you draw from the results of the above experiments? Be as specific as possible.

You know that Mut3 has constitutive expression of the fusion gene based on the results in the 1st table. In the second cross, you find that the Mut3 phenotype is recessive. Therefore, it is a negative regulator of Mut1 expression.

e) Given the available data, draw the two most likely pathways that illustrate the involvement of PhiP, Mut2, and Mut3 in the regulation of MUT1. Assume the regulatory genes operate in series.



You cross the Mut2 strain to the Mut3 strain (both containing the P_{MUT1}- LacZ reporter), sporulate the diploid, and then measure the ß-galactosidase activity of the resulting spores grown on three different types of media. You examine 50 tetrads and observe the following:

# of tetrads		no PhiP	25 mM PhiP	50 mM PhiP
49	**3-**	+	+	+
	3-	+	+	+
	2-	-	-	Arrest
	2-	-	-	Arrest
1	**double mutant**	+	+	+
	3-	+	+	+
	2-	-	-	Arrest
	wt	-	+	+

f) Based on these data, what can you conclude about the relative positions of MUT2 and MUT3 in the pathway? Be as specific as possible.

Because two 2 types (PD and TT) of tetrads are produced, MUT2 and MUT3 must be closely linked. The 49 PD tetrads show two spores that are constitutive at all amounts of PhiP and two spores that are uninducible. In the TT tetrad, one spore must be 2-, one must be 3-, one must be wild-type, and one must be the double mutant. You find that two of the spores are constitutive. Therefore, the double mutant is constitutive and MUT3 acts further downstream in the pathway then MUT2.

g) Which of the two models drawn in **part e** must be correct?

Model 2 must be correct because MUT3 is downstream of MUT2 in model 2.

7.03 Problem Set 6

Due before 5 PM on Monday, November 27

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

1. a) Imagine a continent that has an indigenous population that has allelic variation of a gene that determines the ability of the body to store fat. Before modern times when food was scarce, a relatively rare allele (known as the "thrifty" allele) gives a heterozygous advantage of 2%. Individuals that are homozygous for the thrifty allele, because of health problems such as obesity and diabetes, have a fitness of 0.4. Calculate the expected frequency for the thrifty allele in this population.

b) Explain why the rate of new mutations for the thrifty allele is not relevant for this calculation.

c) Now consider the same continent in modern times in which the population can be thought of having two parts: 10% of the population comes from the indigenous people described above, and 90% of the population has immigrated from Europe where the thrifty allele is so rare that its frequency is effectively 0. Modern high calorie, high fat diet individuals who are homozygous for the thrifty allele are considered to have an inherited obesity related disease. Assuming random mating of the two populations, calculate the frequency of inherited obesity on the continent.

d) What would the frequency of inherited obesity on the continent be if mating between individuals were completely assortative (i.e. no mixing between the immigrant and indigenous populations)?

2. In this problem we will derive general expressions for two of the more practical results from human population genetics – X-linked recessive traits occur more often in males than females and individuals with rare recessive traits often have parents who are related to one another.

a) Derive an expression, as a function of allele frequency (q), for the ratio of affected males to affected females for a X-linked recessive trait.

b) Derive an expression, as a function of allele frequency (q), for the probability that an individual with a rare recessive trait will have parents who are first cousins relative to the probability of first cousin parents in the general population. Assume the *a priori* probability of parents who are first cousins is 0.005. You only need to derive a formula accurate for $q < 10^{-2}$ – make any reasonable simplifying approximations that you need.

7.03 Problem Set 6

Due before 5 PM on Monday, November 27
Hand in answers in recitation section or in the box outside of 68-120

1. a) Imagine a continent that has an indigenous population that has allelic variation of a gene that determines the ability of the body to store fat. Before modern times when food was scarce, a relatively rare allele (known as the "thrifty" allele) gives a heterozygous advantage of 2%. Individuals that are homozygous for the thrifty allele, because of health problems such as obesity and diabetes, have a fitness of 0.4. Calculate the expected frequency for the thrifty allele in this population.

Genotype	Frequency	After selection	ΔFrequency
A/A	p^2	p^2	0
A/a	$2pq = 2q$	$(1 + h)\ 2q$	$2hq$
a/a	q^2	$(1 - s)\ q^2$	$-Sq^2$

$\Delta q = -Sq^2 + hq = 0$ at steady state

$q = h/s = .02/.6 = .033$

b) Explain why the rate of new mutations for the thrifty allele is not relevant for this calculation.

The rate of new mutations typically has an upper limit of 10^{-4}. This value is much smaller than the heterozygote advantage and may be regarded as negligible.

c) Now consider the same continent in modern times in which the population can be thought of having two parts: 10% of the population comes from the indigenous people described above, and 90% of the population has immigrated from Europe where the thrifty allele is so rare that its frequency is effectively 0. Modern high calorie, high fat diet individuals who are homozygous for the thrifty allele are considered to have an inherited obesity related disease. Assuming random mating of the two populations, calculate the frequency of inherited obesity on the continent.

Once the populations mix, there is a new allele frequency, q.

$$q = f(a) = .1\ f(a_{IND}) + .9\ f(a_{EUR}) = (.1 \times .03) + (.9 \times 0) = .0033$$

Since there is random mating, the new genotype frequency $f(a/a) = q^2$

$$f(a/a) = q^2 = (.0033)^2 = 1.1 \times 10\text{^-}5$$

d) What would the frequency of inherited obesity on the continent be if mating between individuals were completely assortative (i.e. no mixing between the immigrant and indigenous populations)?

There is no random mating, so we need to calculate genotype frequencies separately for each group.

f(a/a) in immigrant population = 0

f(a/a) in indigenous population = q^2= $(.033)^2$ = 1 x 10^-3

f(a/a) in total population = q^2 x frequency of indigenous peoples in total population

f (a/a) = 1 x 10^-3 x .1 = 1 x 10^-4

2. In this problem we will derive general expressions for two of the more practical results from human population genetics – X-linked recessive traits occur more often in males than females and individuals with rare recessive traits often have parents who are related to one another.

a) Derive an expression, as a function of allele frequency (q), for the ratio of affected males to affected females for a X-linked recessive trait.

For an X-linked disorder, the frequency of affected males equals the frequency of the allele causing the disorder. Remember, males have one X-chromosome.

Affected males – f (X^aY) = q

Affected females – f (X^aX^a) = q^2

Affected males/Affected females = q/q^2 = 1/q

b) Derive an expression, as a function of allele frequency (q), for the probability that an individual with a rare recessive trait will have parents who are first cousins relative to the probability of first cousin parents in the general population. Assume the *a priori* probability of parents who are first cousins is 0.005. You only need to derive a formula accurate for $q < 10^{-2}$ – make any reasonable simplifying approximations that you need.

