Name:_____

Bio393

Genetic Analysis

Final exam

Wednesday June 10, 2015

Graded exams will be available after 9 AM on Friday June 12th outside Cook 3125. If you have any questions about the grading of the exam, return your exam with a written explanation by Monday June 15th. Grade distribution will be available on the course website.

Thank you for a fun quarter. Enjoy your summer break or your next adventure!

Question 1 (4 points):
You are studying a dominant Mendelian disease via linkage analysis and are focusing on a single marker. Two large families have been genotyped at the same marker and scored for the disease.

In Family I, ten offspring are genotyped: eight children inherited the marker locus and disease locus without recombination; two children appear to be recombinants. You test many values of the recombination fraction (theta) and discover that theta = 0.2 gives the maximum odds ratio, which is 6.87 (LOD = 0.837).

In Family II, 20 offspring are genotyped: 17 children inherited the marker locus and disease locus without recombination; three children appear to be recombinants. You test many values of theta and discover that theta = 0.15 gives the maximum odds ratio, which is 223.4 (LOD = 2.34).

To combine data across Family I and Family II, you multiply odds ratios (add LOD scores). The final estimate of the odds of linkage relative to the null as 1534.8 (LOD = 3.17). Explain what is wrong with this calculation.

Question 2 (4 points):
On a petri dish, *C. elegans* can sense and move toward bacterial cells, which the nematodes eat. This behavior is thought to be mediated by small molecules in high concentrations near bacteria because animals will also move toward a drop of “attractant” chemicals like salt and cyclic nucleotides. This behavior is called chemotaxis. Assume that rapid chemotaxis toward salt is always critical for the nematodes to find food. You mutagenize the laboratory strain of *C. elegans* and identify a mutant that chemotaxes more slowly to salt than the wild type. In parallel, a colleague who has been collecting wild *C. elegans* isolates from rotting fruit shows you that one of her wild strains chemotaxes to salt more slowly than the wild-type laboratory strain does.

You do a genetic mapping experiment with each strain separately: the mutant laboratory strain and the wild isolate. In each case, you identify a single polymorphism that causes slow chemotaxis. In which strain would you expect a more dramatic mutation, like a large deletion or severe amino acid change, and why?

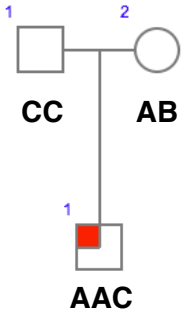
Question 3 (6 points):
Imagine you are doing a genome-wide linkage study in Finnish families looking for the genetic determinants of blood pressure in humans. You have five multi-generational families; each individual is genotyped at 1000 markers and his/her blood pressure is measured. A recent, published study in Icelandic families identified a highly significant locus on chromosome 10 responsible for blood pressure variation. You look through your results and see no significant linkage between the genotype and the disease in your data. Your nearest marker to this locus is 30 cM away.

Give three reasons why you might have failed to find linkage to the chromosome 10 locus. Please explain each reason with one or two sentences.

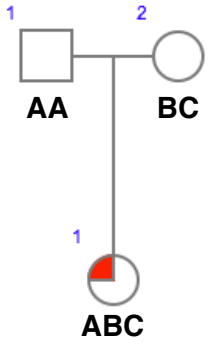
Question 4 (6 points):
Describe three factors that contribute to the ability to detect QTL using genome-wide association studies?

Question 5 (4 points each):
Chromosomal abnormalities cause a large fraction of aborted pregnancies and severe developmental disorders. For the following two parts, please describe what caused the inheritance of the extra chromosome in the affected child?

(a) Down syndrome is caused by inheritance of an extra copy of chromosome 21. A marker on chromosome 21 was genotyped in both parents and child shown below.



