

Name _____ KEY _____

S.I.D. _____

MCB 104 MIDTERM #2 October 23, 2013

*****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 are 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 for full credit. If question says answer in a maximum number of words or sentences, going over this could result in less than full credit.

(Do not write below this line)

PAGE 2	____/15 total
PAGE 3	____/11 total
PAGE 4	____/12 total
PAGE 5	____/12 total
PAGE 6	____/12 total
PAGE 7	____/12 total
PAGE 8	____/11 total
PAGE 9	____/15 total

Total Score _____/100

Name _____ KEY _____

S.I.D. _____

1. (1 point) What does QTL stand for? quantitative trait locus (or loci)

quantitative trait locus (or loci)

2. (3 points) In humans, are regions of linkage disequilibrium larger or smaller in a two generation pedigree compared to the global population? Explain your answer in one sentence.

Larger (1 point). Two generations of recombination does not break up linkage disequilibrium much in a two generation pedigree (1 point), while thousands of generations of historical recombination have reduced linkage disequilibrium in the global population (1 point).

3. (4 points) What are contigs, and how can different contigs be connected together in a genome assembly? Answer in no more than two sentences.

(2 points for accurate contig definition, 1 point each for correctly described paired end reads and scaffolds)

A contig is sequence assembled from contiguous or overlapping DNA sequence (2 points), and different contigs can be connected together into scaffolds using paired-end reads (1 point) where one paired end read of a clone aligns to one contig and the other paired end read of the same clone aligns to another contig (1 point).

4. (3 points) The *CCR5delta32* allele contains a 32 base pair mutation in the human *CCR5* gene, which encodes the receptor for the HIV particle. Having two copies of this allele confers strong resistance to HIV infection. Why is this allele easier to molecularly genotype than a SNP? Answer in no more than two sentences.

Because the allele contains a small deletion, diploid genomic DNA can be molecularly genotyped by simply amplifying by PCR across this sequence, then running the PCR products on a gel.

5. (4 points) How would you determine whether any genes were differentially spliced in different cell types? Answer in no more than two sentences.

Harvest mRNA from different cell types, reverse transcribe mRNA into cDNA, make libraries of cDNA clones, sequence ESTs, then compare the pattern of exons included in each gene across cell types.

Name _____ KEY _____

S.I.D. _____

6. (7 points) Age-related Macular Degeneration (AMD) was found in a GWAS study to be associated with a variant of the Complement Factor H (CFH) gene that has an amino acid change. This allele (called Y402H because the wild-type Y amino acid at position 402 is mutated to H) is extremely common, with the disease risk H allele having a ~20% allele frequency in humans worldwide.

A. (5 points) Assuming Hardy-Weinberg equilibrium, calculate the probabilities of all three genotypes (YY, HH, and HY) in humans.

(must have shown work for full credit, 1 point each for p and q, 1 point for each diploid genotype)

(let p = Y and q = H)

*q = 0.2, so $q^2 = HH = 0.2 * 0.2 = 0.04$; $HY = 2pq$, $H = 1 - 0.2 = 0.8$ so $HY = 2(0.8 * 0.2) = 2(0.16) = 0.32$ and $p = H$ so $p^2 = H * H = 0.8 * 0.8 = 0.64$*

HH = 0.04, HY = 0.32, YY = 0.64

B. (2 points) Two copies of the H allele confer about 7 times the risk to blindness (compared to the risk of YY homozygotes). If MCB104 has 200 students in it, and the allele frequencies in that population are the same as the total world population, how many students would be predicted to be homozygous for this variant?

*$q^2 = 0.2 * 0.2 = 0.04 * 200 = 8$ students*

7. (4 points) In what linear order would these five sequences assemble into a contig?

1. CCGACTCCAGCCTCCACTGCCTCGAGCCCCC
2. GCTCTCCAGTTCTCCGACTCCAG
3. CAGAGAGCACCAGCTCCACAAG
4. CTCCACAAGGGACCTGCTCTCC
5. GCCCCCTGTACGAAGTGGACTCTC

All sequences are on the same strand. Answer just with the linear order of sequence reads from 5' to 3'. Use at least 7 base pairs minimal overlap for this question.

