

**FINAL PROJECT/EXAM**  
**STAT 550 / BIOS T 550: Statistical Genetics I: Spring Quarter 2021**  
**Due by 12:00 p.m., Friday, June 11, 2021**

Name: \_\_\_\_\_

- 
1. Students are not allowed to discuss, collaborate, or work together on this examination.
  2. The exam is due Friday, June 11, 2021 by noon (12:00 p.m.). Exams should be submitted via the Canvas course website.
- 

Question	Points	Score
Question 1	35	
Question 2	65	

1. [35 points] **Haplotypes and Disease Prevalence**

Consider two linked bi-allelic loci (e.g. SNPs) where Locus 1 has alleles  $A$  and  $a$ , and locus 2 has alleles  $B$  and  $b$ . The haplotype frequencies of the two loci for a population are given below. The population is in HWE.

haplotype	frequency
AB	0.2
Ab	0.01
aB	0.3
ab	0.49

- (a) For the two linked SNPs:
- Obtain the linkage disequilibrium coefficient  $D$
  - Obtain the normalized linkage disequilibrium coefficient  $D'$
  - Obtain the squared correlation coefficient  $r^2$ .
  - Comment on the  $D'$  and  $r^2$  values for the two loci and interpret these values.
- (b) Assume that a recessive disease is caused by locus 1 in the population and the genetic model has penetrance values of:  $f_{AA} = 0.1$ ,  $f_{Aa} = 0.005$ , and  $f_{aa} = 0.001$
- What is the prevalence of the disease in the population? Show all work.
  - What is the frequency of allele  $B$  at locus 2 for the affected individuals in the population? What is the frequency of allele  $b$  at locus 2 for the affected individuals in the population? Show all work.
  - What is the frequency of allele  $B$  at locus 2 for the unaffected individuals in the population? What is the frequency of allele  $b$  at locus 2 for the unaffected individuals in the population? Show all work
  - How do your allele frequencies in part (ii) and part (iii) compare to the allele frequencies of locus 2 in the entire population?

## 2. [65 points] Genome-Wide Association Study (GWAS) of Dichotomous Transferrin Serum

Iron is essential for a number of biochemical functions including oxygen transport and oxidative phosphorylation. Excessive iron can cause various disorders, such as iron-overload-related liver diseases, whereas iron deficiency can lead to anemia. Iron status can be assessed by measuring the levels of serum transferrin in the blood. Transferrin is a beta globulin in blood plasma capable of combining with ferric ions and is essential for transporting iron in the body. A 2009 American Journal of Human Genetics article entitled “Variants in TF and HFE Explain ~40% of Genetic Variation in Serum-Transferrin Levels Genome” by Benyamin et al. performed a GWAS of transferrin, and a link to this paper has been provided for you.

You will conduct a GWAS of transferring for case-control phenotype of Transferrin using the PLINK file provided to you for Lecture 20 and write a report about your analysis and results. The PLINK files “Transferrin.bed”, “Transferrin.fam”, and “Transferrin.bim” contain genome-screen data for a sample of individuals with European ancestry. The phenotype file “TrBinaryPheno.txt” contains a dichotomized transferrin variable where individuals who individuals with “high” transferrin measurements were identified to be cases, with a value of 2 for the dichotomous variable, and the controls are individuals who do not have “high” transferrin measurements, and have a value of 1 for this variable. Individuals with missing transferrin measurements have a value of 0 for the dichotomous transferrin phenotype.

The report is to be between 5 and 10 **double spaced pages**, including any tables and figures. The report should include the following sections and adhere to the guidelines given below for each section:

- (a) **Title:** The title of the report may occupy no more than three lines of type. Each line should contain no more than 54 characters, including spaces. The title should convey the conceptual significance of the report to a broad readership.
- (b) **Introduction:** The Introduction should be succinct, with no subheadings, and should present the background information necessary to provide a biological context for the results.
- (c) **Material and Methods:** The Material and Methods section needs to include sufficient detail so that readers can understand how the experiments were performed, and so that all procedures can be repeated, in conjunction with cited references. This section should also include a description of the data, the hypotheses of interest, and the statistical methods employed in the study to analyze the data.
- (d) **Results:** The Results section should include the results of your analysis and should highlight any significant findings. If there are SNPs that have a significant association with the dichotomous transferrin phenotype, you should also determine if these SNPs are in genes.
- (e) **Discussion:** The Discussion section should explain the significance of the results and place them into a broader context. It should not be redundant with the Results section. Include in the discussion a brief description of the Benyamin et al. (2009) and a comparison of your association results to the transferrin GWAS results from this article.