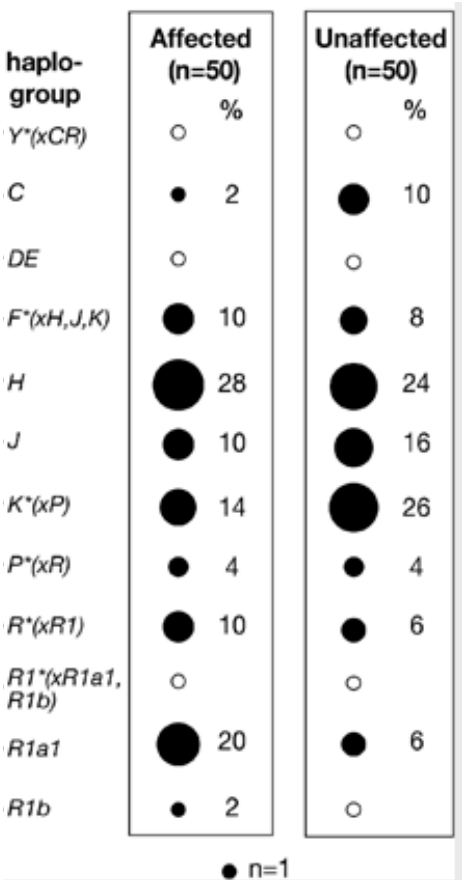
(b) Patau syndrome is caused by inheritance of an extra copy of chromosome 13. A marker on chromosome 13 was genotyped in both parents and child shown below.



Question 6 (3 points each):
(a) What properties of the human genome contribute to the presence of haplotypes?

(b) What properties of human history contribute to the presence of haplotypes?

Question 7 (6 points):
The hairy ears trait has long fascinated geneticists, using evidence from the earliest Italian pedigree in 1907 showing male transmittance to the molecular investigation of southern Indian males in 2004. In the most recent study, 50 affected and 50 unaffected (by visual inspection) southern Indian males were genotyped for markers on the Y chromosome. Twelve distinct Y chromosome haplotypes (haplogroups) were observed in both populations. Please describe whether these data argue for Y linkage and how you arrived at your decision.



Question 8 (16 points):
A psychologist friend of yours heard you studying association analyses for the Bio393 final. He suggested that you help to genetically map belief in a deity for his senior thesis. He collected DNA and genotyped 1000 Northwestern students who attend a religious service weekly (cases) and 1000 students from the atheist club (controls). One marker is nearby the widely purported "faith gene" with the following distribution in genotypes among the students: Cases = 345 AA, 575 AG, 80 GG; Controls = 275 AA, 500 AG, 225 GG.

(a, 4 points) Fill out the 2x2 contingency table for this association test for him.

