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 16 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) _____

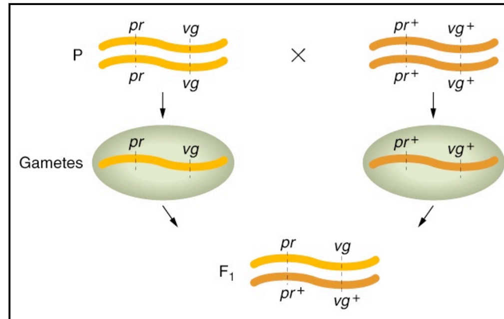
16 (10) _____

TOTAL _____ / 225

NAME: ____

STUDENT ID #: ________________________________

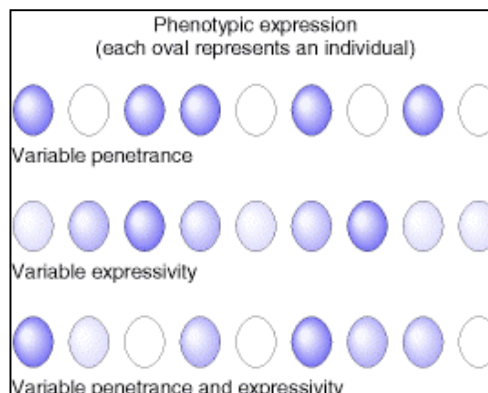
1. (5 points) T.H. Morgan found that two genes – pr and vg (see schematic below) – do not follow Mendel's second law. What specific aspect of meiosis ensures that these two loci follow Mendel's first law?



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)

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.



NAME: ________________________________

STUDENT ID #: ________________________

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?

(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?

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

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?

NAME: ___________________________________

STUDENT ID #: _____________________________

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.

   (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.

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.

NAME: ___________________________________

STUDENT ID #: ____________________________

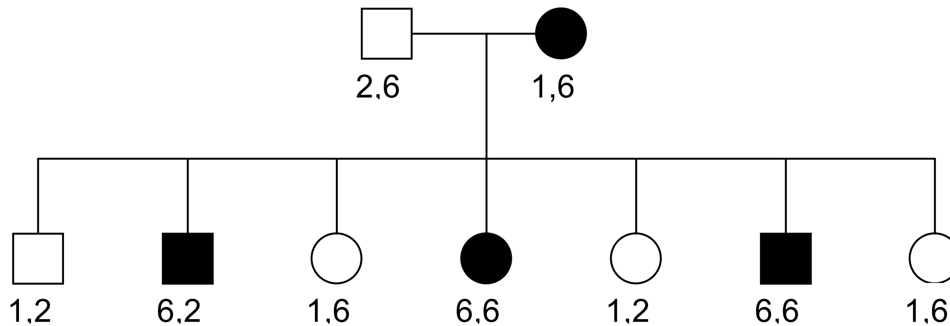
8. Imagine that you are interested in finding the genetic determinants of a severe human disease that strikes in early childhood but does allow patients to live into adulthood. The disease is very rare, indicating that alleles underlying the disease are not present at high frequency in the human population.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

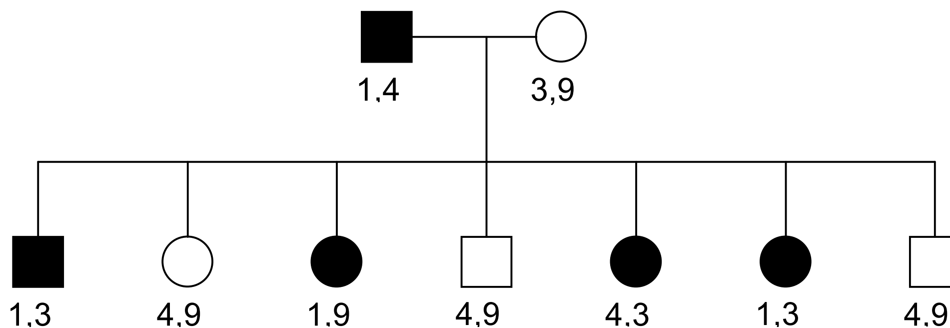
   (a) (5 points) You collect data from six extended families and are planning a linkage study with a dense marker set. The total number of subjects should be plenty for statistical considerations, but no single family is big enough for you to analyze on its own. But one major concern still looms, making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

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

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.



