NAME: ____

STUDENT ID #: ________________________________

MCB 140 FINAL EXAM

Fall 2007

NAME (Please print):____

STUDENT ID #:____

REMINDERS

You have 180 minutes for the 225 point exam.

Print your name and ID# on each page of the exam. You will lose points if you forget to do this.

There are 17 pages total, including this cover page. The last two pages are work space (no problems). All pages must be turned in.

Only the front of each page will be graded. If you use the back of a page, transcribe your answer to the front of the page.

Use a non-programmable calculator.

--------------------------------------(Do not write below this line)-------------------------------------

1 (5) _____

2 (10) _____

3 (10) _____

4 (25) _____

5 (10) _____

6 (20) _____

7 (10) _____

8 (10) _____

9 (10) _____

10 (9) _____

11 (15) _____

12 (15) _____

13 (24) _____

14 (30) _____

15 (10) _____

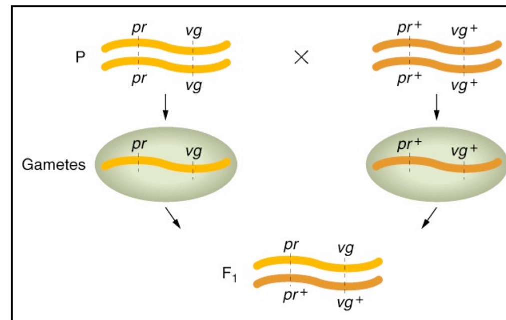
16 (10) _____

TOTAL _____ / 225

NAME: ___________________________________

STUDENT ID #: ____________________________

1. (5 points) T.H. Morgan found that two genes – pr and vg (see schematic below) – do not follow Mendel's second law. What specific aspect of meiosis ensures that these two loci follow Mendel's first law?



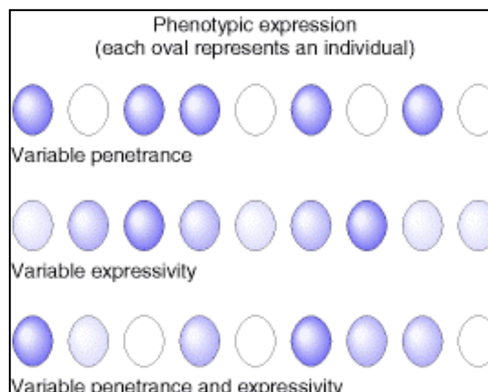
The segregation of nonsister homologs into two daughter cells during meiosis I.

2. (10 points) Vernon Ingram concluded his justly famous 1956 paper on the biochemical difference between sickle cell and wild-type hemoglobin with the following statement: "One can now answer at least partly the question put earlier, and say there there is a difference in the amino-acid sequence in one small part of one of the polypeptide chains. <u>This is particularly interesting in view of the genetic evidence that the formation of hemoglobin S is due to a mutation in a single gene.</u>"

   What was the evidence leading to the conclusion that sickle cell anemia is, in fact, due to a mutation in a single gene? (10 points)

   The inheritance of sickle-cell anemia in families followed simple Mendelian inheritance.

3. (10 points) The inheritance of many traits exhibits variable penetrance and/or expressivity. List one human trait we discussed, and state, what it is characterized by (variable penetrance and/or expressivity; see schematic below). Finally, state the best explanation we currently have on *why* this trait exhibits such variability.



Sample answers:

NAME: ____________________________________

STUDENT ID #: ______________________________

Sickle-cell anemia; variable expressivity; effects of modifier loci
Breast cancer; variable penetrance; environmental effects

4. (25 points total) The three deficiencies *Df1, Df2* and *Df3* remove genes on the *C. elegans* fourth chromosome. In mapping experiments you cross males hemizygous for these three deficiencies (*Df/*+) to hermaphrodites that are doubly mutant for recessive mutations in *dpy-5*, which is on chromosome I, and one of four mutations on chromosome IV: *ced-3, ham-1, unc-30* and *unc-31*. The phenotypes of the progeny of these crosses are shown in the table below.

male parent	Hermaphrodite parent			
	dpy-5; ced-3	*dpy-5; ham-1*	*dpy-5; unc-30*	*dpy-5; unc-31*
Df1/+	+	-	-	+
Df2/+	+	+	-	-
Df3/+	-	+	+	+

"+" indicates complementation; "-" indicates failure to complement

The *dpy-5* mutation causes a Dumpy phenotype; The *ham-1* mutation causes a cell differentiation defect; the *unc-30* and *unc-31* mutations cause an uncoordinated phenotype.