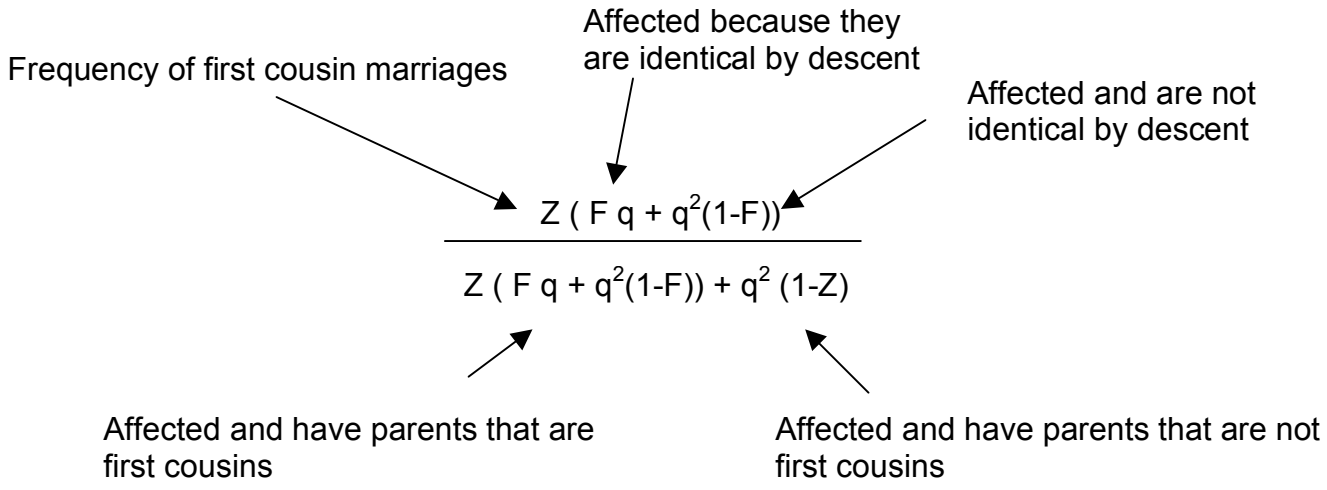
The term that we want is the following:

$$\frac{\text{Probability that an individual with a rare recessive trait will have parents who are first cousins}}{\text{Probability of first cousin parents in the general population}}$$

Let us first solve for the numerator of this term

$$\frac{\text{Percent of affected individuals with first cousin parents}}{\text{Total percent of affected individuals}}$$

This relationship can be described by the following:



Z= Frequency of first cousin marriages
F= Inbreeding Coefficient
q= allele frequency

$$\frac{Z\,(F\,q + q^2(1-F))}{Z\,(F\,q + q^2(1-F)) + q^2\,(1-Z)}$$

This term is the numerator of our final answer. Remember that to get the final answer we must divide this term by the frequency of first cousin marriages in the population (Z). Doing this gives …

$$\frac{F\,q + q^2(1-F)}{Z\,(F\,q + q^2(1-F)) + q^2\,(1-Z)}$$

Simplifying …

$$\frac{F + q\,(1-F)}{Z\,(F + q\,(1-F)) + q\,(1-Z)}$$

More simplifying …

$$\frac{F + q\,(1-F)}{Z\,F + Z\,q - Z\,q\,F + q - Z\,q}$$

and finally

$$\frac{F + q\,(1-F)}{q + Z\,F\,(1-q)}$$

This is the final simplified equation.

Notice as q -> 0 this term approaches $\frac{1}{Z}$

Therefore, as the allele frequency q decreases, the probability that an affected individual has parents that are first cousins increases

If we plug in the values for each of the components we arrive at the following solution:

$$\frac{1/16 + q\,(15/16)}{q + (1/200)\,(1/16)(1-q)} = \frac{1/16 + q\,(15/16)}{q + ((1-q)/3200)}$$

7.03 PROBLEM SET 7

BASED ON LECTURES 30-36
THIS PROBLEM SET WILL NOT BE GRADED

1) Cancer is a term used to describe a number of diseases characterized by unregulated cell growth. Cancers typically are associated with genetic changes, which can range from point mutations to large-scale chromosome abnormalities. The net effect of such mutations generally is the release of cells from their normal growth constraints.

The results from a sarcoma study involving ten individuals recently were reported. Wild-type and tumor cells were analyzed to determine the genotype of a gene involved in cell cycle regulation. In all cases, wild-type cells were heterozygous, carrying a wild-type allele and a previously uncharacterized allele. In contrast, all tumor cells were homozygous with two copies of the uncharacterized allele.

a) Based on the above information, is it more likely that this gene is an ongocene or tumor suppressor gene? Explain.

The unknown allele identified in this study was sequenced. A mutation was detected in the promoter region of one study participant. In the other nine study participants, sequencing disclosed a mutation in the coding region of this allele.

b) Describe how these mutations could result in a cancer phenotype.

c) Offer an explanation for the different frequencies observed between the two types of mutations detected in this study.

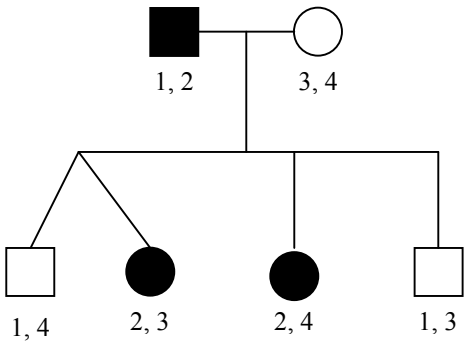
The results from a carcinoma study recently were reported. Of the ten individuals examined, nine were homozygous recessive for a gene involved in the detection of altered DNA.

d) Is this gene more likely to be an oncogene or tumor suppressor gene? Explain.

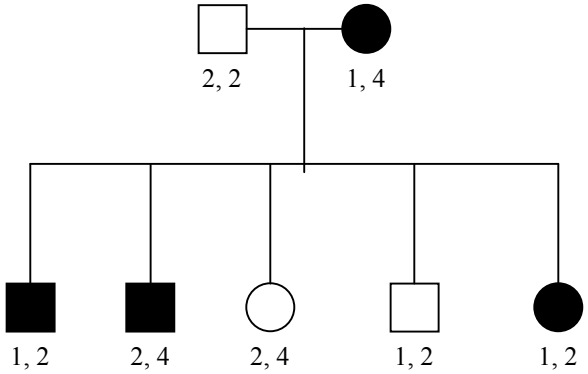
e) Propose a hypothesis to account for how the homozygous recessive condition could lead to carcinoma development.

2) You have been interested in a rare genetic disorder for some time. After conducting an exhaustive search, you identified three families with this disorder in the US. Preliminary findings suggest that the gene causing this disorder may be linked to a specific SSR. This SSR locus has four alleles, each with a different number of repeats. To begin your analysis, you constructed a pedigree for each of the three families. The pedigrees are listed below (affected individuals are darkened and the SSR genotype is listed below each individual).

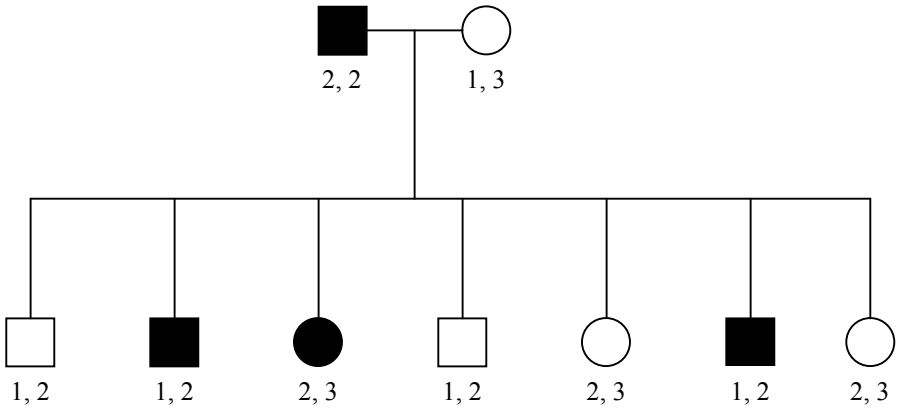
FAMILY #1



FAMILY #2



FAMILY #3



a) What is the most likely mode of inheritance for this disorder?