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

Name _____ KEY _____

S.I.D. _____

8. (9 points) The gene *Sonic Hedgehog* (*SHH*) is required for many developmental processes in vertebrates, including development of the limbs. A limb specific enhancer of the *SHH* gene maps 1 million base pairs away from the *SHH* gene, and mutations in this enhancer cause polydactyly (extra fingers) in human patients. This enhancer is required for *SHH* limb expression and limb development, as mice homozygous for deletions of this enhancer have extremely reduced limbs. These mice live much longer than mice homozygous for loss-of-function mutations in the exons, which die as early embryos due to defects in multiple other tissues (e.g. the brain and spinal cord) that express *SHH*.

A. (2 points) How would you ask which cells in a mouse embryo are transcribing the *SHH* gene? Answer in three words.

in situ hybridization

B. (4 points) Given the information in the paragraph above, what would you expect the result in part A above would be in wild-type mice, and in mice homozygous for the limb enhancer deletion? Your answer should mention three different tissues and be no more than two sentences.

In wild-type mice, SHH expression would likely be detected in cells in or near the limbs, brains, and spinal cord. Since the limb enhancer is required for SHH expression, in mice homozygous for the limb enhancer deletion, SHH transcripts would be detected in or near the brain and spinal cord, but not limbs.

C. (3 points) How would you test the *SHH* limb enhancer from a patient with a mutation in their limb enhancer and a limb phenotype for enhancer function in a mouse embryo? Answer in two sentences.

Make reporter gene constructs with patients' mutated limb enhancers ligated to a minimal promoter and a reporter gene. These reporter gene constructs could then be injected into one cell mouse embryos and transgenic animals screened for reporter gene expression in the limbs.

9. (3 points) You pay to have your own genome genotyped for 300,000 tagSNPs, and are sent a long list of disease risks based on a subset of your specific SNP genotypes and many published human GWAS studies. Are any of these SNPs causative for diseases you might have or get? Answer yes, no, maybe, and explain why in no more than two sentences.

Maybe (1 pt). The SNPs are simply tagging blocks of genetic variation, so could just be linked to the causative mutation (1 pt). However, the SNP could itself be causative (1 pt).

Name _____ KEY _____

S.I.D. _____

10. (6 points) Kernel color in wheat is a simple quantitative trait, controlled by two unlinked loci, each with two alleles that make purple pigment. Plants homozygous for purple alleles at both loci (AABB) have purple kernels. Plants homozygous for white alleles at both loci (aabb) have white kernels. The total number of A and B alleles determines kernel color in a simple additive way: one allele of either A or B makes kernels light red, two alleles of either A or B makes kernels red, and three alleles of A or B makes kernels dark red. You cross AABB purple plants to aabb white plants and get all red AaBb F1 hybrids. You then backcross these AaBb F1 hybrids to AABB purple plants.

What are the expected ratios of F2 genotypes and phenotypes, using the Aa Bb convention above?

(3 points for correct genotype ratios, 3 points for correct phenotype ratios)

AaBb x AABB

F2 phenotype and genotype ratios:

1/4 purple, (1/4 AABB)

1/2 dark red, (1/4 AaBB + 1/4 AABb)

1/4 red, (1/4 AaBb)

(0 light red Aabb or aaBb)

(0 white, aabb)

11. a. (3 points) Two SNPs are in absolute linkage disequilibrium. SNP1 has allele frequencies of 70% T and 30% C, and SNP2 has allele frequencies of 30% A and 70% T. Calculate the observed frequencies of haplotypes in this population.

1.5 pts/each haplotype frequency

30% CA

70% TT, SNPS are in complete linkage disequilibrium, so CT and TA are never observed.

b. (3 points) If SNP1 has allele frequencies of 70% T and 30% C and SNP2 has allele frequencies of 30% A and 70% T, calculate the observed frequencies of haplotypes in the absence of linkage disequilibrium.

