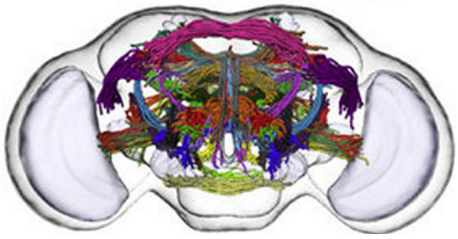
Bio393: Genetic Analysis **Name:**_____
Problem Set #4
Due on Friday, June 2, 3 PM

Question 1:
Drosophila extend their proboscis to eat or drink. This behavior is controlled by a complex system of connected neurons and muscles. The four images below show the process of proboscis extension (temporally from left to right).

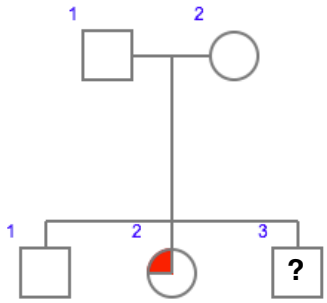


(a) You identified a mutant with a recessive phenotype of much reduced proboscis extension that you named *hungry*, and you are interested to figure out in which set of neurons shown in different colors (left) the gene regulates proboscis extension. Using a series of promoters that express gene products in these regions, please describe how you would determine in which set of neurons the *hungry* gene acts. Also, describe any caveats.

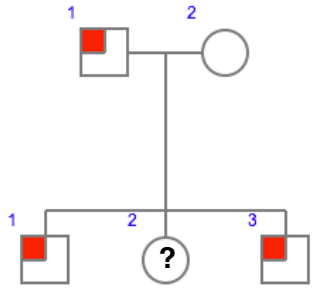
(b) You determine that *hungry* acts in the purple set of neurons to regulate proboscis extension. Using the same promoter and optogenetic tools, describe how you would use specific tools to determine if those neurons inhibit or promote proboscis extension.

Question 2:
Each of the families below exhibits a different, extremely rare genetic disorder. Individuals expressing the trait (the disorder) are indicated by symbols with red sections. Assume that no new mutations have arisen in any of the individuals shown. Consider the following possible modes of inheritance: (i) X-linked recessive with complete penetrance, (ii) autosomal recessive with complete penetrance, (iii) autosomal recessive with 70% penetrance, (iv) autosomal dominant with complete penetrance, (v) autosomal dominant with 70% penetrance. For each pedigree state which, if any, of these five modes of inheritance are not possible. For the modes of inheritance that are possible, calculate the probability that the individual indicated by a "?" is affected.

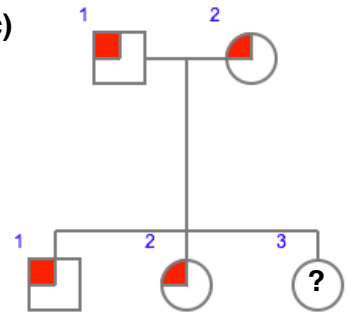
(a)



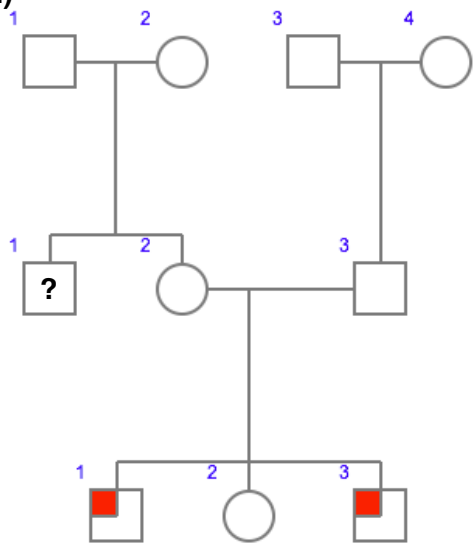
(b)



(c)



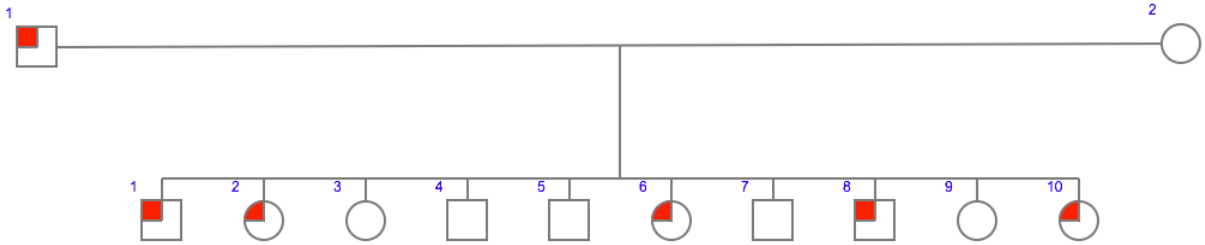
(d)

