Name:_____

Bio393

Genetic Analysis

Final exam

Thursday June 8, 2017

Graded exams will be available after 4 PM on Friday June 9th outside Cook 3125. If you have any questions about the grading of the exam, return your exam with a written explanation by Monday June 12th at 9 AM. The grade distribution will be available on the course website.

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

Question 1 (5 points):
Recurring behavioral disorders were observed in some male members of a large pedigree extending over several generations. The males were mildly mentally retarded and, especially when under stress, were prone to repeated acts of aggression, including sex offenses, attempted murder, and arson. An X-linked gene, MAO, coding for the enzyme monoamine oxidase, which participates in the breakdown of neurotransmitters, was found to be defective in the affected men in this pedigree. Other researchers found abnormal levels of monoamine oxidase in some unrelated men with similar behavioral issues, even though the MAO gene was not defective in these cases. Does this evidence support the hypothesis that defective monoamine oxidase is responsible for the behavioral disorder? Please explain your answer.

The evidence is correlative within one population but not in the other population. It is possible that in the large family with the defective MAO gene that MAO is the causal gene underlying this behavioral trait. However, again, the evidence is only correlative. One would need to connect loss of MAO gene function using genetic experiments to this behavior. In humans, this goal is difficult if not impossible.

Question 2 (5 points):
Edward B. Lewis, Christianne Nusslein-Volhard, and Eric Weichaus shared the 1995 Nobel Prize in Physiology or Medicine for their work on the developmental genetics of *Drosophila*. In their screen for developmental genes, Nusslein-Volhard and Weichaus initially identified 20 lines bearing maternal-effect mutations that produced embryos lacking anterior structures but with the posterior structures duplicated. When Nusslein-Volhard mentioned this result to a colleague, he was astonished to hear that mutations in 20 genes could give rise to this phenotype. Explain why his astonishment was completely unfounded and showed a failure to understand genetics.

Nusslein-Volhard never said that it was 20 mutations in 20 different genes! The 20 lines that they recovered from their screen likely have independent mutations, but those mutations could be in the same genes. Her colleague probably does not understand how mutagenesis works (random generation of mutations throughout genomes).

Question 3 (25 points):
You are interested in vertebrate eye development and function. Therefore, like any good geneticist, you find the most tractable system to study your question – zebrafish.

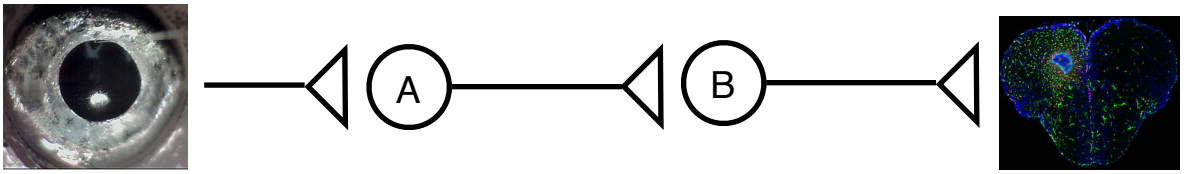
(a, 5 points) You perform a mutant screen for fish that are unable to see certain colors and obtain three mutants in three separate complementation groups. Explain whether your screen is saturated and how you would know.

The screen is not saturated because we only have single mutations in single genes. In order to be saturated, we would expect to identify many independent mutations in the same gene.

(b, 6 points) Over the next three years, you map and clone the three genes. You chose zebrafish because of the plethora of genetic tools available. Using promoters that express your gene of interest in the different parts of the zebrafish retina, describe an experiment to test for the function of the *cb* gene in the blue, magenta, red, orange, or green neurons shown in the retina below.

In colorblind mutant fish, we use transgenesis to introduce different promoters that drive the wild-type copy of the colorblind gene in different neuron types. If the colorblind gene can rescue the mutant phenotype of the colorblind mutant when expressed in a specific neuron type, then the fish will have normal vision. The caveat of this experiment is that overexpression could ectopically bypass the colorblind mutation in certain neuron types.

(c, 14 points) Using channelrhodopsin to manipulate neuronal activity in specific parts of the zebrafish, you would like to determine where and how your three genes act. You perform the following experiments and measure visual activity. Write out where *cb* and *rb* act in the rudimentary neuronal circuit below and the reasoning for your conclusions.