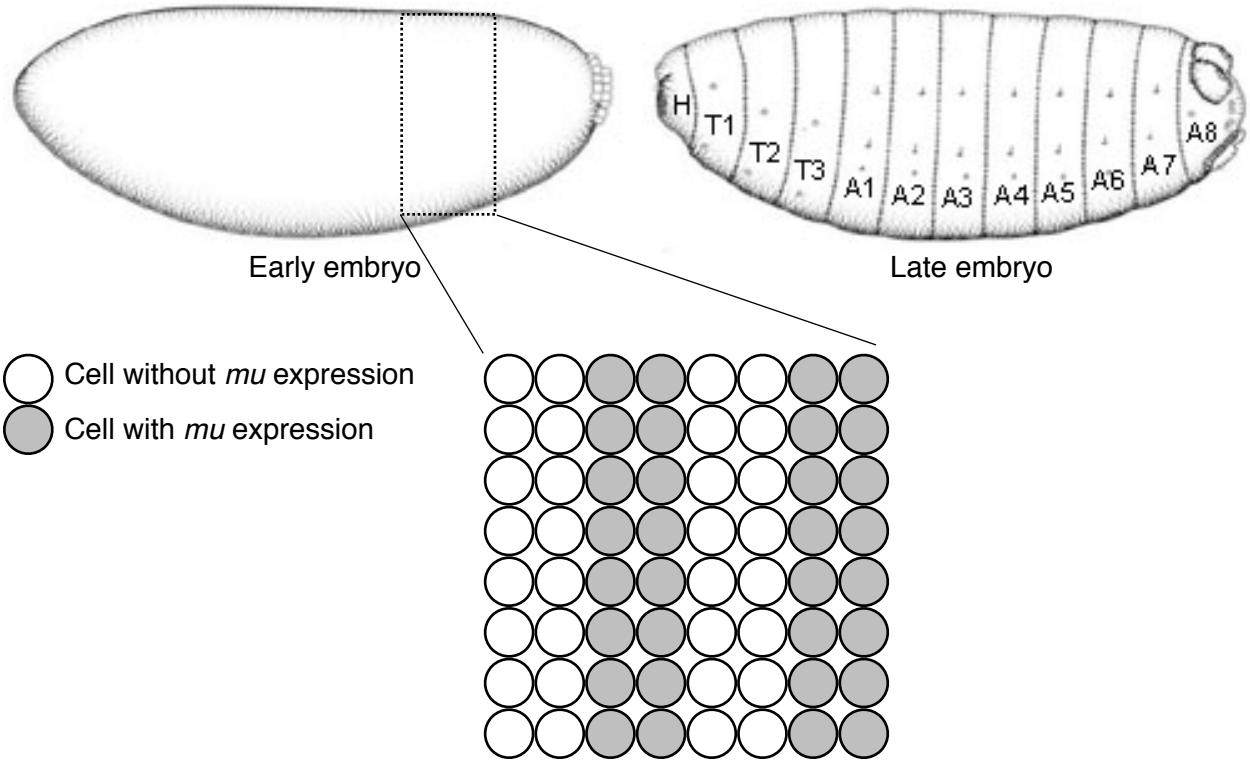
		Cases	Controls
Genotype	A		
	G		

(b, 4 points) This chi-squared test returned a *p*-value of 7.2E-12, which is significant given multiple testing correction. Why do we have to do multiple testing correction?

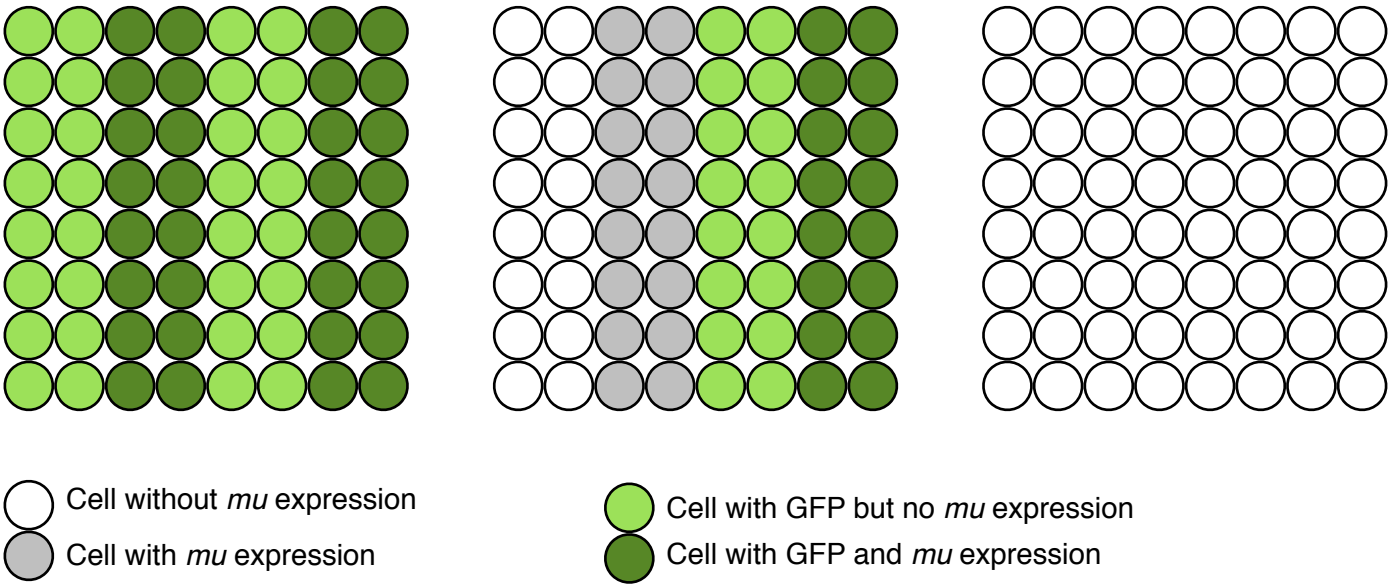
(c, 4 points) The group of students genotyped and phenotyped for their beliefs represented a diverse collection of undergraduates. Why might this statistically significant association be a spurious result of experimental design?

