Name _____

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 _____

S.I.D. _____

1. (1 point) What does QTL stand for?

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.

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.

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.

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

Name _____

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.

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?

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

1. CCGACTCCAGCCTCCACTGCCTCGAGCCCCCC
2. GCTCTCCCAGTTCTCCGACTCCAG
3. CAGAGAGCACCAGCTCCACAAG
4. CTCCACAAGGGACCTGCTCTCCC
5. GCCCCCCTGTACGAAGTGGACACTCTC

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.

Name _____

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.

B. (4 points) Given the information in the paragraph above, what would you expect the result in part A above to 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.

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.

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.

Name _____

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 AaBb convention above?

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.

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.

Name _____

S.I.D. _____

12. (6 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.

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.

13. (4 points) Define what GWAS stands for and list three key steps explained in lecture on how one is carried out:

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.

Name _____

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.

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.

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.

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.

Name _____

S.I.D. _____

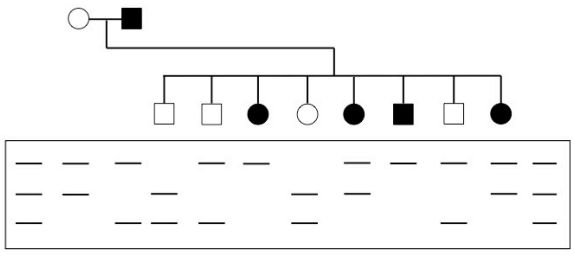
18. (7 points) Below is a matrix of results from complementation tests between fruitflies with different mutations that affect eye color:

| 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.

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.

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 woman?

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?

Name _____

S.I.D. _____

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:

| Allele | Cancer | Controls |
|---|---|---|
| T | | |
| C | | |

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.

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?* Answer in one or two sentences.

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.