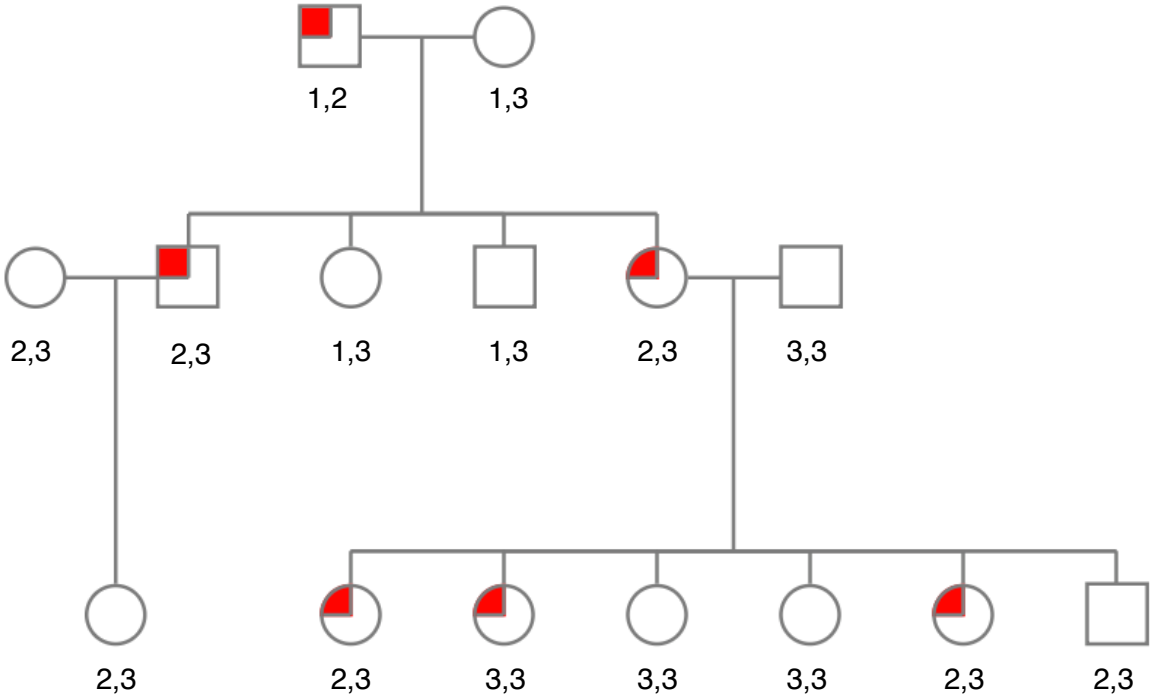
Eye **Brain**

Channelrhodopsin stimulation	Genetic background	Visual activity
None	WT	YES
A neuron	WT	YES
B neuron	WT	YES
None	*colorblind*	NO
A neuron	*colorblind*	NO
B neuron	*colorblind*	YES
None	*red-blind*	NO
A neuron	*red-blind*	YES
B neuron	*red-blind*	YES

Activating the B neuron bypasses the broken circuit in colorblind mutant fish, indicating that the colorblind gene (cb) must act previous to the B neuron in either the A neuron or in the eye.

Activating the A or B neurons bypasses the broken circuit in red-blind mutant fish, indicating that the red-blind gene (rb) must act previous to both the A and B neurons likely in the eye.

Question 4 (30 points):
A rare dominant disorder has the following inheritance pattern. Individuals marked with red have the disorder. You would like to test linkage between a marker (with three different alleles) and the disorder-causing allele (D). Choose a theta that will maximize the LOD score for this pedigree and fill in the equation for the LOD score calculation.