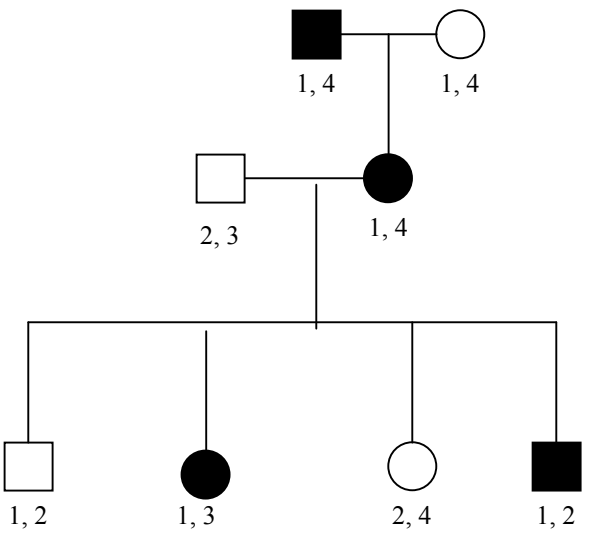
We will use information in the pedigree to calculate a LOD score for each family.

b) Which of these families can be used to calculate a LOD score to test for linkage between the SSR marker and the gene of interest? Explain.

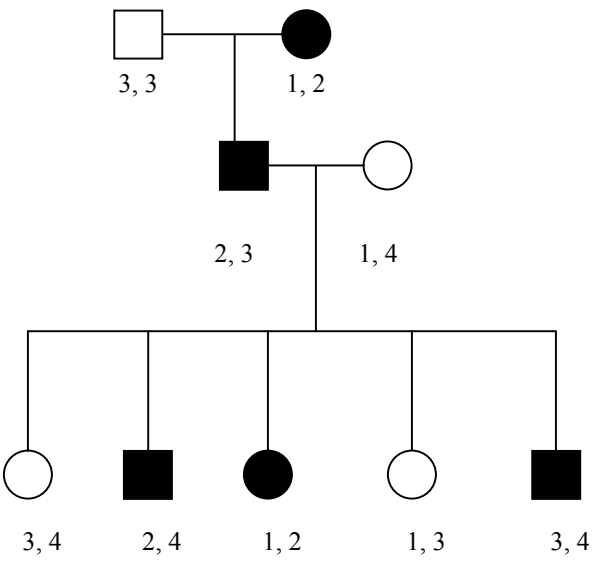
c) Calculate the LOD score at theta (θ) values of (.05, .1, .2, .3, and .4). Which θ value shows the highest odds of linkage for each family?

Three families with this disorder also were identified in Argentina. In contrast to the US families, genetic information is available for three generations in these families. The pedigree for each family is listed below.

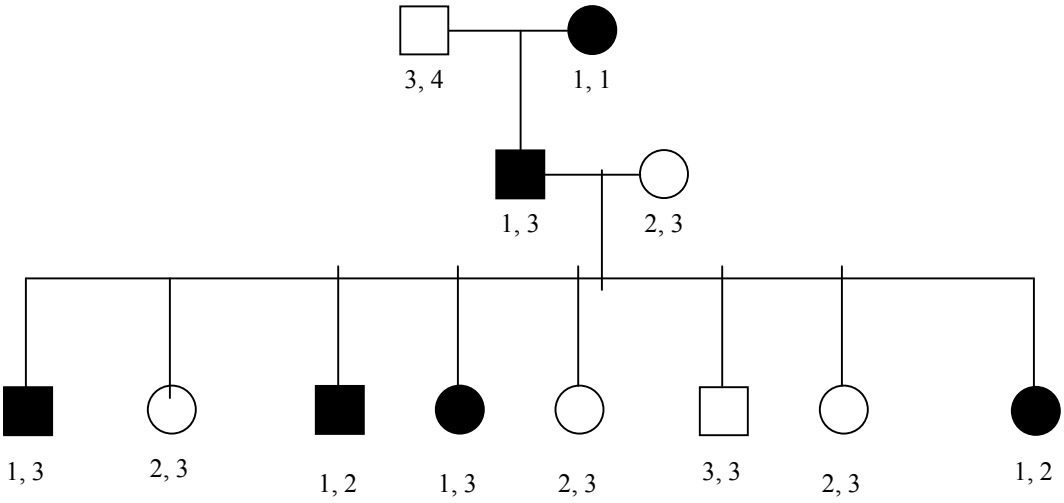
FAMILY #1



FAMILY #2



FAMILY #3



d) Which of these families can be used to calculate a LOD score? Which is the relevant parent in each pedigree?

e) Calculate the LOD score at theta (θ) values of (.05, .1, .2, .3, and .4). Which θ value shows the highest odds of linkage for each family?

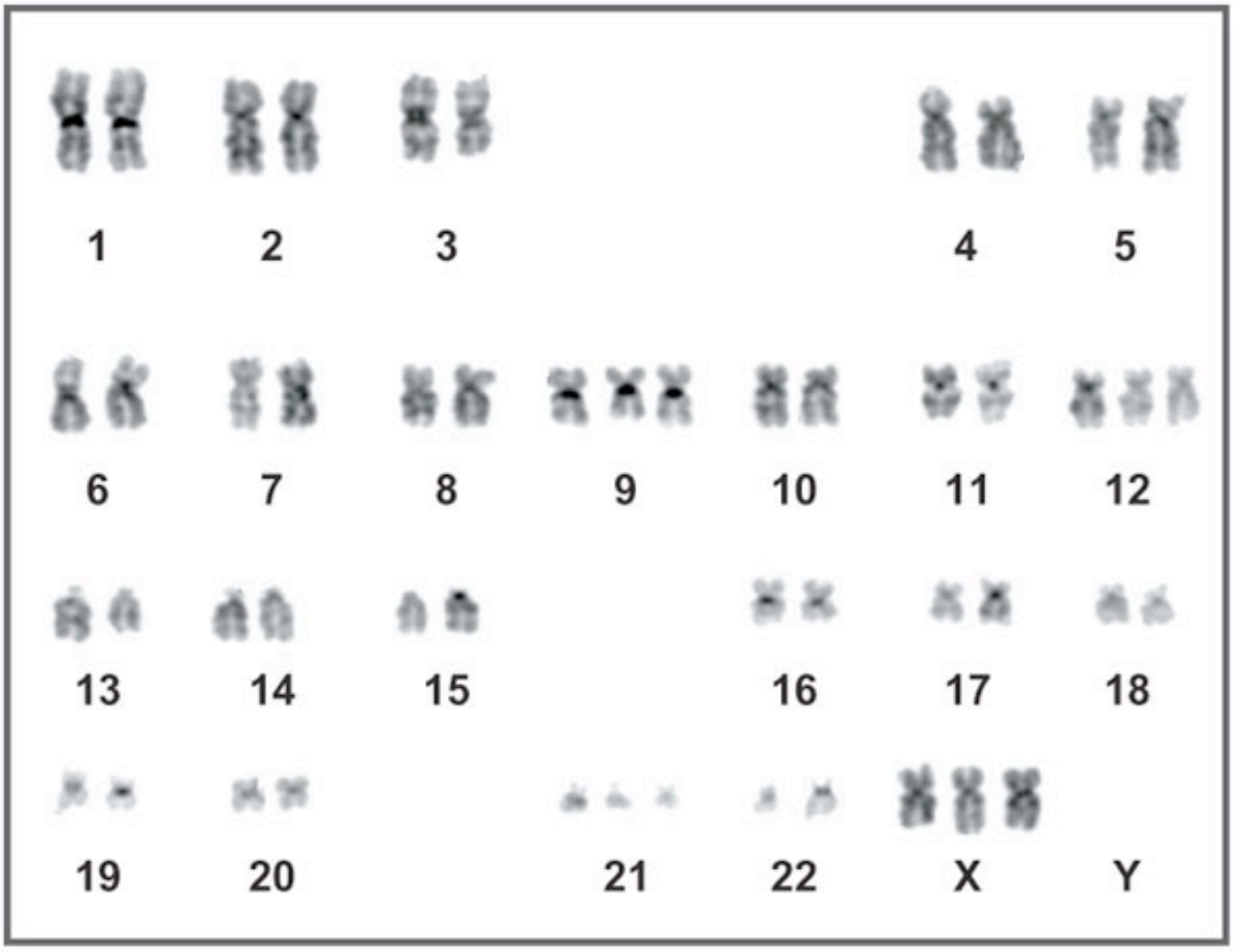
f) Plot the LOD score for each theta value for the US and Argentine families. Is the curve similar for the various families?