(d, 4 points) The HapMap Consortium found that the G allele at this marker is most often found in the African population. Do you think that this allele could have been introduced into our species from interbreeding with Neaderthals? Why or why not?

Question 9 (28 points):
You are studying pattern formation of the *Drosophila* embryo. Previous studies established that certain genes are expressed in pattern-specific ways, even before the time that patterns can be distinguished morphologically. One such gene is called *mixed up* (*mu*).



Mutations in the gene *disorganized* (*dis*) cause abnormal morphological patterning in the embryo. To determine whether *dis* affects the pattern of *mu* expression, you create genetic mosaics by transplanting nuclei between wild-type and *dis* embryos. Wild-type cells are labeled using the *GFP* gene with a constitutive promoter driving gene expression. In three representative mosaic embryos, you see the patterns of *mu* and *GFP* expression indicated below.



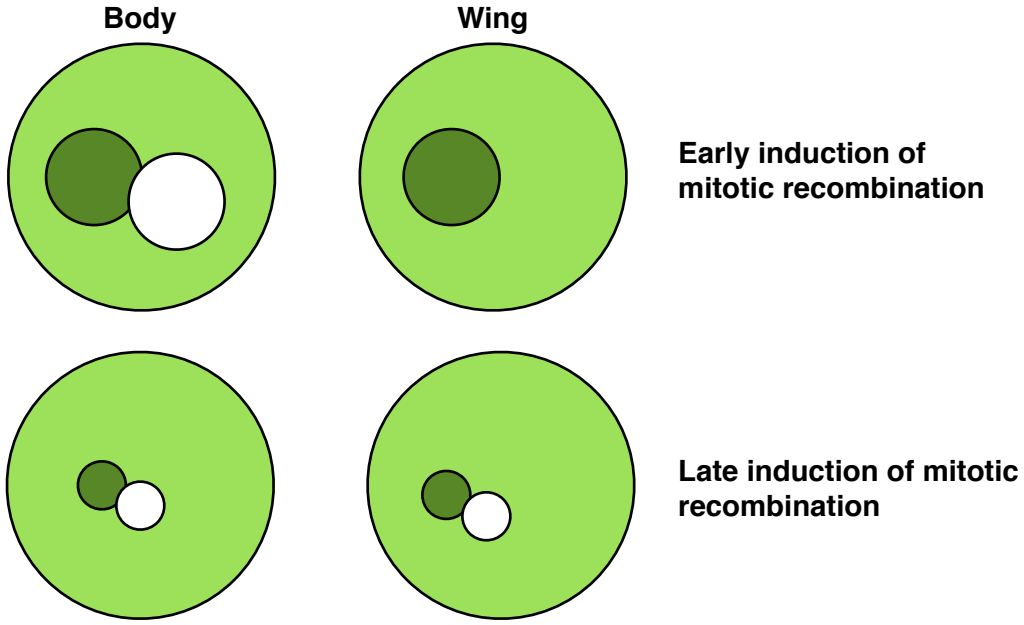
(a, 8 points) What conclusions can you draw about the role of *disorganized* in controlling *mixed up* gene expression?

(b, 6 points) Other data suggest that the *confused* (*cf*) family of genes might regulate *mu* gene expression. Two *cf* family genes exist (*cf1* and *cf2*). Previous genetic screens for embryonic patterning mutants failed to identify any *cf* mutants. Assuming the EMS screen was saturated, please suggest three other explanations for why *cf* mutants were not found.