0.5 pts/each haplotype frequency, 1 pt. for showing work

If allele frequencies are 0.7 T and 0.3 C at SNP1 and 0.3 A and 0.7 T at SNP2, haplotype frequencies in the absence of linkage disequilibrium would be:

*0.7*0.3 = 0.21 TA*

*0.7*0.7 = 0.49 TT*

*0.3*0.3 = 0.09 CA*

*0.3*0.7 = 0.21 CT*

Name _____ KEY _____

S.I.D. _____

12. (5 points) Shown below are 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	A	G	C	C	T	A	T
C	G	T	C	T	C	G	A	G	T	A	C
A	C	G	T	A	A	T	C	C	G	G	T
A	G	G	C	T	A	G	C	C	T	A	T
A	C	G	T	A	A	T	C	C	G	G	T
A	C	G	T	A	A	T	C	C	G	G	T
C	G	T	C	T	C	G	A	G	T	A	C
A	G	G	C	T	A	G	C	C	T	A	T
C	G	T	C	T	C	G	A	G	T	A	C
C	G	T	C	T	C	G	A	G	T	A	C

A. (2 points) Do the observed genotypes suggest a recombination hotspot? Answer yes or no, and explain your answer in one sentence.

No. All SNPs are in linkage disequilibrium (LD), at no position are all possible pairs of dinucleotide haplotypes observed.

B. (4 points) How many distinct alleles are shown here? List alleles by their first two SNPs, and calculate the allele frequencies of each allele in this population.

Three alleles: AG is at 0.3 frequency, CG 0.4, and AC 0.3.

13. (4 points) Define what GWAS stands for and list three key steps explained in lecture on how one is carried out: (1 pt for correct definition, 1 pt. for each step)

Genome wide association study

1. collect DNA from large number of cases and controls (or phenotyped or diseased/not diseased) individuals
2. genotype genome-wide SNPs
3. test for associations between genotype at each SNP and disease/trait/phenotype

14. (2 points) If you used Sanger sequencing to sequence with two vector primers the paired end reads of one 4 kb clone from a genomic library, would you get the entire 4 kb of sequence or would you have a gap in the sequence of the clone? Answer this question and explain your answer in no more than one sentence.

You would have a gap in the sequence (1 pt), because you only get about 1 kb of sequence with Sanger sequencing (0.5 pt), so by sequencing 1 kb from either end, there'd be a gap of about 2kb between these sequences (0.5 pt).

Name _____ KEY _____

S.I.D. _____

15. (5 points) In cats, the extent of linkage disequilibrium varies over 20-fold across the genomes of different breeds, with Burmese cats having over 20 times greater linkage disequilibrium than Siberian cats.

A. (3 points) Considering only the relative size of linkage disequilibrium in Burmese and Siberian cats, which breed would give better genomic resolution from GWAS studies and why? Explain in one or two sentences.

Given the larger blocks of LD in Burmese cats than Siberian cats, an associated haplotype would be smaller in Siberian cats, thus genetic resolution would be better in Siberian cats, because more generations of historical recombination have occurred in this breed.

B. (2 points) The Burmese cat breed is known to have a common recessive mutation called “Burmese Head Defect” that when homozygous causes a severe and lethal craniofacial birth defect. You cross a known heterozygous carrier cat to a cat of unknown genotype and observe a litter of two normal kittens. Is the cat of unknown genotype heterozygous for this mutation? Answer yes, no, or maybe and explain in one or two sentences.

Maybe. Since mutation is recessive, you'd expect to see 25% with phenotype if cross fails to complement. However, with just two kittens your sample size is not conclusive.

16. (3 points) Using only paired end read data from genomic DNA shotgun sequencing, how could chromosomal translocations be detected? Answer in one or two sentences.

If paired end reads from the same clone aligned to two different chromosomes in the reference assembly, a translocation could be detected. Also full credit if answer is a single paired end read spans a translocation breakpoint (start of sequence from one paired end read matches one chromosome and then suddenly switches to matching another chromosome).

17. A. (4 points) In dogs, achondroplasia (short limb bones) results from a haplotype containing a retrotransposed copy of the *Fibroblast Growth Factor 4 (FGF4)* gene. In the genome of dog breeds with achondroplasia, a second copy of the *FGF4* gene is found, and appears to have resulted from retrotransposition (meaning mRNA was made into cDNA, and this cDNA inserted randomly somewhere into the dog genome). What genomic sequence is present in the normal, endogenous *FGF4* gene but missing in the retrotransposed copy? What additional sequence is present in the retrotransposed *FGF4* gene that is not present in the endogenous gene? Answer in one or two sentences.