g) Do the LOD scores for the US families suggest linkage? Do the LOD scores for the Argentine families suggest linkage? Explain.

h) Can the results from the Argentine and US families be combined? If so, what is your final conclusion regarding linkage between the gene of interest and SSR marker?

3) Chromosome abnormalities are associated with a number of disorders including some cancers. Duplications, for example, have been detected in some carcinomas and inversions in some lymphomas. It is believed that the analysis of such abnormalities could provide important clues to key oncogenic events and perhaps suggest potential avenues for cancer treatment.

The karyotype for a specific type of carcinoma is displayed below.



a) Identify the chromosome abnormalities evident in this karyotype.

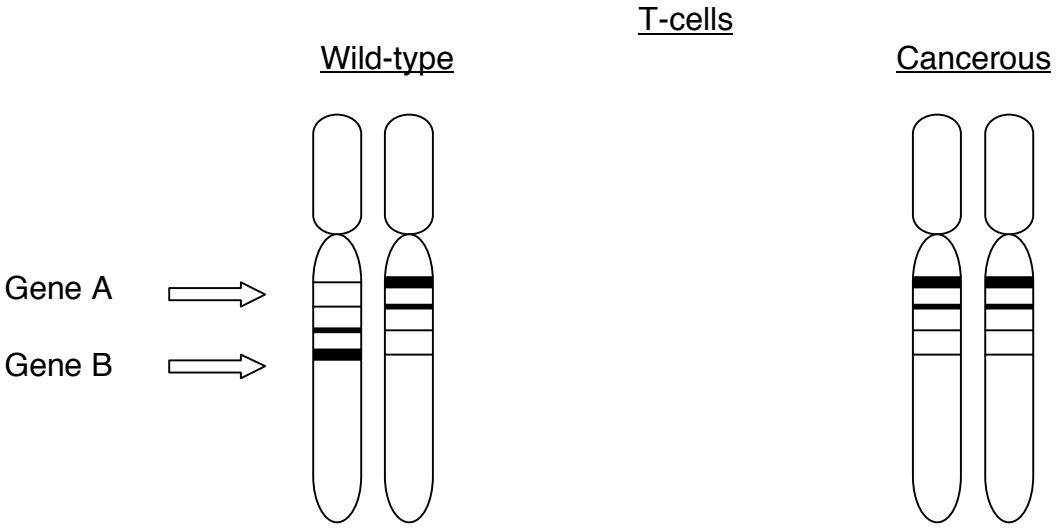
Sixteen individuals with a specific carcinoma were screened for chromosomal aberrations. In eleven of these individuals, the short arm of chromosome eight is longer in wild-type cells compared to cancer cells.

b) Which type of abnormality exists in the cancer cells? Explain.

c) Describe how the chromosome abnormality listed in **part "b"** could lead to carcinoma development.

d) In each of the eleven individuals described above, a subset of cancer cells was aneuploid. Why wasn't aneuploidy observed in all cancer cells?

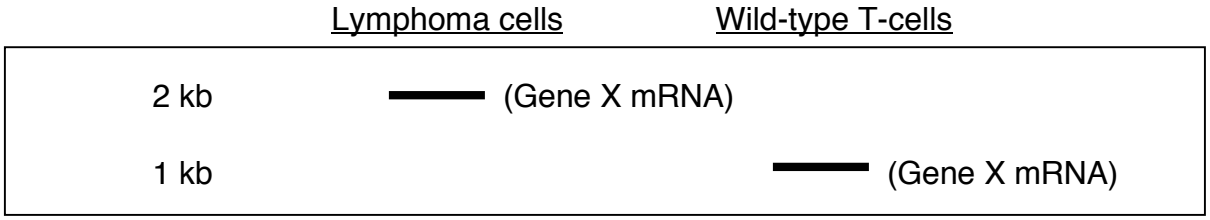
One goal of an on-going lymphoma study is to determine if a correlation exists between a chromosome #2 aberration and lymphoma development. Some early results from this study are listed below.



	Wild-type	Cancerous
Gene A Expression	+	+
Gene B Expression	+	-

e) Is there a chromosome abnormality? Describe a mechanism to account for the development of the cancer phenotype.

Wild-type T-cells express Gene X at high levels. A type of T-cell lymphoma is heterozygous for a specific translocation that is absent in wild-type cells. Despite the translocation, lymphoma cells continue to show high levels of Gene X expression. The results from Northern blot analysis of mRNA isolated from lymphoma or wild-type T-cells are shown below).



f) Propose a hypothesis to account for the results shown above.

The Gene X transcripts from lymphoma or wild-type T-cells were translated using an *in vitro* system. The resulting polypeptides were assayed for tyrosine kinase activity (see results below).

	Transcripts from lymphoma cells	Wild-type transcripts
Activity Level	+++	-

g) Based on the above results, propose a mechanism to account for development of the cancer phenotype.

7.03 PROBLEM SET 7

BASED ON LECTURES 30-36
THIS PROBLEM SET WILL NOT BE GRADED

1) Cancer is a term used to describe a number of diseases characterized by unregulated cell growth. Cancers typically are associated with genetic changes, which can range from point mutations to large-scale chromosome abnormalities. The net effect of such mutations generally is the release of cells from their normal growth constraints.

The results from a sarcoma study involving ten individuals recently were reported. Wild-type and tumor cells were analyzed to determine the genotype of a gene involved in cell cycle regulation. In all cases, wild-type cells were heterozygous, carrying a wild-type allele and a previously uncharacterized allele. In contrast, all tumor cells were homozygous with two copies of the uncharacterized allele.

a) Based on the above information, is it more likely that this gene is an ongocene or tumor suppressor gene? Explain.

Given this information, it seems more likely to be a tumor suppressor gene. The unknown allele may contain a loss of function mutation. If this were the case, then the cancer cells would lack a functional allele resulting in unregulated growth.

The unknown allele identified in this study was sequenced. A mutation was detected in the promoter region of one study participant. In the other nine study participants, sequencing disclosed a mutation in the coding region of this allele.

b) Describe how these mutations could result in a cancer phenotype.

The promoter region mutation may prevent transcription, so the absence of gene product could lead to unregulated cell growth. In contrast, the coding region mutation may abolish the wild-type activity of the encoded polypeptide, preventing it from executing its role in cell cycle regulation.

c) Offer an explanation for the different frequencies observed between the two types of mutations detected in this study.