(a) (5 points) In these crosses, why are all of the hermaphrodites made homozygous for the *dpy-5* mutation?

To distinguish self (*dpy-5/dpy-5)* from cross (*dpy-5/+)* progeny.

(b) (5 points) What are the phenotypic classes and sexes of the progeny on the plate produced when *Df1*/+ males are crossed with *dpy-5, unc-31* hermaphrodites?

Dpy, Unc hermaphrodites

Wild-type (or NonDpy, NonUnc) males and hermaphrodites

(c) (5 points) What are the phenotypes of the progeny produced when *Df1*/+ males are crossed with *dpy-5, unc-30* hermaphrodites?

NAME: ________________________________

STUDENT ID #: ________________________

Dpy, Unc hermaphrodites

Unc (or NonDpy, Unc) males and hermaphrodites

Wild-type (or NonDpy, NonUnc) males and hermaphrodites

(d) (10 points) Draw a map showing the positions of the three deficiencies and the *ced-3, ham-1, unc-30* and *unc-31* genes. If there are any ambiguities, describe them.

ced-3 ***ham-1*** ***unc-30*** ***unc-31***

Df1 ______________ (spanning ham-1 to unc-30)

Df2 ______________ (spanning unc-30 to unc-31)

Df3 _____ (covering ced-3)

OR

ham-1 ***unc-30*** ***unc-31*** ***ced-3***

Df1 ______________ (spanning unc-30 to unc-31)

Df2 ______________ (spanning ham-1 to unc-30)

Df3 _____ (covering ced-3)

5. (10 points) Gene therapy is the name for techniques in which DNAs are used as treatment for human disease: the artificial gene is transcribed and translated inside the patient's tissues to produce exogenous protein. Huntington's disease is a genetic disorder caused by an extra-long allele of the huntingtin gene which encodes extra glutamine residues in the protein sequence. The long glutamine-repeat proteins ultimately aggregate in the brain. A colleague of yours proposes to make his fortune by designing a gene therapeutic of wild-type huntingtin protein for the treatment of Huntington's patients. His new methods guarantee that a single copy of the wild-type gene will be introduced into each neuron. Even so, you have to break it to him that his scheme will not work as a complete cure for Huntington's. Why not?

   Huntington's alleles follow a dominant genetic model. If you introduce a wild-type copy of the gene, this will not prevent the mutant copies in the patient's genome from encoding proteins that continue to aggregate.

6. The aggregation of the yeast prion Sup35 impairs translation termination. In aggregation-prone (PSI+) strains, during translation of a given gene, the ribosome "reads through" the primary stop codon, translating nucleotide triplets in the sequence

beyond the stop codon as though they were ordinary coding sequence. This results in a protein identical to wild-type except that it contains additional amino acids at the 3' (C-terminal) end.

(a) (10 points) Wild isolates (genetically distinct strains) of yeast have different phenotypes when in the PSI+ state, even though the extent of Sup35 aggregation is the same. Some strains show strong resistance or sensitivity to environmental treatment, while others show unusual colony morphologies. Explain why the phenotypic effect of Sup35 aggregation is not the same for all strains.

Strains carry different alleles in the sequence beyond the stop codon. This causes sequence differences in the new C-terminal tails of the proteins which can confer changes in function.