Present in the genome of the endogenous gene but missing from the retrotransposed gene are the introns. Present in the retrotransposed gene but not present in the endogenous gene is the polyA tail.

18. (7 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								-

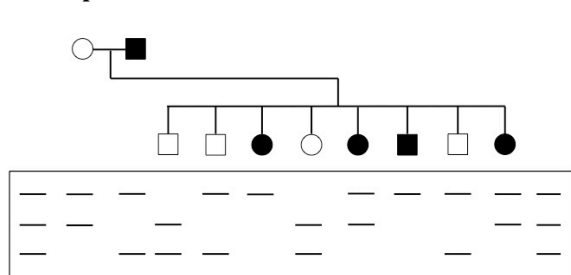
A. (4 points) Based upon these data, how many genes are represented by these mutations? Answer with a number and explain in one sentence how you came up with that number.

Five genes – white, cherry, coral, and buff are all allelic so all represent one gene, but garnet, ruby, vermillion, and carnation complement all other mutations, so are each a separate gene.

B. (3 points) Based upon these data, is the genome saturated for mutations affecting eye color? Answer yes or no, and explain your answer in one sentence.

No. Four of the five genes are represented by only one mutation, so the genome is not saturated, and only would be so if multiple alleles of all genes had been found.

19. (4 points) In trying to determine the genetic basis of a human disease, you genotype a human pedigree shown below, where an autosomal dominant phenotype present in one parent is transmitted to four of eight offspring. Your molecular genotyping assay is a microsatellite known to be tightly linked to the disease locus. You amplify the microsatellite with PCR and size-separate by electrophoresis. Molecular genotypes are shown beneath the pedigree. In the following questions, the lanes on the far left and far right of the pedigrees are DNA size fragments of (from top to bottom): 200 bp, 150 bp, and 100 bp.



A. (2 points) In this pedigree, what are the odds that children of the affected male child (black rectangle) would have the disease if he had children with an unaffected spouse?

50% (since he's heterozygous for the disease allele)

B. (2 points) What are the odds that a child of the last unaffected male (last white square to the right) and a homozygous wild-type mother will be affected? Zero. *This male has mom's 200 bp allele that is not linked to disease allele (since he got 100 bp paternal allele, you know 200 bp allele in this male must be from mom).*

20. (15 points) In a recent GWAS for colon cancer, the SNP rs16969681 was genotyped in cases and controls. The results are shown in the following table:

genotype	Cancer	Controls
TT	50	40
TC	500	320
CC	500	840

A. (4 points) For both cancer and controls, calculate the total number of T and C alleles, showing your work: (you had to show your work for full credit)

Allele	Cancer	Controls
T	$(50 \times 2) + 500 = 600$	$(40 \times 2) + 320 = 400$
C	$(500 \times 2) + 500 = 1500$	$(840 \times 2) + 320 = 2000$

B. (6 points) Set up the equation to calculate the allelic odds ratio (for T against C) for cancer risk? Show all of your work in setting up all equations for full credit.

odds of getting disease with T = $600/400 = 1.5$ (2 pts)

odds of getting disease with C = $1500/2000 = 0.75$ (2 pts)

odds ratio for T against C = $(600/400)/(1500/2000) = 1.5/0.75 = 2$ (full credit if all equations set up correctly) (2 pts)

C. (2 points) If the p-value for a chi-squared test for the above data was 0.007, what would you conclude about the significance of this association *if this was the only SNP you genotyped*?

It would be significant, because a single hypothesis is being tested, and this p-value is smaller than 0.05.

D. (3 points) If the p-value for a chi-squared test for the above data was 0.007, what would you conclude about the significance of this association if this was *one of 100,000 SNPs* tested? Explain your answer in no more than two sentences, and include a mathematical calculation.

(1 point for right answer, 1 point for correct calculation, 1 point for explaining multiple hypotheses or Bonferroni correction needed)

Since 100,000 hypotheses are being tested, the p-value has to be Bonferroni adjusted, correcting for multiple hypothesis testing. This adjusted p-value would be $0.05/100,000 = 5 \times 10^{-7}$. Since 0.007 does not pass this threshold, this association is not significant.