Nine of the individuals in the study are from the same family.

One allele is more common in the population than the other.

The coding region may contain a hotspot for mutation.

If the coding region were considerably longer than the promoter region, then the probability of a coding region mutation event would be comparatively greater.

The results from a carcinoma study recently were reported. Of the ten individuals examined, nine were homozygous recessive for a gene involved in the detection of altered DNA.

d) Is this gene more likely to be an oncogene or tumor suppressor gene? Explain.

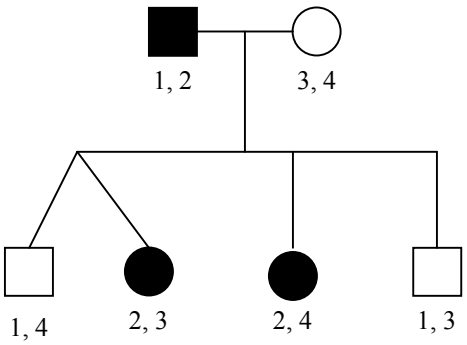
It is more likely a tumor suppressor gene. If this were the case, then the homozygous recessive individuals would be without any functional gene product. The absence of functional gene product could lead to tumor formation.

e) Propose a hypothesis to account for how the homozygous recessive condition could lead to carcinoma development.

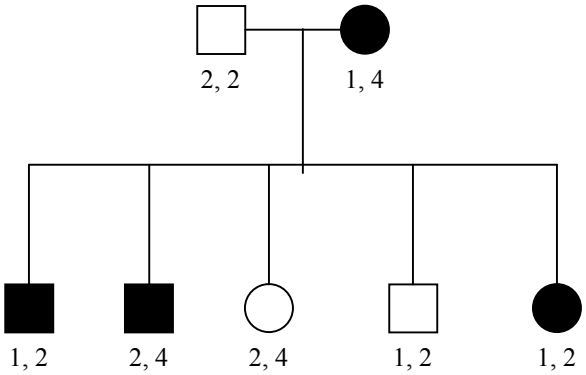
The absence of gene product prevents the cell from detecting DNA damage due to mutation. This defect leads to the accumulation of deleterious mutations, potentially ranging from point mutations to chromosome abnormalities. These mutations eventually will lead to cell-cycle de-regulation.

2) You have been interested in a rare genetic disorder for some time. After conducting an exhaustive search, you identified three families with this disorder in the US. Preliminary findings suggest that the gene causing this disorder may be linked to a specific SSR. This SSR locus has four alleles, each with a different number of repeats. To begin your analysis, you constructed a pedigree for each of the three families. The pedigrees are listed below (affected individuals are darkened and the SSR genotype is listed below each individual).

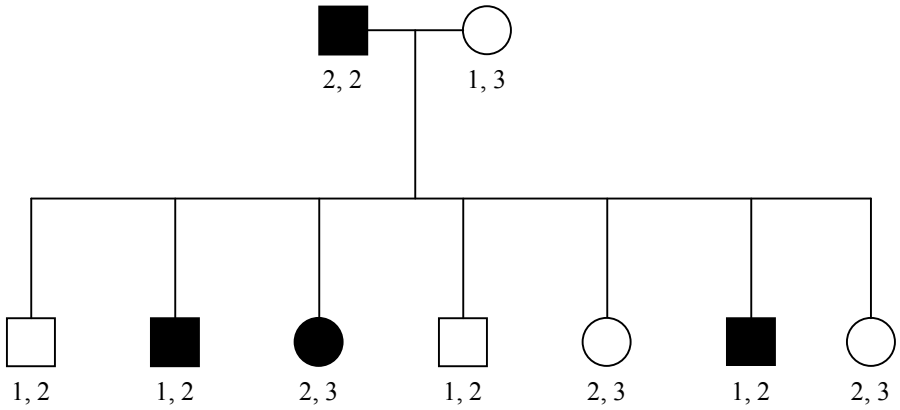
FAMILY #1



FAMILY #2



FAMILY #3



a) What is the most likely mode of inheritance for this disorder?

Based on its appearance in both generations, it would seem to be autosomal dominant.

We will use information in the pedigree to calculate a LOD score for each family.

b) Which of these families can be used to calculate a LOD score to test for linkage between the SSR marker and the gene of interest? Explain.

Families one and two can be used. The relevant parents in each of these families are heterozygous for the SSR and exhibit the disorder.

c) Calculate the LOD score at theta (θ) values of (.05, .1, .2, .3, and .4). Which θ value shows the highest odds of linkage for each family?

Calculating the LOD scores for the US families requires accounting for both possible phases in the relevant parent, as the phase is unknown.

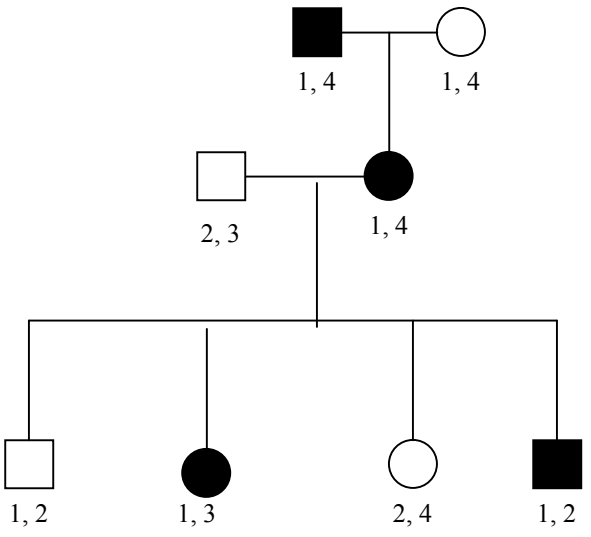
US FAMILIES

THETA	FAMILY1	FAMILY2
0.05	0.8139	-1.4424
0.1	0.7201	-0.8873
0.2	0.5171	-0.3876
0.3	0.2978	-0.1514
0.4	0.0939	-0.0354

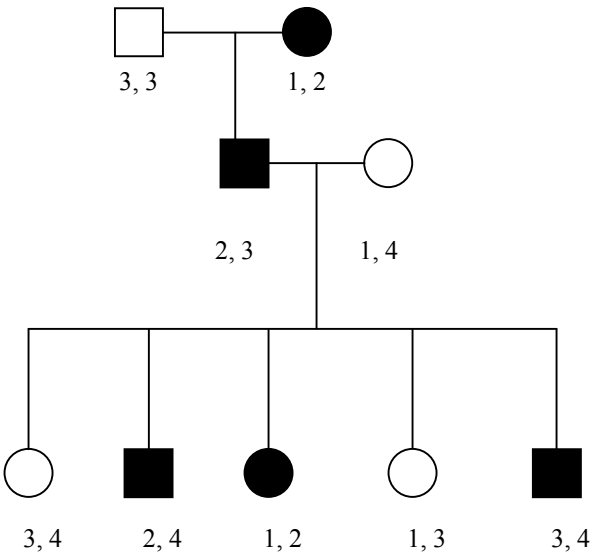
Among the theta values tested, a theta value of 0.05 showed the highest odds of linkage for family #1. Keep in mind, though, that this low LOD score does not suggest linkage. For family #2, a theta value of 0.4 showed the highest odds of linkage. Again, this low LOD score does not suggest linkage between the disorder-causing gene and the SSR.