(b) (10 points) Genetically distinct strains, when they are PSI- (no Sup35 aggregation), tend to have barely noticeable phenotypic differences—much less dramatic than when the same strains are converted to PSI+. The reason for this contrast has to do with random mutations and their elimination by selection, but where exactly? Give an evolutionary explanation for the fact that polymorphisms whose effects are observable in PSI+ strains have more dramatic effects than polymorphisms whose effects are observable normally.

Mutations that arose in the sequence beyond the stop codon, during the evolution of the strains, usually were not translated and probably had no phenotypic effect. This means they could be maintained at appreciable frequency due to genetic drift. By contrast, dramatic mutations that arose elsewhere in the genome, like in promoter or coding sequences, usually had deleterious effects and were culled by natural selection; the occasional mutation with very strong beneficial effect will have swept through the population. Only more subtle mutations in functional regions, which have weaker phenotypic effects (good or bad), show up at appreciable frequency.

7. (10 points) You are interested in the genetic determinants of lifespan in *C. elegans*. You already have in the lab two genetically distinct natural isolates which differ in lifespan, and a large number of F2 progeny from a cross between them, genotyped at densely spaced markers; heritability of lifespan is high in the cross, so you know the trait could be mapped. However, you also have in the lab a large collection of mutagenized animals (derived from a single parent) which could be used in a genetic screen. For each experiment—linkage mapping of natural variants and genetic screening—give one reason why you might fail to identify genes underlying lifespan differences.

Linkage mapping fails whenever the effects of any individual locus are too small to be statistically significant. This could be because the trait is genetically complex or because the threshold for significance is "too" high, possibly due to the multiple testing problem.

A screen may fail if it has not been taken to saturation, such that there do exist genes in which mutations would affect phenotype, but such mutants were not examined. A screen will also fail if no single mutation is sufficient to affect the phenotype. In the latter scenario, a mutation may land in a gene in a pathway related to the trait, but

other genes in the pathway have redundant functions and so cover up (are epistatic to) the effect of the mutation; a strain harboring mutations in multiple genes in the pathway will show effects on the trait, but multiple mutations rarely come up or are analyzed in a screen of this type.

8. Imagine that you are interested in finding the genetic determinants of a severe human disease that strikes in early childhood but does allow patients to live into adulthood. The disease is very rare, indicating that alleles underlying the disease are not present at high frequency in the human population. When you do see cases, they are invariably clustered in families with multiple affected individuals.

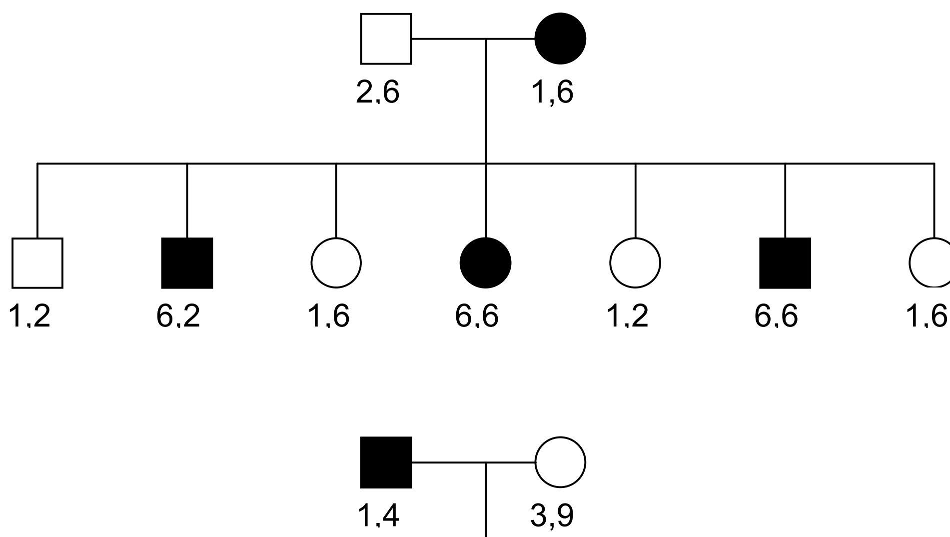
   (a) (5 points) You collect data from six extended families and are planning a linkage study with a dense marker set. The total number of subjects should be plenty for statistical considerations, but no single family is big enough for you to analyze on its own. But one major concern still looms, making you worry that may not get any significant linkage signal for this disease; what is it?

