

Name _____ KEY _____

S.I.D. _____

MCB 104 MIDTERM #2 March 18, 2014

*****IMPORTANT REMINDERS*****

Print your name and ID# on every page of the exam.
You will lose 0.5 point/page if you forget to do this.

If you need more space than is available on a page, continue your answer on the back of the same page. The pages will be separated for grading, so no points will be given for answers continued on the back of a different page.

This is a closed book, closed note exam. No calculators, phones or any electronic device is allowed.

Look through the entire exam before starting. You should have 9 numbered pages, including this cover page. You do not have to start with Question 1. Read each question entirely before beginning. Write legibly. Show all of your work for formulas/math questions.

(Do not write below this line)

PAGE 2	_____ /14 total
PAGE 3	_____ /12 total
PAGE 4	_____ /12 total
PAGE 5	_____ /12 total
PAGE 6	_____ /11 total
PAGE 7	_____ /12 total
PAGE 8	_____ /14 total
PAGE 9	_____ /13 total

Total Score _____ /100

Name _____ KEY _____

S.I.D. _____

1. (4 points) QTL mapping and GWAS are both genetic methods used to try to link genotype and phenotype. What do QTL and GWAS stand for, and in one sentence, how do they differ in what group of DNAs are studied? underlined terms are required for points

Quantitative trait locus (or loci). Genome Wide Association Studies (1 point each). QTL mapping studies DNAs from a genetic cross or pedigree (1 point), while GWAS studies DNAs from a population (1 point).

2. (3 points) Comparing QTL mapping and GWAS, are patterns of linkage equilibrium/disequilibrium between markers on the same chromosome similar or different? Answer and in one sentence, explain your answer.

Different (1 point). For markers on the same chromosome, regions of linkage disequilibrium are much bigger in QTL mapping than in GWAS (2 points).

3. (3 points) Comparing QTL mapping and GWAS, are patterns of linkage equilibrium/disequilibrium between markers on different chromosomes similar or different? Answer and in one sentence, explain your answer.

Similar (1 point). Markers on different chromosomes would be expected to be in linkage equilibrium in both QTL mapping and GWAS (2 points) (due to Mendel's second law of independent assortment).

4. (3 points) In a genome assembly, which (on average) are longer (in total number of base pairs), scaffolds or contigs? Answer, and then in one sentence explain your answer.

Scaffolds are longer (1 point), because scaffolds consist of contigs connected by paired end reads (2 points). Full credit if worded differently but answer that scaffolds are made up of contigs.

5. (1 point) When you sequence paired end reads from a genomic DNA library or from a cDNA library what is the source of the DNA sequences that the sequencing primers are complementary to (i.e. to what DNA sequences do the sequencing primers hybridize)?

the vector (1 point) (full credit for universal sequencing primer)

Name _____ KEY _____

S.I.D. _____

6. (3 points) In no more than two sentences, explain the difference between a contig and a scaffold.

(1 point each for accurate contig and scaffold definition, 1 point for correctly described paired end reads connecting contigs into scaffolds.) A contig is sequence assembled from contiguous or overlapping DNA sequence. A scaffold uses paired end read information (where one end of a clone aligns to one contig and the other end of the same clone aligns to another contig) to join adjacent contigs into scaffolds.

7. (3 points) What does EST stand for and how would you generate the molecules for a library to collect ESTs from? Your answer should mention two different types of nucleic acid and one enzyme covered in lecture.

Expressed Sequence Tag (1 point). mRNA can be reverse transcribed into complementary DNA (cDNA) (1 point) using reverse transcriptase (1 point).

8. (3 points) How can aligning EST reads to genomic DNA sequence reveal the positions of introns?

If an EST read aligned perfectly to two adjacent stretches of genomic DNA sequence, but was missing a stretch of genomic DNA between these stretches, that would reveal an intron (in between two exons, which are the regions that align). Full credit if correctly explain that if a single EST read aligns to two regions of the genomic DNA with a missing gap in between, then the missing gap is the intron.

9. (2 points) Most GWAS studies have found mutations in exons that affect the amino acid sequence of proteins. Based upon what you learned in lecture, is this statement true or false? Answer, and then explain your answer in no more than one sentence.

False. Most GWAS studies have found association between non-coding regions (intergenic or intronic being most common) and human disease. No credit if just said false without correctly answering about non-coding regions usually being associated.