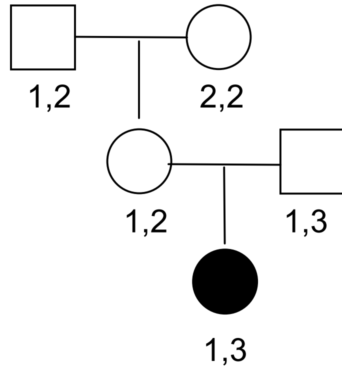


Imagine that you pursue molecular confirmation experiments and find a large gene deletion that is truly causal for the disease, closely linked to this marker. All the affected individuals in both families have the deletion, and all the healthy individuals have a full-length copy of the gene. But the 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.

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.

NAME: ______________________________

STUDENT ID #: ______________________



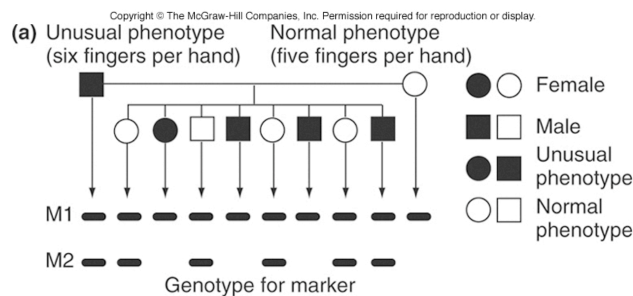
For each individual, determine whether the phase between alleles at the disease locus and alleles at the marker is known. Do not count recombinations—just determine whether there is a single possible arrangement of alleles at the two loci (the haplotype) on each homolog for:

a. Individual II-1

b. Individual II-2

c. Individual III-1

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.

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.

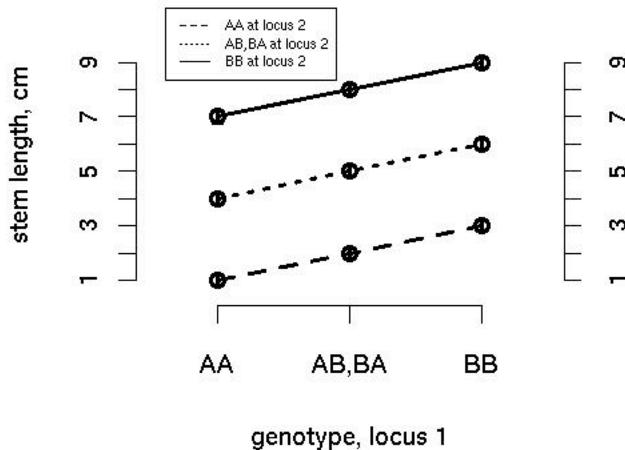
    Imagine that you have lots of mouse crosses in the lab and can study a disease, with 10,000 genotyped markers, in multiple pedigrees. In fact, there are so many pedigrees that you have one with five meioses, one with six, one with seven, one with 105 meioses, one with 106, and so on. Imagine that for 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.

NAME: ______________________________

STUDENT ID #: ________________________

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 the answer.*



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

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

(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?

(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?