   Locus heterogeneity—the disease may be segregating in each family for a different genetic reason.

   (b) (5 points) Speculate on the evolutionary history of the alleles that cause the disease in these families.

   The mutations probably arose in the recent ancestors of the modern-day families, rather than having arisen many generations ago and been passed throughout the larger population.

9. (10 points) You're using linkage analysis to study a human disease that is caused by a single locus following a dominant genetic model. You focus on a single restriction fragment length polymorphism marker that shows the best evidence for linkage. For this marker, the lengths in kilobases of the fragments recognized by the probe in each individual are shown below in two families. For each pedigree, the affected parent is a heterozygote at the disease locus and phase is known: in the mother of the top family, the disease allele is on the same homolog as the 6 allele of the marker, and in the father of the bottom family, the disease allele is on the same homolog as the 1 allele of the marker.



NAME: ______________________________

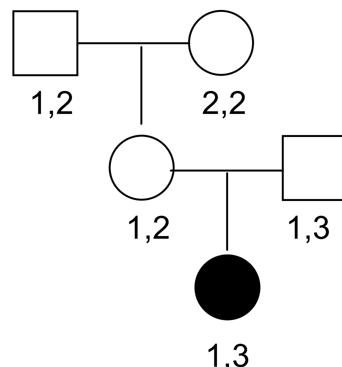
STUDENT ID #: ___________________________

Imagine that you pursue molecular confirmation experiments and find a large gene deletion that is truly causal for the disease, closely linked to this marker. All the affected individuals in both families have the deletion, and all the healthy individuals have a full-length copy of the gene. But the pedigree shows that a different marker allele is linked to the disease mutation in each family (allele 6 in the top family, allele 1 in the bottom family). How can this be? Propose an explanation for the fact that the alleles of the marker that co-segregate with the disease mutation differ between the families, even though the disease mutation itself is the same.

The deletion at the disease locus may have been introduced independently in two recent ancestors, one of each family. Allele 6 at the marker arose long ago and was passed down to an ancestor of the top family, in whom the deletion at the disease locus arose (or was introduced by recombination) more recently; allele 1 at the marker also arose long ago and was passed down to an ancestor of the bottom family, in whom the deletion at the disease locus arose (or was introduced) more recently.

Equivalently, the polymorphisms giving rise to the marker may have arisen independently in each family: the deletion mutation arose long ago in a single individual and was passed down to ancestors of each family, but allele 6 at the marker arose (or was introduced by recombination) in a recent ancestor of the top family, and allele 1 arose (or was introduced) in a recent ancestor of the bottom family.

10. (3 points each) Consider the following human pedigree, in which the segregating disease is controlled by a single locus which follows a recessive model, and the labels represent alleles observed at a RFLP marker linked to the disease locus. Assume that individual I-2 is homozygous wild-type at the disease locus.



For each individual, determine whether the phase between alleles at the disease locus and alleles at the marker is known. Do not count recombinations—just determine whether there is a single possible arrangement of alleles at the two loci (the haplotype) on each homolog for:

NAME: ___________________________________

STUDENT ID #: _____________________________

a. Individual II-1

Yes, the disease allele is on the same homolog as the 1 allele inherited from I-1.

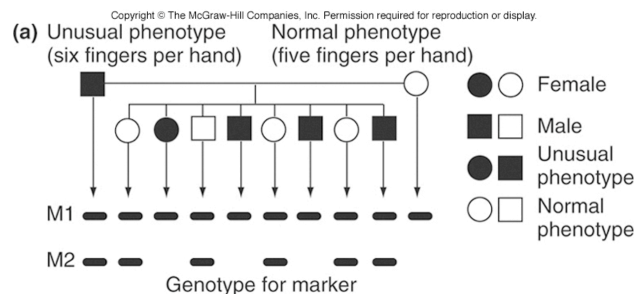
b. Individual II-2

No

c. Individual III-1

Yes, there are two disease alleles so there is only one possible arrangement.