Three families with this disorder also were identified in Argentina. In contrast to the US families, genetic information is available for three generations in these families. The pedigree for each family is listed below.

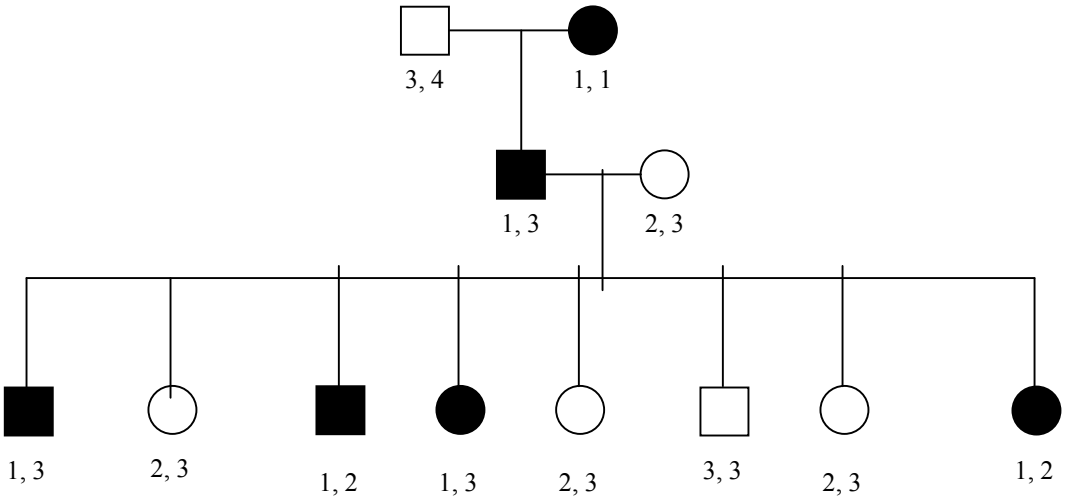
FAMILY #1



FAMILY #2



FAMILY #3



d) Which of these families can be used to calculate a LOD score? Which is the relevant parent in each pedigree?

All three families can be used to calculate LOD scores. In family 1, the mother in the second generation is the relevant parent. As her phase is unknown, both possible phases must be accounted for in the LOD score calculation. For family 2, the mother in the first generation is relevant to establish the phase of the father in the second generation, who is the relevant parent. For family 3, the mother in

the first generation is relevant to establish the phase of the father in the second generation, who is the relevant parent.

e) Calculate the LOD score at theta (θ) values of (.05, .1, .2, .3, and .4). Which θ value shows the highest odds of linkage for each family?

The phase of the relevant parent in family #1 is unknown and must be accounted for in the LOD score calculation.

