GENE 210: Personalized Genomics and Medicine
Spring 2013 Final Exam
Due Tuesday, May 28 2013 at 10 am.

Stanford University Honor Code

The Honor Code is the University's statement on academic integrity written by students in 1921. It articulates University expectations of students and faculty in establishing and maintaining the highest standards in academic work:

• The Honor Code is an undertaking of the students, individually and collectively:
  – that they will not give or receive aid in examinations; that they will not give or receive unpermitted aid in class work, in the preparation of reports, or in any other work that is to be used by the instructor as the basis of grading;
  – that they will do their share and take an active part in seeing to it that others as well as themselves uphold the spirit and letter of the Honor Code.

• The faculty on its part manifests its confidence in the honor of its students by refraining from proctoring examinations and from taking unusual and unreasonable precautions to prevent the forms of dishonesty mentioned above. The faculty will also avoid, as far as practicable, academic procedures that create temptations to violate the Honor Code.

• While the faculty alone has the right and obligation to set academic requirements, the students and faculty will work together to establish optimal conditions for honorable academic work.

Signature

I attest that I have not given or received aid in this examination, and that I have done my share and taken an active part in seeing to it that others as well as myself uphold the spirit and letter of the Stanford University Honor Code.

Name: Stefano Rensi SUNet ID: srensi

Signature: Stefano Rensi

Some questions may have multiple reasonable answers: if you are unsure, provide a justification based in genetics and cite your sources (SNPedia is fine, journals are better); as long as the justification is sound, you will receive full credit.

If you are unsure which SNP(s) are associated with a trait, you may consult any reference you like.

A family of 3 (mother/father/daughter) has come to you to find out what they can learn from their genotypes. The parents were both adopted, so they do not know any of their family history. You have sent their DNA to LabCorp, which ran their genotypes on a custom 1M OmniQuad array, and they've returned the results at: http://www.stanford.edu/class/gene210/files/final/final_patients.zip *(X points)*

1. A mislabeling in the lab has caused the samples to be shuffled around and they are simply labeled: 'patient1.txt,' 'patient2.txt,' and 'patient3.txt.' Determine which sample is the mother's, the father's and the daughter's. *(15 points)*

Mother - Patient 1
Father - Patient 3
Daughter - Patient 2

Evidence: Patient 3 has a Y chromosome. For rs5940403 in the X chromosome patient 3 is hemizygous C, Patient 1 is CT, and Patient 2 is CC. Thus, patient 2 must be the daughter, as patient 1 could not have inherited a T allele from patients 2 or 3. Also, on a PCA plot for ancestry, Patient 2 lies between patients 1 and 3.

2. What can you tell about the ancestry of the parents? *(15 points)*

On the PCA plot, both of the parents and the child fall in the European Cluster region, with the father being more European, and the mother being more near eastern. In a more detailed regional PCA plot, the father falls in the northern European/French cluster, while the mother sits at the intersection of the southern and northern European clusters, a region containing individuals of northern Italian, Tuscan and French Ancestry. In the Chromosome painting view, these observations are confirmed with all chromosomes being classified as CEU when compared with the Hapmap 2 genomes, and as TSI when compared with Hapmap 3 genomes. The mother is most likely has Ashkenazi Jewish ancestry as she is a carrier for the BRCA1 Ashkenazi haplotype.

3. The parents are concerned about their daughter's chance for getting breast cancer. You investigate the genomes of the father, mother and the daughter and provide genetic counseling for the family. *(15 points total)*

A. What is the lifetime risk for breast cancer for the overall population of Europeans?

The average lifetime breast cancer risk for European women is about 1 in 8, or 12.5%.

B. Does the genotype of the mother or daughter (at rs77944974) alter their risk of breast cancer? Explain briefly, providing data on the most important risk alleles and their effect on risk for breast cancer.

The mother is a heterozygous carrier (D,I) of the 185delAG BRCA1 mutation. For this particular SNP the risk allele is the deletion (D) allele. According to SNPpedia, the mutation is dominant, with the presence of a single risk allele increasing lifetime breast cancer risk from ~13% to ~80%, and ovarian cancer risk from ~2% to ~50%. According the 23andMe lifetime risk is increased to ~60% for breast cancer, and ~40% for ovarian cancer. The daughter homozygous (I,I) and not a carrier of the of the 185delAG BRCA1 mutation.

C. Briefly outline what advice you would give to the mother about her risk for breast cancer, based on your analysis?

*I would inform he mother the she has a **significantly** elevated risk of developing breast or ovarian cancer. I would counsel her about the risks, costs, and benefits of different possible courses of action, such as increased surveillance or a preventive mastectomy and/or hysterectomy.*

D. Briefly outline what advice you would give to the daughter about her risk for breast cancer, based on your analysis?

I would inform the daughter that she is not carrier of the 185delAG BRCA1 mutation, thus she does not appear to be have a significantly elevated risk of breast cancer. I would remind her that the overwhelming majority of breast cancer cases are not genetically inherited, that breast cancer is still one of the most common cancer types in women, and encourage her to eat well, exercise, and perform regular BSEs as she gets ages to mitigate her risk (as well as promote good overall health).