11. (15 points) This human pedigree, familiar to you from lecture, depicts the segregation of a phenotype controlled by a single locus that follows a dominant model, and the segregation of two alleles, M1 and M2, at a RFLP marker:



Assume that the marker is close to the disease locus. Assume also that at the disease locus, the father is a heterozygote and the mother is homozygote wild-type. In lecture we swept under the rug the fact that phase is not known in the father. In reality, because there is equal a priori probability of each phase in the father, the true expression for the odds that the disease locus and the marker are separated by genetic distance r, relative to the odds that the disease locus and the marker are unlinked, is

$$\text{Odds} = 0.5*[(1-r)^7 r^1]/[0.5^8] + 0.5*[(1-r)^1 r^7]/[0.5^8].$$

Plugging in r = 0.1, the first term comes out to 6.122 and the second comes out to 1.15 x 10^{-5}. Explain in genetic terms why the second term contributes so little to the total odds of linkage.

It is much less likely to have the marker linked to the disease locus and yet have 7 recombinations. In other words, the phase in which the disease allele in the father is on the same homolog as allele 2 of the marker is much less likely.

12. (15 points) In the last quiz, you considered a mouse pedigree of 10 meioses in which a disease is segregating. You saw that if you test 10,000 markers which are truly unlinked to the disease (and independent), one would expect ~10 markers to show spurious evidence of complete linkage, *i.e.* zero apparent recombinants.

NAME: ______________________________

STUDENT ID #: ________________________

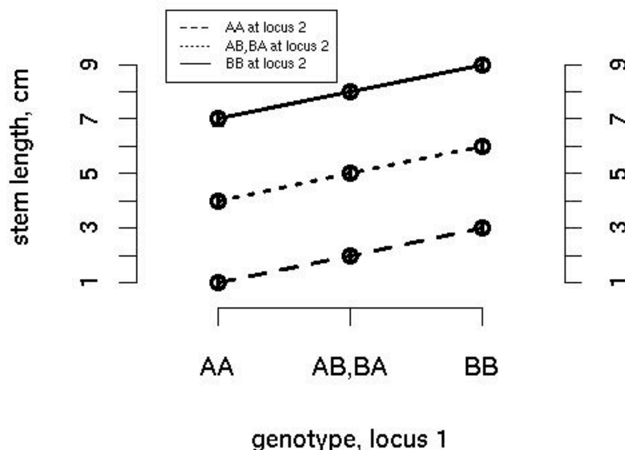
Imagine that you have lots of mouse crosses in the lab and can study a disease, with 10,000 genotyped markers, in multiple pedigrees. In fact, there are so many pedigrees that you have one with five meioses, one with six, one with seven, one with 105 meioses, one with 106, and so on. Imagine that for each such pedigree you see exactly one marker whose inheritance relative to the disease gives zero recombinants. In each case, you don't know whether this is a real indication of a truly causal locus.

In order to satisfy the demands of reviewers for an article you're writing, you need to select only a single pedigree for analysis. Again, any pedigree you choose will have a single marker showing complete linkage. But the reviewers demand that you guarantee a 1% chance that this marker is a false positive. How many meioses do you want in your chosen pedigree in order for the analysis to satisfy this constraint? (Hint: under the assumption that all markers are unlinked, you want to expect 0.01 of them to show complete linkage.) You can leave your answer in terms of logs and/or exponents.