(c, 6 points) Using a loss-of-function allele of *cf1*, you perform a genetic screen and identify a new mutation that causes abnormal patterning in the embryo indistinguishable from that caused by a lack of *mu* expression. Without describing specific crosses, describe how you would determine if the effects of this new mutation depended upon the presence of a *cf1* mutation.

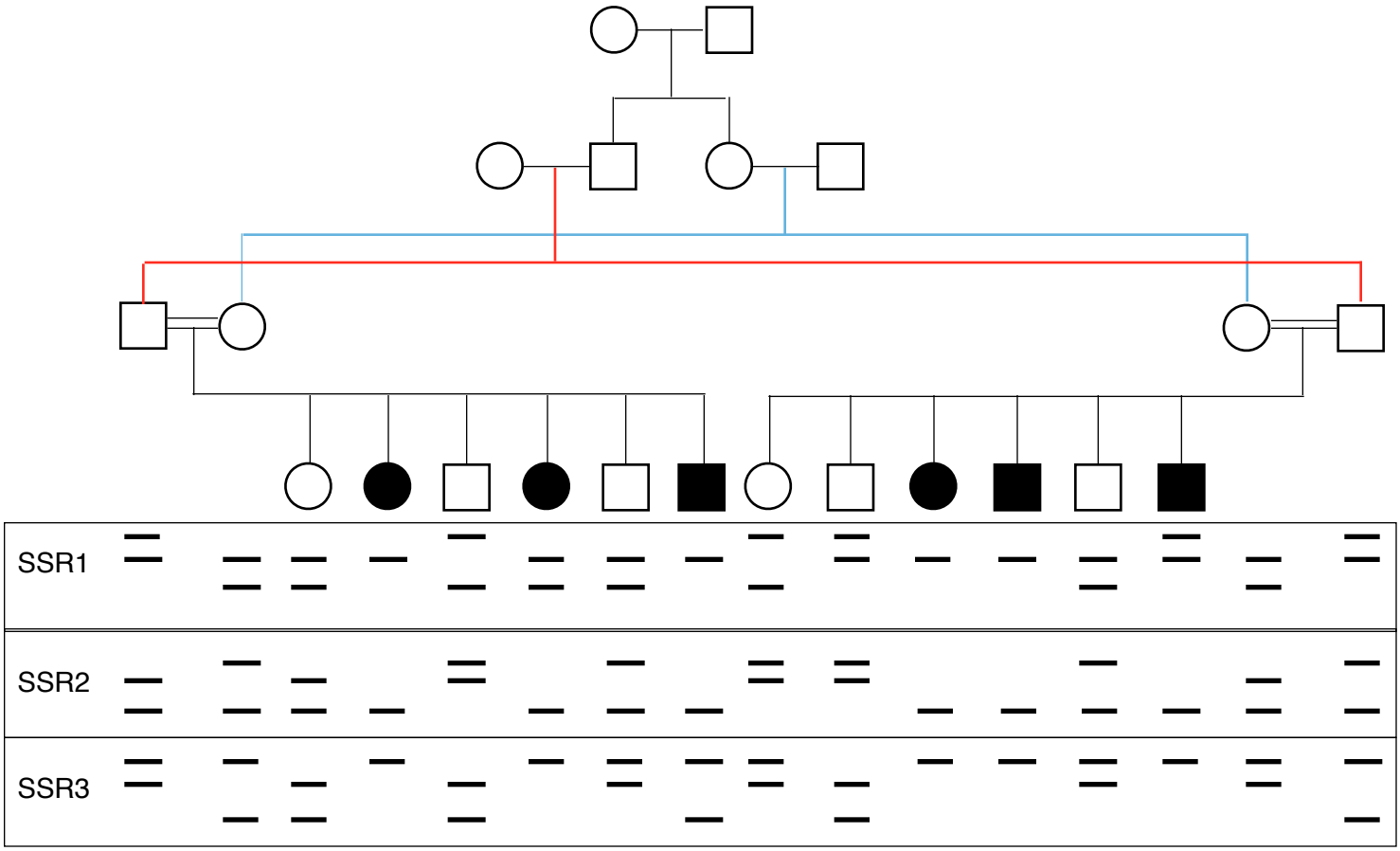
(d, 8 points) Assume your new mutation is in *cf2*. Both *confused* genes act during wing development. You wish to determine the time of action of the *cf* genes during wing formation. You induce mitotic recombination either early or late in wing development, using a strain that is heterozygous for both *cf1* and *cf2* along with construct that ubiquitously expresses GFP near both *confused* genes. Assume that mitotic recombination does not separate these three alleles.