We have three families where we can evaluate linkage. Generation II with individuals II-2 through II-5 are informative, but we do not know the phase of the parents. Generation III is interesting. Individual III-1 is not informative so does not go into the calculation. Individuals III-2 through III-7 are informative and we know the phase of parent II-5. We will add two likelihood equations at the same theta to get the best LOD score for this pedigree. We have 10 informative gametes and two of the offspring are recombinants (III-3 and III-7), so the best theta is 2/10 or 0.2).

$$LOD_{0.2} = \log_{10} \frac{0.5*((1-0.2)^4 * (0.2)^0 + (1-0.2)^0 * (0.2)^4)}{(0.5)^4} + \log_{10} \frac{(1-0.2)^4 * (0.2)^2}{(0.5)^6}$$

Question 5 (6 points):
A patient comes into your medical office presenting his 23andme results and an extreme sense of worry. He is completely confused about how 23andme determined that he has a reduced risk for gout, especially because he loves fatty foods. 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

23andme calculated his gout risk by looking at his genome-wide genotype. This individual shares haplotypes (markers) with individuals that have gout less often than the average person.

The alleles present in an individual might provide some protective benefit for certain diseases.

Question 6 (6 points):
Circle the correct answer.

(a) Which type of variant is <u>easiest</u> to map using family-based analyses?

1 - Rare variants with variable penetrance
2 - Common variants with variable penetrance
3 - Rare variants with high penetrance
4 - Common variants with high penetrance

(b) Which type of variant is <u>easiest</u> to map using population-wide analyses?

1 - Rare variants with variable penetrance
2 - Common variants with variable penetrance
3 - Rare variants with high penetrance
4 - Common variants with high penetrance

Question 7 (17 points):
Crohn's disease affects about 0.001% of the European population. Given the increased prevalence of the disease in identical twins as opposed to fraternal twins, you suspect that this disease has a genetic cause. You perform a genome-wide association in a population of one million Europeans to identify genetic markers for Crohn's disease.

(a, 5 points) Explain why you studied a single population (Europeans) and did not take a mixed sample of individuals from various countries of origin.

A mixed sample of individuals from various countries of origin would confound the association mapping approach because variants unique to each country or region would correlate with the disease of interest by chance. By focusing solely on the European population, variants that can define other countries or regions of origin will not be considered. This approach could identify variants that correlate with Crohn's disease in European ancestry, but these variants might not be correlated in other human populations. Additionally, different loci could cause the disease in various populations, complicating the ability of a GWA to detect one significantly correlated marker in a mixed population.

(b, 6 points) You performed the association mapping and found that the T allele at a SNV, *rs4077515*, is highly correlated with the disease. However, the ratio of the T to the other allele in the control population is approximately 1:1. Explain how individuals with the T allele in the control population may not be affected by the disease.

Individuals with the T allele in the unaffected population could not have the disease because the disease could be complex with many environmental and genetic effects. For example, an intergenic suppressor could be present in unaffected individuals, which reduces the likelihood of disease in those people. Also, the disease allele could have arisen in an individual with the T allele haplotype, but other individuals have the T allele without the disease-causing allele. The two are linked, but the T allele itself may not be causative. Finally, individuals in the control population with the T allele could have the disease-causing allele but low penetrance could make these individuals appear unaffected.

You decide to study 2000 individuals from your earlier association mapping that carry the T allele at *rs4077515* (1000 of which have Crohn's disease and 1000 of which do not have Crohn's disease). You identify a new SNV that is correlated with Crohn's disease in this smaller population. Of the affected individuals, 45% have the A allele and 55% have the G allele at the new SNV site. Of the unaffected individuals, 95% have the A allele and 5% have the G allele at the new SNV site.

(c, 6 points) Fill in the contingency table to perform a chi-squared test for the new SNV in this smaller population.

	cases	*controls*
A	*450*	*950*
G	*550*	*50*

Question 8 (6 points):
Genome sequencing of a Melanesian population revealed a hominid (but not *H. sapiens*) region in their genomes. This region is not Neanderthal nor Denisovan in origin. If you were to sequence the genomes of billions of people and only find this mysterious region in the Melanesian population, what conclusion could you make?

This hominid genomic region found only in the Melanesian population suggests that long ago this hominid species interbred with ancestral humans. This interbreeding event could have occurred before the ancestors of Melanesians colonized Melanesia and only Melanesians with this genomic interval survive today, potentially because of some fitness advantage. Alternatively, the interbreeding happened only on the Melanesian islands. Over many generations, the part of the genome from this ancient hominid recombined with human sequences and got progressively smaller.