4. Weeks later, the father (a 42 year old, 185 cm in height, 80 kg in weight, not taking any other medication) is rushed to the hospital with a stroke. What dose of warfarin would be given from a clinic that does not perform genetic testing? What dose of warfarin would be given from a clinic that does perform genetic testing? Explain the genetic basis for modifying the warfarin dose of the father given his genotype. *(5 points)*

39.37 mg/wk Clinical Dose
24.46 mg/wk PGx Dose

5. In her next visit, you observe that the mother has high cholesterol. Would you prescribe simvastatin (Zocor) to the mother? Why or why not? *(5 points)*

No I would not. I would prescribe an alternative. The mother is homozygous CC at rs4149056. This genotype is associated with an elevated risk of simvastatin-related myopathy.

6. You counsel the family about the risk for type 2 diabetes for their daughter. You analyze the daughter's genome on genotation.com. You need to explain the results to the family, and how this influences the daughter's risk for Type 2 diabetes. *(15 points total)*

A. What is the likelihood of type 2 diabetes prior to genetic testing?

23.7%

B. What is the likelihood of type 2 diabetes following analysis of the daughter's genotype using Genotation?

44.2%

C. How many SNPs were used to assess the risk for type 2 diabetes?

15 SNPs

D. How were the SNPs combined to give the overall score? Which SNP had the greatest influence on diabetes risk? Explain briefly.

The combinatorial effect of the SNPs was estimated by taking their product of their likelihood ratios. rs9465871 had the largest individual effect with a LR of 1.5.

E. What advice can you provide to the family to help mitigate the chance of their daughter developing type 2 diabetes?

Maintain healthy diet and exercise habits, eat foods low on the glycemic index, make sure to get regular HBA1c tests when getting checkups at the hospital.

7. The following two SNPs were shown to be associated with risk for type 2 diabetes in two GWAS studies. *(15 points total)*

snp	odds ratio	p-value	cases	controls
rs4402960	1.14	8.9×10^{-16}	14586	17968
rs7754840	1.28	3.5×10^{-7}	1921	1622

A. Which SNP has a larger effect size on risk for type 2 diabetes? Explain your answer.

rs7754840 will have the bigger effect as it has a larger odds ratio

B. Which SNP is most statistically significant for risk for type 2 diabetes; i.e. which SNP is most likely to have a true association?

rs4402960 is the more significant risk as it has a much lower p-value and also a much higher statistical power.

C. Is the SNP with the biggest effect size on risk for type 2 diabetes always going to be the SNP that is most statistically significant? Why or why not?

No. Statistical power is a function of both effect size and sample size. A smaller effect/risk with a very large sample size may be more significant than a large effect with a small sample size. And this isn't even factoring in a prior if we want to go the Bayesian route.

D. rs7754840 is a SNP that lies within the CDKAL1 gene. This SNP was identified because it was contained on the Illumina Chip used for genotyping in the GWAS study. Does this result indicate that rs7754840 is the causal mutation? Does this result indicate that CDKAL1 is involved in type 2 diabetes? Explain why or why not.

No GWAS does not imply causality, only that the SNP is in high LD (in the study popul-1 sibly very close to the causal variant. However, this does imply that CDKAL1 is involved in type 2 diabetes as the vast majority of the SNPs in high LD with the associated variant (the candidates for the causal SNP) are in the CDKAL1 operon.

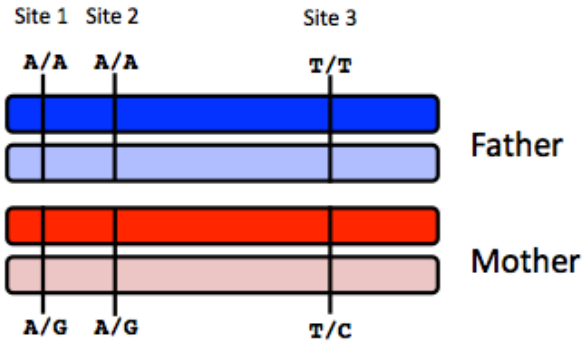
8. The two parents are considering having another child. You analyze their genomes and then counsel them on their chance of having a child with one of the following diseases: hemochromatosis (rs1800562), Alzheimer's disease (specifically, look for APOE4 status), breast cancer (BRCA1 status; rs77944974), cystic fibrosis (rs113993960) and sickle cell anemia (rs334).