LOD Score

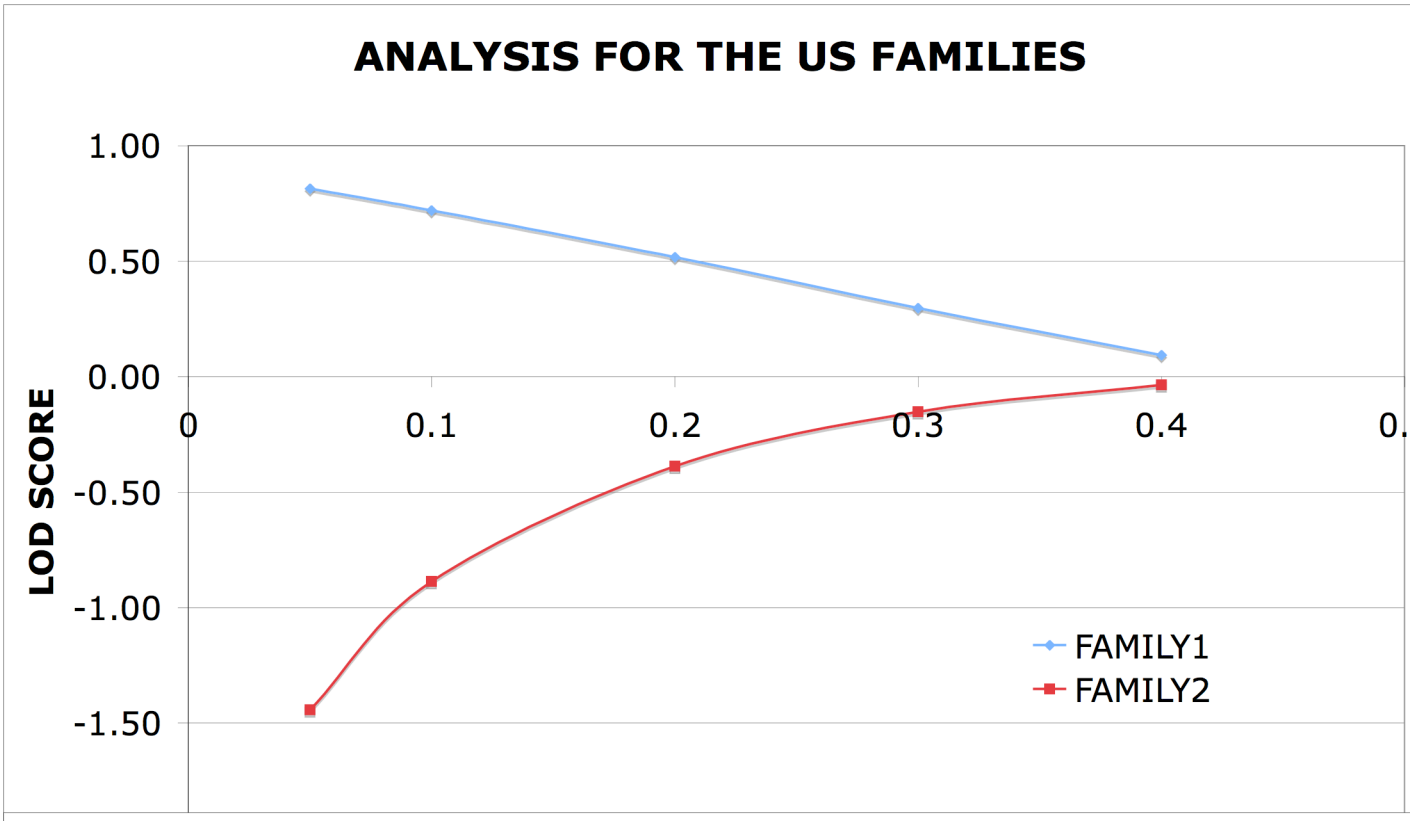
ARGENTINA

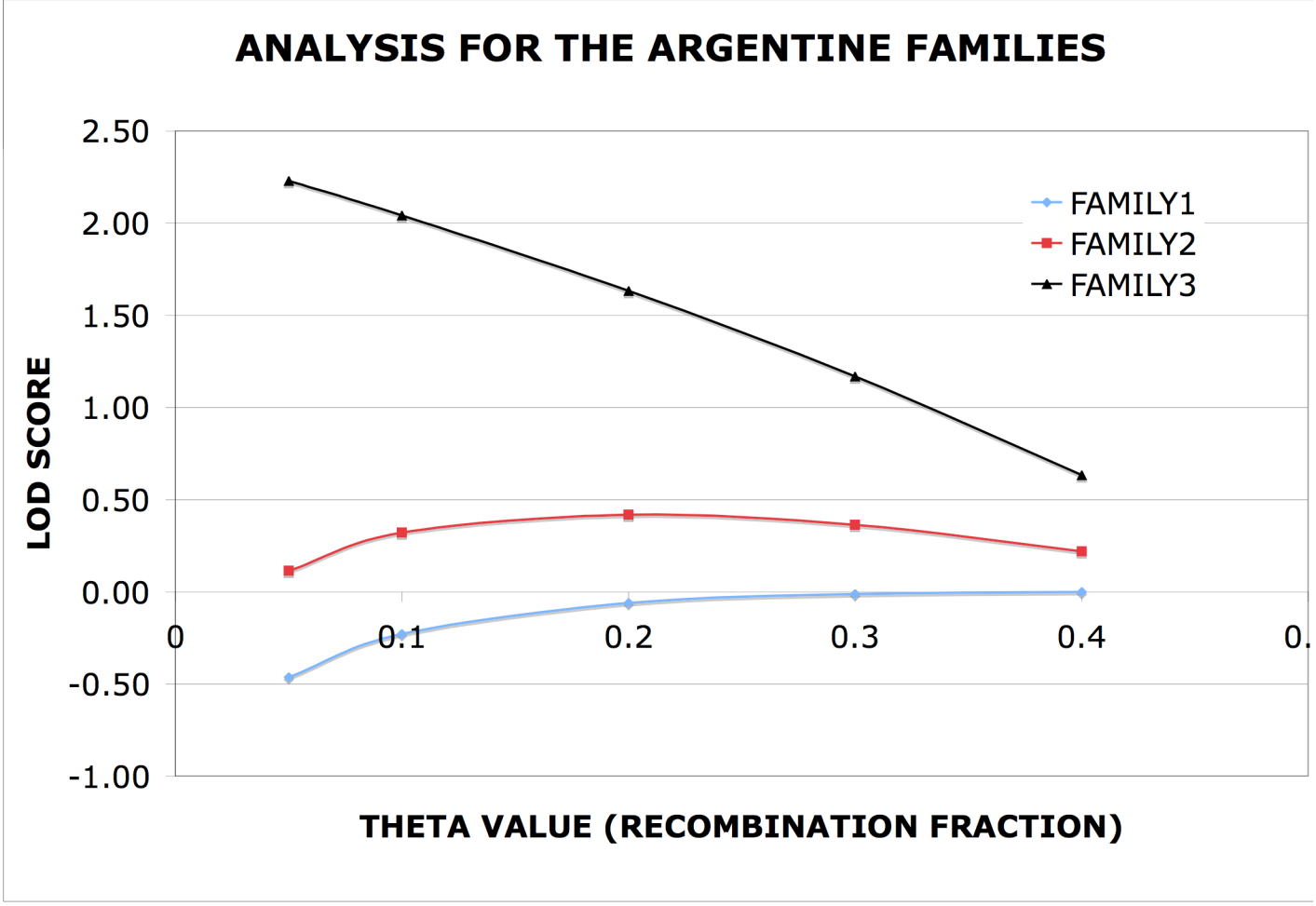
THETA	FAMILY1	FAMILY2	FAMILY3
0.05	-0.4635	0.115	2.23
0.1	-0.2288	0.3221	2.042
0.2	-0.0602	0.4185	1.6329
0.3	-0.0112	0.3627	1.169
0.4	-0.0007	0.2198	0.6334

For family 1, a recombination fraction of 0.4 showed the highest odds yet does not suggest linkage. A recombination fraction of 0.2 yielded the highest odds for family 2 and does not suggest linkage. In contrast, family 3 showed the highest odds of linkage at a recombination fraction of .05 without suggesting linkage. Based on this analysis, linkage between the two loci is unlikely for all three families. Due to the independence of the pedigrees, though, the LOD scores at each theta value can be pooled across families. The combined LOD scores still fall below the conventional standard (LOD score of ≥ 3) necessary to suggest linkage.

f) Plot the LOD score for each theta value for the US and Argentine families. Is the curve similar for the various families?

Except for the Argentine families #1 and #2, the curves differ substantially. The curves illustrate the low odds of linkage between the two loci under investigation.


ANALYSIS FOR THE US FAMILIES
LOD SCORE
1.00
0.50
0.00
-0.50
-1.00
-1.50
0
0.1
0.2
0.3
0.4
FAMILY1
FAMILY2


ANALYSIS FOR THE ARGENTINE FAMILIES
LOD SCORE
2.50
2.00
1.50
1.00
0.50
0.00
-0.50
-1.00
0
0.1
0.2
0.3
0.4
FAMILY1
FAMILY2
FAMILY3
THETA VALUE (RECOMBINATION FRACTION)

g) Do the LOD scores for the US families suggest linkage? Do the LOD scores for the Argentine families suggest linkage? Explain.

For the US families, the LOD scores associated with each theta value do not suggest linkage. In addition, combining the LOD scores for families 1 and 2 at each theta value fails to suggest linkage.

For the Argentine families, none of the LOD scores for the theta values tested suggest linkage. Pooling the LOD scores likewise fails to suggest linkage at any of the tested theta values.

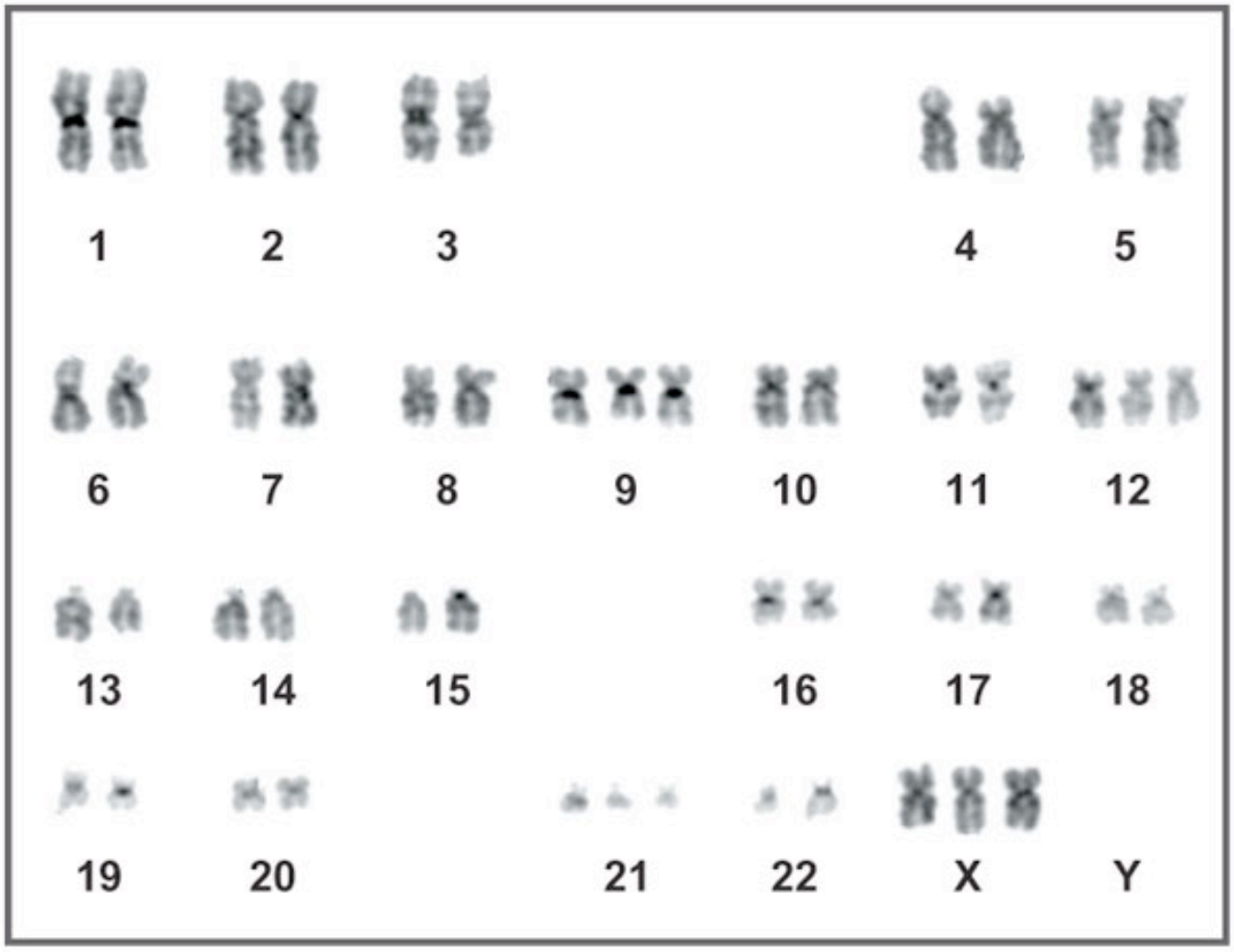
h) Can the results from the Argentine and US families be combined? If so, what is your final conclusion regarding linkage between the gene of interest and SSR marker?