You want to study enough meioses such that the number of markers showing complete linkage (zero recombinants) by chance is 0.01. If you study *x* meioses, the likelihood of a single marker showing complete linkage is 0.5^x. The expected number of markers showing complete linkage is $10,000 * 0.5^x$. Thus,

$10,000 * 0.5^x = 0.01$
$0.5^x = 10^{-6}$
$\log(0.5^x) = \log(10^{-6})$
$x*\log(0.5) = \log(10^{-6})$
$x = \log(10^{-6})/\log(0.5) \sim 20$ meioses

13. (3 points each, 24 total) In class we discussed an F2 cross in which a quantitative trait is controlled by two loci under a purely additive, incomplete dominance model. Imagine that you are studying stem length in green bean plants in such a cross. The grandparents A and B of the cross are homozygotes for the two loci, *i.e.* one has genotype AA/AA and the other BB/BB. The F2's have all 16 combinations of alleles at the two loci. The allele effects are shown below. Note that the *x* values represent genotype at the first locus and the line styles represent genotype at the second locus. *You will not get credit unless you show the computations that allow you to arrive at your answer.*



NAME: ______________________________

STUDENT ID #: ______________________

(a) What proportion of F2's have genotype AA at the first locus and AB at the second?

2/16 (two cells of the Punnett square)

(b) What proportion of F2's have genotype AA at the first locus (and any genotype at the second)?

1/4 (four cells of the Punnett square)

(c) Using the values on the *y*-axes of the plot, what is the average stem length among all F2's with genotype AA at the first locus?

1/4 of these F2's have stem length = 1 cm
1/2 of these F2's have stem length = 4 cm
1/4 of these F2's have stem length = 7 cm
so the average is (1/4)*1 + (2/4)*4 + (1/4)*7 = 16/4 = 4 cm.

(d) Likewise, what is the average stem length among all F2's with genotype AB at the first locus? With genotype BB at the first locus?

AB: (1/4)*2 + (2/4)*5 + (1/4)*8 = 20/4 = 5 cm
BB: (1/4)*3 + (2/4)*6 + (1/4)*9 = 24/4 = 6 cm

(e) Given your answers in (c) and (d), what is the average effect of gaining a single B allele at the first locus?

Increase of 1 cm in stem length.

(f) What is the average stem length among all F2's with genotype BB at the second locus (and any genotype at the first)?

(1/4)*7 + (2/4)*8 + (1/4)*9 = 32/4 = 8 cm

(g) Which has a stronger effect on stem length, variation at locus 1 or locus 2, and why do you say so?

Gaining a B allele at locus 2 is associated with a 3-cm increase in stem length, whereas gaining a B at locus 1 is associated with a 1-cm increase; variation at locus 2 has a stronger effect.

(h) How would the plot look different if gaining a B allele at locus 1 were associated with shorter stem length rather than longer stem length?

The horizontal lines (representing the effect of locus 1 as genotype changes) would slope downward instead of upward.