10. (1 point) Many SNPs tested for association with disease, even in very large samples of thousands of cases and controls, show no evidence for association. What would be the odds ratio for SNPs with zero evidence of association?

One

Name _____ KEY _____

S.I.D. _____

11. (4 points) In the United States, about 1 baby every 10,000 births is affected with phenylketonuria (PKU), an autosomal recessive disorder resulting from the inability to break down the amino acid phenylalanine. What is the allele frequency of the disease allele in the United States, and what percentage of Americans would be predicted to be heterozygous carriers for the disease allele?

*Given that $q^2 = 1/10,000 = 0.0001$ (1 point). So $q = \text{the square root of } 0.0001 = 0.01$ (1 point)
Since $p+q=1$, $p = 1-0.01 = 0.99$. Percentage of heterozygous carriers given by $2pq = 2*(0.99)(0.01)$ (2 points) = $0.99 * 0.02 = 0.0198 = \sim 2\%$*

Full credit for setting up equations correctly, even if incorrectly did final calculation.

12. (2 points) The rs6983267 SNP upstream of the *MYC* gene associated with cancer that was discussed in lecture has a worldwide allele frequency of 50% G and 50%T. If MCB104 has 200 students in it, how many students would be predicted to be homozygous for the cancer risk allele? (assume the MCB104 population is similar to the worldwide population to answer)

*$p=0.5$, $q=0.5$. q^2 would be $0.5*0.5=0.25$ or 25% (1 point). 25% of 200 = 50 students homozygous for risk allele (1 point).*

13. (6 points) In humans, color blindness is an X-linked recessive trait. In one population, 10% of men in a population are affected.

A. (2 points) What is the frequency of the color-blinded allele in this population (assume random mating)?

Men have one X chromosome, so if 10% of men have recessive trait, then $q = 0.1$ (10%).

B. (2 points) What percentage of women in this population are expected to be color-blind?

If $q=0.1$, then $q^2 = 0.1 \times 0.1 = 0.01$, or 1 percent

C (2 points) What percentage of women in this population are expected to be heterozygous carriers of the color-blind allele?

If $q=0.1$ and $p+q=1$ then $p=0.9$. Percentage of heterozygotes given by $2pq$, or $2(0.9)*(0.1) = 0.18$ or 18 percent.*

Name _____ KEY _____

S.I.D. _____

14. (4 points) In what linear order would these five sequences assemble into a contig?

1. GCCGAAAGCGCCGCTGTGGCTGAGA
2. ACTGGGTTGTTACATTCAAAAAAACT
3. TGGCTGAGAGCGAAGTTTCAGAGACTCTTA
4. GGAAGGCTACTGGCCGAAAG
5. GAGACTCTTATTTAAACTGGGTTG

All sequences are on the same strand. Answer just with the linear order of sequence reads from 5' to 3'. Use at least 8 base pairs minimal overlap for this question.

4-1-3-5-2 (partial credit if part of sequence is correct, 0 pts for 1 bp overlaps)

15. (4 points) QTL mapping muscle mass phenotypes in pigs identified the genomic region including the *Insulin Growth Factor 2* gene. Further fine mapping studies suggested the phenotype maps to a non-coding region nearby the *IGF2* gene. This pig sequence is highly conserved in the mouse genome. Before testing for enhancer activity of this non-coding sequence you want to know in what cells in developing mouse embryos the *IGF2 mRNA* is produced, and also in what cells the *IGF2 protein* is present. Describe two experiments, one to detect *IGF2 mRNA*, and another to detect *IGF2 protein*.

To detect IGF2 mRNA, you could perform in situ hybridization with a labeled antisense IGF2 riboprobe (or cDNA probe). To detect IGF2 protein, you could perform immunolocalization experiments, using an antibody that binds to the mouse IGF2 protein.

For full credit had to describe the experiment mentioning riboprobe and antibody – half credit if only said in situ hybridization and immunolocalization (or immunohistochemistry or use and antibody specific to IGF2 protein).

16. (4 points) The *even-skipped* gene is expressed in seven stripes in early *Drosophila* embryos. In no more than three sentences, how would you test non-coding regions of DNA around the *even-skipped* gene for enhancer activity? Your answer should include three different regions of DNA.