Yes, the LOD scores can be combined. All five of the pedigrees are assumed to be independent since no one individual can be placed in more than one pedigree.

Combining the LOD scores between the US and Argentine families does not suggest linkage between the disorder-causing gene and the SSR.

3) Chromosome abnormalities are associated with a number of disorders including some cancers. Duplications, for example, have been detected in some carcinomas and inversions in some lymphomas. It is believed that the analysis of such abnormalities could provide important clues to key oncogenic events and perhaps suggest potential avenues for cancer treatment.

The karyotype for a specific type of carcinoma is displayed below.



a) Identify the chromosome abnormalities evident in this karyotype.

Chromosomes 9, 12, and 21 are trisomic as is the X-chromosome.

There appears to be either a duplication or deletion in chromosome 22 as well.

Sixteen individuals with a specific carcinoma were screened for chromosomal aberrations. In eleven of these individuals, the short arm of chromosome eight is longer in wild-type cells compared to cancer cells.

b) Which type of abnormality exists in the cancer cells? Explain.

The data suggest that the cancer cells are homozygous for a deletion in the short of chromosome eight.

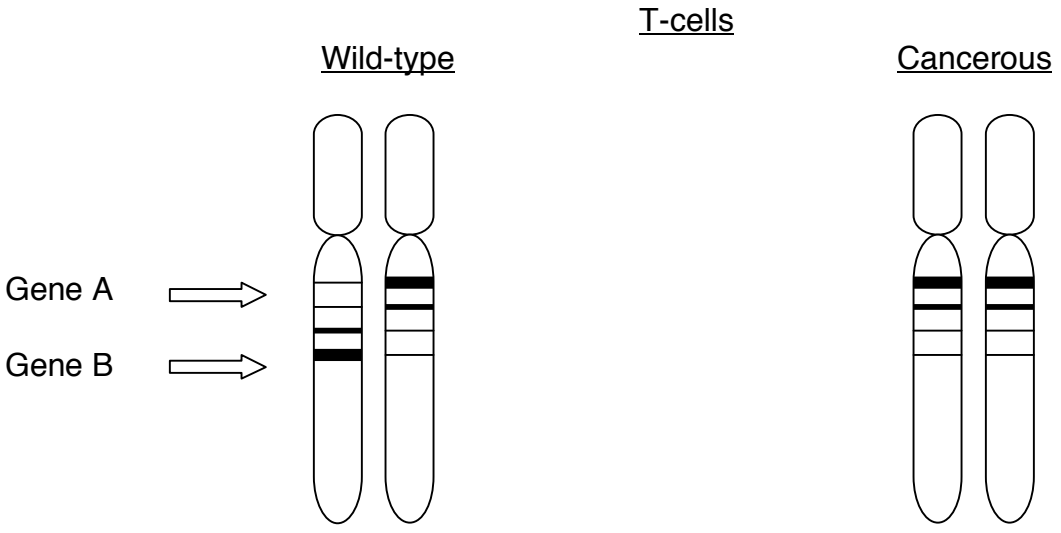
c) Describe how the chromosome abnormality listed in **part "b"** could lead to carcinoma development.

The deletion event may have eliminated a key tumor suppressor gene. As the cancer cells are homozygous for the deletion, there would be no functional gene product to participate in cell cycle regulation.

d) In each of the eleven individuals described above, a subset of cancer cells was aneuploid. Why wasn't aneuploidy observed in all cancer cells?

Tumors consist of clonal populations of cells that are continually mutating. Thus, not all cells of the tumor share the same genotype. In this case, some of the cancer cells have one or more mutations that result in aneuploidy.

One goal of an on-going lymphoma study is to determine if a correlation exists between a chromosome #2 aberration and lymphoma development. Some early results from this study are listed below.

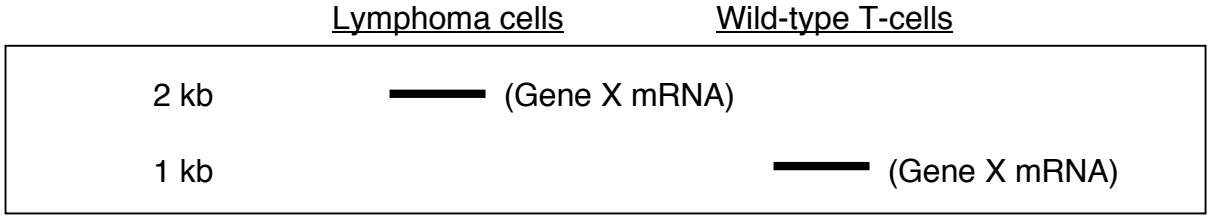


	Wild-type	Cancerous
Gene A Expression	+	+
Gene B Expression	+	-

e) Is there a chromosome abnormality? Describe a mechanism to account for the development of the cancer phenotype.

The results show that an inversion has occurred. The cancer cells are homozygous for this inversion in chromosome two. In the cancer cells, the inversion is associated with the absence of Gene B expression. One hypothesis is that Gene B is a tumor suppressor gene, and the inversion silenced its expression. Without any Gene B product, the cell cycle no longer is regulated in a wild-type manner.

Wild-type T-cells express Gene X at high levels. A type of T-cell lymphoma is heterozygous for a specific translocation that is absent in wild-type cells. Despite the translocation, lymphoma cells continue to show high levels of Gene X expression. The results from Northern blot analysis of lymphoma and wild-type T-cells are shown below).



f) Propose a hypothesis to account for the results shown above.

The elongated transcript in lymphoma cells indicates that the translocation caused a gene fusion between Gene X and another gene.

The Gene X transcripts from lymphoma or wild-type T-cells were translated using an *in vitro* system. The resulting polypeptides were assayed for tyrosine kinase activity (see results below).

	Transcripts from lymphoma cells	Wild-type transcripts
Activity Level	+++	-

g) Based on the above results, propose a mechanism to account for development of the cancer phenotype.

The translocation event resulted in a gene fusion between the highly expressed Gene X and a tyrosine kinase gene. The high expression of the gene fusion resulted in constitutive, or at least more frequent, tyrosine kinase activity. As tyrosine kinases play key roles in cell cycle regulation, the cancer phenotype may be due to the over-stimulation of cell division.