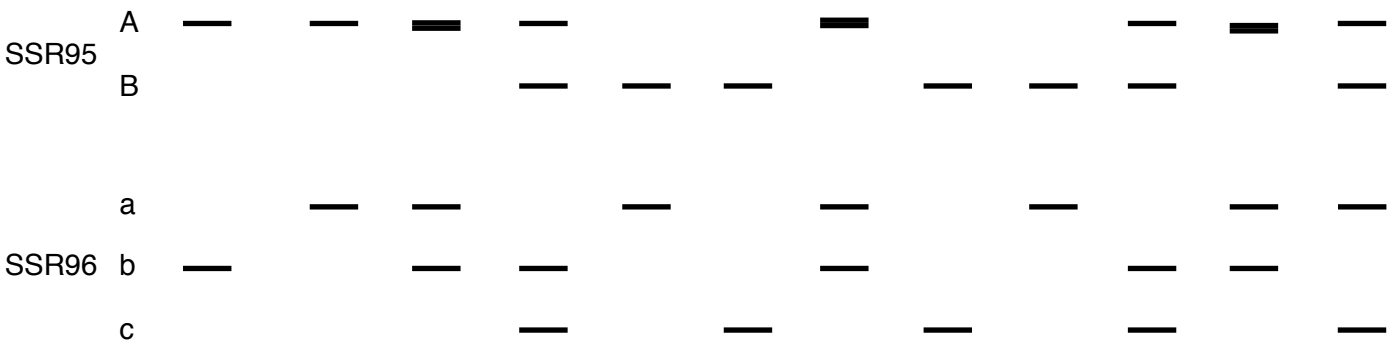


Question 3:
You set out to genetically map color blindness with respect to short-sequence repeat (SSR) markers. Color blindness shows X-linked recessive inheritance and therefore is usually found in males. However, the mutant allele frequency is sufficiently high that colorblind females do occur.

Alleles: + (normal) cb (associated with color blindness)

Here is a family in which some individuals are affected:





(a) Diagram the two possible phase relationships between the SSR95 and SSR96 alleles in the mother.

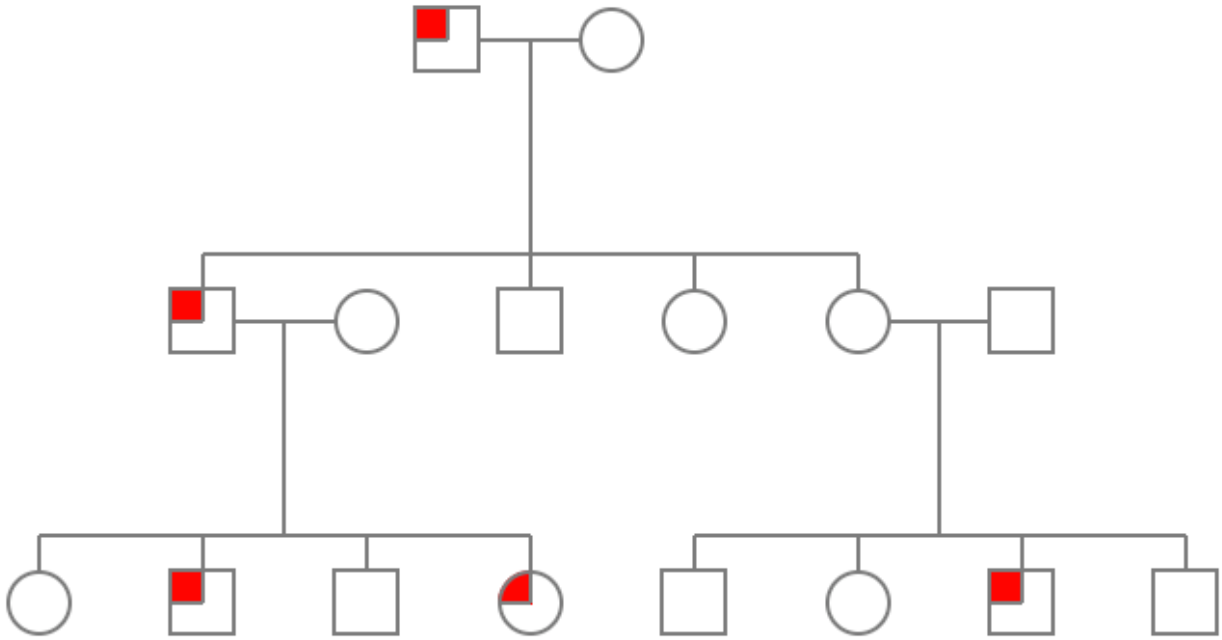
(b) Calculate the LOD score for linkage at $\theta = 0.1$ between SSR95 and SSR96 in this family.