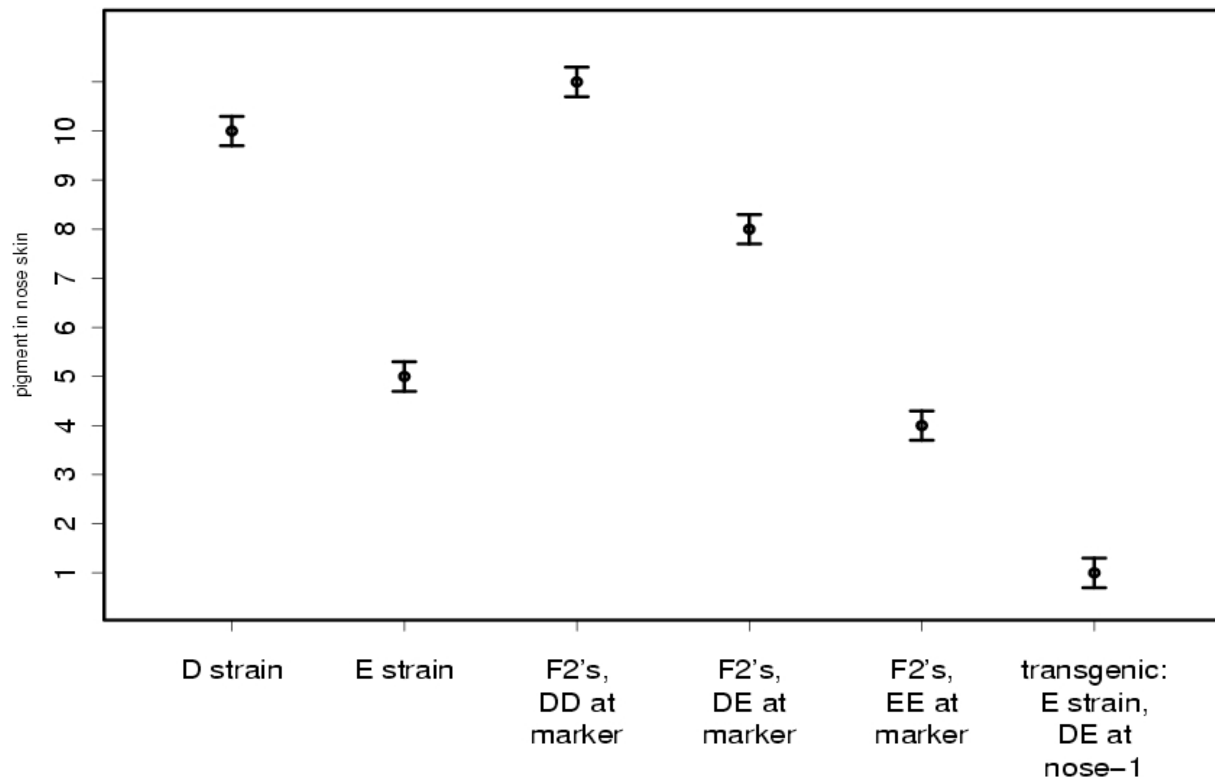
14. You are studying pigmentation in mouse nose skin. Mouse strain D has a pure black nose and strain E has a grey nose. You cross strains D and E and generate F2 progeny, finding that their noses are black, white, and various shades of grey. You genotype the progeny, conduct linkage mapping using nose skin pigmentation as a quantitative trait, and find a strong, significant peak to a marker on mouse chromosome 5. The relation between marker genotype and phenotype in the F2's is

NAME: ______________________________

STUDENT ID #: ______________________________

shown in the middle three columns of the figure below. F2's who are homozygotes for the D allele at the marker have darker noses than F2 heterozygotes, which have darker noses than F2's homozygous for the E allele.

Looking in the region around the marker, among other genes you find one called nose-1, whose sequence is well conserved across mammals and has been shown in previous mouse studies to affect nose skin coloration when deleted. Sequencing shows that both copies of nose-1 in the E strain have a valine amino acid at a position in the protein sequence where the D strain, and all other mammals, have a glycine. You muster your research funds to create a transgenic mouse by allelic replacement. This strain is identical to the E strain but contains, at the endogenous nose-1 locus, a single copy of nose-1 from the D strain, *i.e.* it has the glycine at the variable site in one copy of the genome. Results are shown in the farthest right column of the figure below: the transgenic animal has a nose with very low pigmentation, almost white.



(a) (2 points) What is the effect of introducing a D allele at nose-1 into the E strain background?

Pigment decreases by 4 units.

(b) (2 points) What is the effect of gaining a D allele at the marker among the F2's?

Pigment increases by 4 units.

(c) (6 points) Compare your answers from (a) and (b). Why is this comparison

NAME: ______________________________

STUDENT ID #: ______________________

The net effect of a D allele at the marker is in the opposite direction from the effect of a D allele at nose-1.

(d) (10 points) Without invoking epistasis, dominance, or loci on other chromosomes, propose a hypothesis to explain the puzzle in (c).

If nose-1 were the only gene at the locus that varies between the strains and is causal for pigmentation differences, then the allelic replacement of this gene alone would have agreed with the net effect of the locus shown by the marker. There must be other genes at the locus responsible for the net effect. If you studied each of these genes in isolation, you would see that for the majority of cases, the D allele would be causal for higher pigment levels; the sum total of the effects of all variants in the region should agree with the results in the F2's.