You could make reporter gene constructs where non-coding regions of DNA were ligated upstream of a minimal promoter and a reporter gene (e.g. lacZ or GFP), then inject these into fly embryos to assay reporter gene expression. 1 point each for non-coding DNA, promoter, and reporter gene, and 1 point for describing putting construct into fly embryos to look for reporter gene expression.

Name _____ KEY _____

S.I.D. _____

17. (9 points) Like in humans, in one species of lemurs blue eye color has evolved. The blue-eyed black lemur (*Eulemur macaco flavifrons*) is closely related to the orange-eyed black lemur (*Eulemur macaco macaco*).

You sequence the *OCA2* enhancer containing the blue-eyed associated SNP in humans in both orange and blue eyed lemurs to ask whether this same SNP is mutated in blue-eyed lemurs. You see from the sequence that it is not, both blue and orange eyed lemurs have the same nucleotide at this SNP (which is the same nucleotide as in all non-blue eyed humans).

A. (3 points) What can you conclude about whether lemurs are blue-eyed due to changes in *OCA2*?

Only that the same SNP is not causative in both humans and lemurs (1 point). Based upon this alone, you can not conclude anything more about whether OCA2 regulates blue eye color in lemurs. Only 2 points if answered nothing, full credit requires mentioning that can only conclude the SNP causative in humans is not causative in lemurs.

In the wild, these two lemur subspecies occasionally interbreed. You collect DNA and phenotype eye color (which is either blue or orange) from a very large two-generation lemur pedigree that is segregating the blue-eyed phenotype. You want to ask whether, as in humans, the blue-eyed phenotype is due to the *OCA2* genomic region.

B. (3 points) If this blue-eyed trait was not due to *OCA2* (or any gene linked to *OCA2*), what would you expect to observe between genotypes of molecular markers ~1 centiMorgan from lemur *OCA2* and the phenotype?

Genotype and phenotype would not be associated (i.e. the phenotypes would not correlate with genotype). Full credit if explained in some way that genotype and phenotype would not be associated/correlated/etc. or that the phenotypic distributions would not be different by genotypic class.

C. (3 points) If this blue-eyed trait was due to *OCA2* (or any gene tightly linked to *OCA2*), what would you expect to observe between genotypes of molecular markers ~1 centiMorgan from lemur *OCA2* and the phenotype?

Genotype and phenotype would be associated (i.e. the phenotypes would correlate with genotype). Full credit if explained in some way that genotype and phenotype would be associated/correlated/etc. or that the phenotypic distributions would be different by genotypic class.

18. (2 points) The mutation causing sickle cell anemia is a SNP in the hemoglobin gene. If you amplified wild-type and mutant sickle cell anemia alleles by PCR, and size-separated the resulting fragments, would you be able to distinguish the two alleles? Answer yes or no, and explain your answer in one sentence.

No. SNPs change a single nucleotide to a different nucleotide, so do not result in a change in DNA sequence length, so the PCR products would be identical in size.

Name _____ KEY _____

S.I.D. _____

19. (4 points) Shown below is a matrix of human SNP genotypes. Each row is a haploid human chromosome, and each column is a specific SNP position, aligned in physical order.

A	G	G	C	T	C	G	A	G	T	A	C
C	A	T	A	T	C	G	A	G	T	A	C
C	A	T	A	A	A	T	C	C	G	T	T
A	G	G	C	A	A	T	C	C	G	T	T
C	A	T	A	A	A	T	C	C	G	T	T
C	A	T	A	T	C	G	A	G	T	A	C
C	A	T	A	T	C	G	A	G	T	A	C
A	G	G	C	T	C	G	A	G	T	A	C
A	G	G	C	A	A	T	C	C	G	T	T
C	A	T	A	A	A	T	C	C	G	T	T

Do the observed genotypes suggest a recombination hotspot? Answer yes or no, and explain your answer in one sentence. If you answered yes, label the recombination hotspot with a line in the above matrix.

Yes, linkage disequilibrium is observed between the four SNP columns on the left and within the eight SNP columns on the right, but linkage equilibrium is observed between SNP columns 4 and 5. Therefore a recombination hotspot exists between SNPs 4 and 5. One point for vertical line between columns 4 and 5, three points for correctly explaining patterns of LD.

20. (4 points) In a forward genetic screen, how would you know whether you had saturated the genome? Answer in no more than one sentence.