For each of these five diseases, what is the chance that the child will have that disease? Briefly explain your answer. *(15 points total)*

> ***Hemochromatosis*** *– There is no risk for the child as the mother is homozygous GG and the disease only effect (AA) homozygotes*
> ***Alzheimer's Disease*** *– The -1 e a significantly increased risk of Alzheimer's. The father is double homozygous APOE4, while the mother is apoE3/E4. There is a 75% chance the child will be apoE4/E4, and a 25% chance the child will be apoE4/E3, both of which present an increased risk of Alzheimer's.*
> ***Breast Cancer*** *– There is a 50% chance that their next child will inherit the BRCA1 mutation from its mother. In the case of a male child, the child will be a carrier, but most likely unaffected. If female, the child will have a significantly increased risk of breast cancer. Overall, there is a 25% chance that the child is both female, and receives the BRCA1 mutation.*
> ***Cystic Fibrosis*** *– The child has an elevated risk because both parents are carriers of the cystic fibrosis gene. The child has a 25% chance of having cystic fibrosis.*
> ***Sickle Cell Anemia*** *– The child has no chance of having sickle cell anemia as neither parent has sickle cell trait.*

9. Prenatal genetic diagnosis (15 points total)

A) A pregnant woman seeks non-invasive prenatal genetic testing and provides a sample of plasma. You isolate the cell-free DNA (cfDNA) from the maternal plasma and determine that 10% of it is derived from the fetus. You perform whole genome sequencing on genomic DNA samples from the mother and father. Next you perform whole genome sequencing on the cfDNA isolated from maternal plasma. For each of the sites below, you obtain 100X coverage (i.e., 100 reads for each site). Fill in the **expected** read counts in the tables below. Use the parental genotypes below and the observed allele counts for the cfDNA sequencing to infer the genotype of the fetus at each of three sites and fill them in the table.



Site 1

	A reads observed	A reads expected
If mother transmits A	59	55
If mother transmits G	59	50

Site 2

	A reads observed	A reads expected
If mother transmits A	52	55
If mother transmits G	52	50

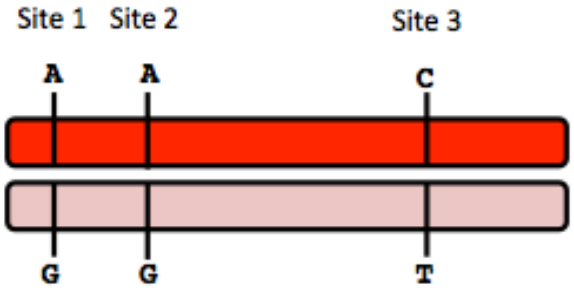
Site 3

	T reads observed	T reads expected
If mother transmits T	49	55
If mother transmits C	49	50

Infer fetal genotype:

Site 1	Site 2	Site 3
AA	GA	CT

B) You worry that your call at site 3 might not be accurate. In order to improve the accuracy of your fetal genotyping, you use parental haplotype blocks. Re-evaluate your fetal genotype inference based on the maternal haplotypes below.



Re-evaluated fetal genotype inference:

Site 1	Site 2	Site 3
AA	AA	CT

10. Neurodegenerative disease genetics (15 points total)

A) Mutations in several genes connected to production of amyloid-beta (Ab) peptides are associated with early onset Alzheimer disease. These include mutations in APP (amyloid-b precursor protein), and presenilin 1 (PSN1) and presenilin 2 (PSN2). APP is the protein from which Ab peptides are derived and PSN1 and PSN2 are components of gamma-secretase, the enzymatic complex that cleaves APP to generate Ab peptides. So far, all Alzheimer disease-linked APP mutations lead to increased production of Ab peptides as does Down Syndrome (trisomy 21), since the *APP* gene is located on chromosome 21. Thus, it appears that increased levels of Ab peptides could lead to disease.

Researchers from the company deCODE Genetics in Iceland analyzed whole-genome sequence data from 1,795 elderly Icelanders and identified a coding mutation (Ala673Thr) in APP that protects against Alzheimer disease and cognitive decline in the elderly without Alzheimer disease. They found that the protective Ala673Thr variant was significantly more common in a group of over-85-year-olds without Alzheimer disease (the incidence was 0.62%) — and even more so in cognitively intact over-85-year-olds (0.79%) — than in patients with Alzheimer's disease (0.13%). Based on what you know about Alzheimer disease genetics:

A) In <u>one or two sentences</u>, propose a mechanism by which this mutation might protect against Alzheimer disease.

The mutation could work by inhibiting conversion of APP to Ab, by making it less susceptible to cleavage by PSN1 and PSN2, alternatively, it might act be making APP more susceptible to degradation by competing proteases, thereby pushing molecular flux into alternate pathways.

B) In <u>one or two sentences</u>, suggest an experiment to test your hypothesis.

The alternate flux hypothesis could be tested by transferring the protective gene into mice, and then knocking out (either genetically or with pharmacological agents) proteases that compete with PSN1 and PSN2 for APP. Increased resistance to proteolytic cleavage by PSN1 and PSN2 can be tested by expressing APP (mutant and wildtype), PSN1, and PSN2 in a eukaryotic expression chassis (such as saccharomyces), purifying the proteins, adsorbing PSN1 and PSN2 onto a carbon electrode with some redox mediator, adding the APP in some buffer solution, and then reading the current generated by the resulting redox reaction (the electrode acts as an electron source/sink for the proteolytic cleavage).

11. Extra credit question available at
http://www.stanford.edu/class/gene210/web/html/extracredit.html (*13 pts*).