(e) (10 points) What further experiment would you do to test the hypothesis in (d)? (Assume that you have already tested a strain with two copies of nose-1 from the D strain in the E background, as well as strains with one and two copies of nose-1 from the E strain replaced into the D strain background. Don't propose either of these; we already assume there is no epistasis.)

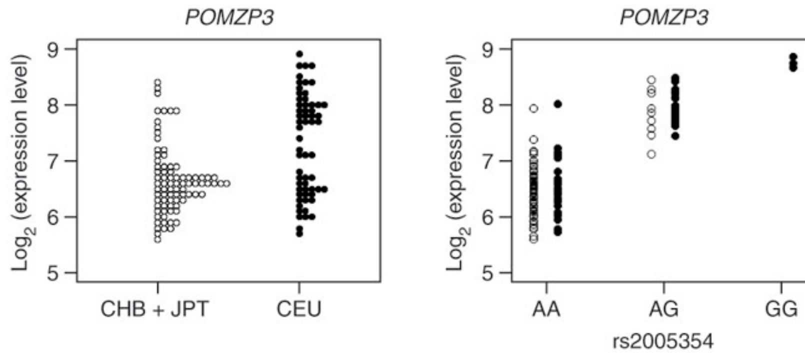
Taking a large chunk of the genome from the D strain—comprising nose-1 plus multiple additional neighboring genes harboring polymorphisms—and introducing this into the E strain background should give a result that agrees with linkage. That is, having the whole locus from the D strain is causal for elevated pigment levels. Also, doing the experiment on one gene from the region at a time will identify which sites from the D strain are causal for higher *versus* lower pigmentation. You could also achieve this with further generations of crossing to give recombinants between the multiple linked genes.

15. (10 points) The current boom in human disease association studies has resulted directly from advances in genotyping technology which allow thousands of markers to be scored in a sample. Before this advance, with only a few hundred markers in hand, an association study was unlikely to succeed even if the primary assumption—that people with the disease all got the same disease allele from the same distant ancestor—were met. Why do you need so many more markers to study a disease using an association study than you would to study the same disease *via* linkage mapping with the same number of subjects?

Alleles at the disease locus are not correlated with alleles at markers far from the disease locus when you study unrelated people. This is because recombination has broken up the relationship between the disease locus and distant markers as the former was passed down to modern-day patients. These patients do not carry the entire ancestor chromosome on which the disease mutation arose; they are likely only to carry a chunk of the chromosome, containing the disease mutation and some sites nearby. In order to boost the likelihood that a marker will fall close enough to the disease locus to show signal in the study, one needs a lot of markers.

16. (10 points) A recent study by Spielman *et al.* used association mapping to search for the genetic determinants of differences between humans in the mRNA expression levels of thousands of genes. The study was done using two populations: CEU (European Americans in Utah) and CHB + JPT (Han Chinese in Beijing and Japanese

in Tokyo). Association calculations were done separately in each population. Results for mRNA expression of the gene POMZP3 are shown below. Levels of this transcript were mapped by association methods to the marker rs2005354 in each population. The left panel shows expression levels of POMZP3 in the two populations. The right panel shows the relationship between POMZP3 levels and genotype at the marker; dark circles represent CEU individuals and unfilled circles represent CHP + JPT individuals.



From the left panel, it is clear that CEU individuals on average express the gene at higher levels than do CHB + JPT individuals (the mean expression level among all individuals in the sample is higher in CEU). Based on the right panel, what is the genetic reason for this population difference? Be clear in your answer, bearing in mind that genotypes given in the right panel are at a marker, not the true causal variant.

The allele of the causal site associated with the G allele at the marker causes higher expression of POMZP3. This allele is rare among CHB + JPT individuals, so their POMZP3 expression is lower.

NAME: ___________________________________

STUDENT ID #: ____________________________

NAME: ______________________________________

STUDENT ID #: _______________________________