If you no longer found new loci (genes), but just more alleles of the same loci (genes), then you would know the genome was saturated.

21. (4 points) The sequencing of tumor genomes is often done to look for chromosomal rearrangements in somatic tumor cells. What are two ways that paired end reads from a genomic library can reveal a translocation? (Hints: the first way uses only single sequencing reads, and the second way uses information from both sequencing reads from a clone. Both ways involve aligning the tumor genome sequence to the normal reference human genome sequence.)

The first way is if a single paired end read spans a translocation breakpoint (one end of sequence matches one chromosome and then suddenly switches to matching another chromosome) (2 points). The second way is if paired end reads from the same clone aligned to two different chromosomes in the reference assembly (2 points).

22. (8 points) Below is a matrix of results from complementation tests between fruitflies with different mutations that affect eye color:

Mutation	white	garnet	ruby	vermillion	cherry	coral	buff	carnation
white	-	+	+	+	-	-	-	+
garnet		-	+	+	+	+	+	+
ruby			-	+	+	+	+	+
vermillion				-	+	+	+	+
cherry					-	-	-	+
coral						-	-	+
buff							-	+
carnation								-

A. (3 points) You intercross a fly that is heterozygous for *ruby* with a fly that is heterozygous for *vermillion*. What would be the phenotypes of the offspring, and in what expected ratios?

100% wild type, as these mutations complement (i.e. are in different genes)

B. (3 points) You intercross a fly that is heterozygous for *cherry* with a fly that is heterozygous for *buff*. In what proportion of offspring would you expect to see any eye color phenotype?

25 percent since these mutations fail to complement

C. (2 points) You isolate a new allele that causes white eyes, but unlike all the other alleles, the new allele is dominant not recessive. You set up complementation tests with flies heterozygous for this dominant allele and flies heterozygous for each of these eight recessive alleles. What would be the phenotypic ratios in the offspring of these crosses?

50% have white eyes, 50% normal in all eight crosses (Dominant allele will cause white eye phenotype with only one copy).

23. (6 points) You are studying a rare but simple autosomal dominant human disease.

A. (2 points) How would you know whether two tightly linked (each ~1 centiMorgan away) molecular markers flanked this disease locus? (i.e. vs. both markers being 5' or 3' to the disease causing mutation)?

The recombinants would be different individuals if the two markers flanked the locus, but the same individuals if markers did not flank. Only underlined first part required for 2 points.

B. (4 points) You eventually find a SNP that tracks perfectly with the disease – in zero of 100 affected patients is recombination observed between disease and this SNP. Is this SNP the causative mutation? Answer yes, no, or maybe and explain your answer in no more than two sentences.

Maybe (1 point). The SNP may just be in strong linkage disequilibrium with the disease causing mutation (or on the same haplotype) (3 points).

Name _____ KEY _____

S.I.D. _____

24. (13 points) In a recent GWAS for lung cancer, the SNP rs8034191 was genotyped in cases and controls. The results are shown in the following table:

genotype	Cancer	Controls
TT	10	5
TC	200	100
CC	900	950

A. (4 points) For both cancer and controls, calculate the total number of T and C alleles, showing your work:

Allele	Cancer	Controls
T	$(10 \times 2) + 200 = 220$	$(5 \times 2) + 100 = 110$
C	$(900 \times 2) + 200 = 2000$	$(950 \times 2) + 100 = 2000$

B. (6 points) What is the allelic odds ratio (for T against C) for cancer risk? Show all of your work, including all of the steps of your calculation.

$$\text{odds of getting cancer if a patient has the T allele} = 220/110 = 2$$

$$\text{odds of getting cancer if a patient has the C allele} = 2000/2000 = 1$$

$$\text{odds ratio (for T against C) for cancer risk} = 2/1 = 2$$

(2 pts for calculating each odds, 2 points for correctly setting up odds ratio – must show all work for full credit)

C. (3 points) In a GWAS, if you genotyped 500,000 SNPs, and found one SNP with a p-value of association to disease equal to 0.04, would you conclude the association between genotype and phenotype was significant? Answer yes or no, and explain your answer in no more than two sentences.

No (1 point). Since 500,000 SNPs are being tested, the Bonferroni-corrected (or multiple hypothesis corrected) p-value of significance should be $0.05/500,000$, or much smaller than 0.04.