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

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

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

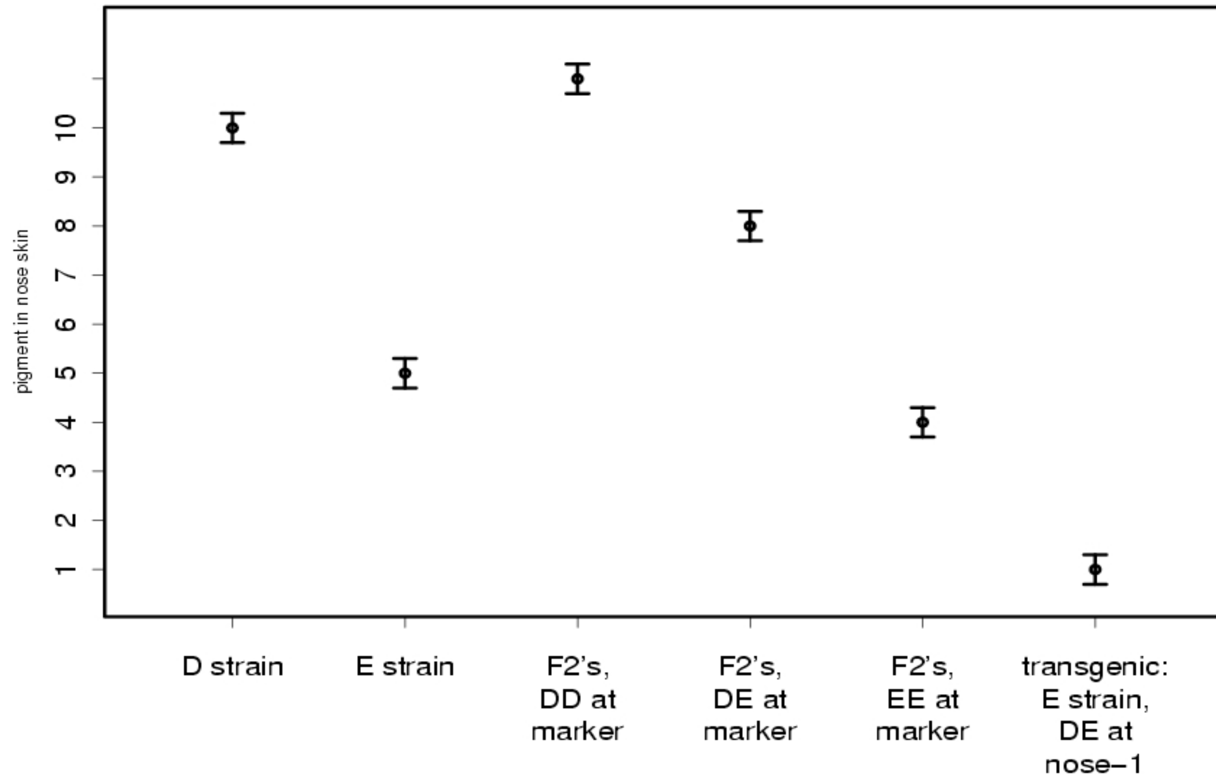
(h) How would the plot look different if gaining a B allele at locus 1 were associated with shorter stem length rather than longer stem length?

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 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.

NAME: ______________________________

STUDENT ID #: ___________________________



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

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

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

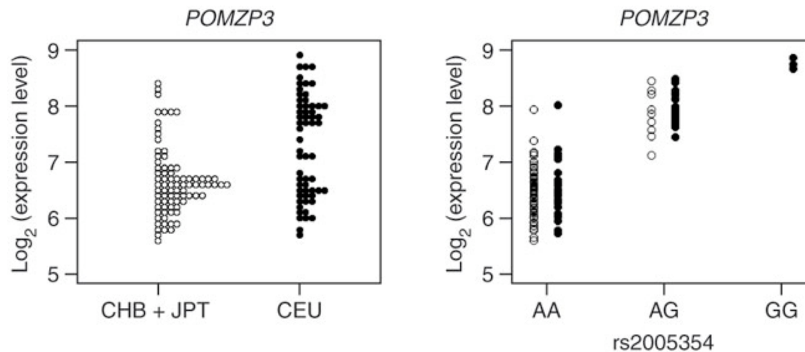
(d) (10 points) Without invoking epistasis, dominance, or loci on other chromosomes, propose a hypothesis to explain the puzzle in (c).

(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.)

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?

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.