(c) Identify a value of θ at which this family will yield a higher LOD score for linkage between SSR95 and SSR96. Calculate the LOD score for linkage between SSR95 and SSR96 at that new θ value.

(d) Diagram the two possible phase relationships between the SSR95 and color blindness alleles in the mother.

(e) Calculate a LOD score for linkage at $\theta = 0.1$ between SSR95 and color blindness in this family.

Question 4:
Syndactyly is a rare genetic condition inherited as an autosomal dominant trait. Unusually however, a person who has the defective allele responsible for syndactyly (N) does not always express the trait. The diagram below shows a pedigree of a family with syndactyly.



You find no history of syndactyly in the ancestors of individual I-2, II-2, and II-6. Assuming no new mutations exist, explain why or why not it is reasonable to conclude the following:

(a) Individual II-5 has the genotype nn at the syndactyly locus

(b) Individuals II-4 and II-5 may have the same genotype at the syndactyly locus

(c) Individual II-2 has the genotype Nn at the syndactyly locus

(d) Individuals III-2 and III-7 have different genotypes at the syndactyly locus

Question 5:
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 a marker allele and a disease-causing allele 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 a marker allele and a disease-causing allele 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).

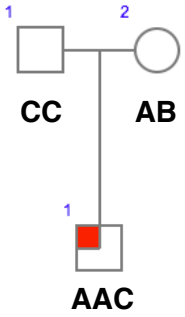
To combine data across Family I and Family II, you multiply odds ratios (add LOD scores). The final estimate of the odds of linkage relative to the null as 1534.8 (LOD = 3.18). Explain what is wrong with this calculation.

Question 6:
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.

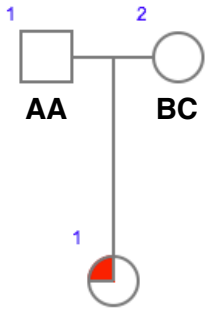
Give three reasons why you might have failed to find linkage to the chromosome 10 locus. Please explain each reason with no more than one to two sentences.

Question 7:
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 8:
You are studying a rare recessive disease that you have mapped approximately by linkage to simple sequence repeat (SSR) markers. In an effort to localize the disease locus more precisely, you decide to look for linkage disequilibrium (LD) with respect to two dimorphic DNA-based markers (designated A and B) known to be in the vicinity of the disease gene. You first examine a relatively isolated Scandinavian population in which the frequencies of alleles A1 and A2 are 0.9 and 0.1 respectively, and the frequencies of B1 and B2 are both 0.5. By examining the DNA from individuals in the population who have the disease it is possible to determine the frequency of each haplotype, as shown in the table below.

Haplotype	Number of individuals with the disease
A1 B1	10
A1 B2	90
A2 B1	1
A2 B2	10

(a) (i) What can you say about possible linkage disequilibrium between each of the markers and the disease causing allele in this population? (ii) Assume the disease causing allele arose after both of the markers (A and B) were present in the population. Which of the two DNA-based markers is likely to be closer to the disease locus? (iii) Assuming that the disease allele arose only once in this population, what can you say about the haplotype context in which the original disease mutation arose?

(b) Next you examine the genotypes of individuals with the same disease in a large African population. In this population the frequencies of alleles A1 and A2 are both 0.5, and the frequencies of B1 and B2 are also both 0.5. The frequencies of the each haplotype for individuals with the disease in the African population are shown in the table below.

Haplotype	Number of individuals with the disease
A1 B1	26
A1 B2	24
A2 B1	28
A2 B2	22

Give two different explanations for why the linkage disequilibrium results differ between the African and Scandinavian populations.

Question 9:
You are running a case-control GWAS for Type 2 Diabetes. Of the 500,000 variants you test, one variant (rs4514, which has 2 alleles, A and G) near the *sweetums* gene has good separation between cases and controls. You have 1000 cases, (480 of which are AA, 400 are AG, and 120 are GG at rs4514) and 1000 controls, (360 of which are AA, 440 are AG, and 200 are GG at rs4514).

(a) Using a chi-squared test, what is the p-value of the association of these alleles with the disease.

(b) Given that you did 500,000 tests, what is your (Bonferroni) corrected threshold for p-value significance (initial α=0.05)? Does the rs4514 variant pass "genome-wide significance" for association with Type 2 Diabetes?

(c) What is the odds ratio of this variant in a risk for Type 2 Diabetes?