Genotype		
+	*cf1*	*cf2*
GFP	*cf1+*	*cf2+*



Top pictures represent sections from either body or wing in which mitotic recombination was induced early. Bottom pictures represent sections from either body or wing in which mitotic recombination was induced late. The shading of each spot denotes the level of GFP expression: darker is high expression, lighter is moderate expression, and white represents no GFP expression. Propose an explanation for these patterns. What do the data suggest about the time of action for the *confused* genes?

Question 9 (28 points):
You are studying a family with a rare genetic disease. The pedigree is given below along with the segregation of alleles for three linked simple sequence repeats. The order of the three markers is SSR1, SSR2, and SSR3. The double horizontal lines denote consanguineous marriages. The children of generation II are connected with red or blue lines to clarify relationships.



Individuals III-1 and III-4 inherit the middle allele of SSR1, the bottom allele of SSR2, and the top allele of SSR3 from their father individual II-2. Individuals III-2 and III-3 inherit the middle allele of SSR1, the bottom allele of SSR2, and the top allele of SSR from their mother individual II-3.

(a, 6 points) Circle the progeny in generation IV that are informative.

(b, 16 points) For each SSR, calculate the LOD score for the best theta given the data above. Remember that the LOD equation is the log of the odds ratio that you see linkage between a marker and the disease-causing allele. For a recessive disorder, each chromosome must be evaluated. Think of it as every affected individual as the likelihood of two chromosomes coming together. Use the form of the odds ratio (below) to calculate LOD score and show your work on the following page.

$$\text{Odds ratio} = \frac{\text{Likelihood(theta)}}{\text{Likelihood(0.5)}} = \frac{0.5*((1\text{-theta})^{P} * (\text{theta})^{R} + (1\text{-theta})^{R} * (\text{theta})^{P})}{(0.5)^{P+R}}$$

P = Number of parental chromosomes
R = Number of recombinant chromosomes

SSR1

SSR2

SSR3

(c, 4 points) Have we established linkage between the disease gene and any of the three SSR markers? Why or Why not?

Question 10 (10 points):
A patient comes in to your medical office presenting his 23andme results and an extreme sense of worry. Both his father and his grandfather died of prostate cancer, and he is worried that his days are numbered given his genome results.

(a, 5 points) Please calculate his risk of prostate cancer given the results below.

NAME	AVG. RISK	COMPARED TO AVERAGE
Atrial Fibrillation	27.2%	1.25x
Prostate Cancer ♂	17.8%	1.33x
Gallstones	7.0%	1.58x
Exfoliation Glaucoma	0.7%	2.90x
Ulcerative Colitis	0.77%	1.30x

(b, 5 points) Further down on the form you see that he has reduced risk for Alzheimer's disease. He is completely confused how 23andme determined that result. Please briefly explain in words how his risk can be less than 1x and how they calculated it.

NAME	YOUR RISK	AVG. RISK	COMPARED TO AVERAGE
Gout	17.1%	22.8%	0.75x
Venous Thromboembolism	9.0%	12.3%	0.73x
Alzheimer's Disease	4.3%	7.2%	0.60x
Age-related Macular Degeneration	3.1%	6.5%	0.48x
Melanoma	2.2%	2.9%	0.75x