GENE 210: Personalized Genomics and Medicine
Spring 2013 Final Exam
Due Thursday, June 13, 2012 at midnight.

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: Gergana Vandova SUNet ID: 05706546

Signature: *Gergana Vandova*

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.

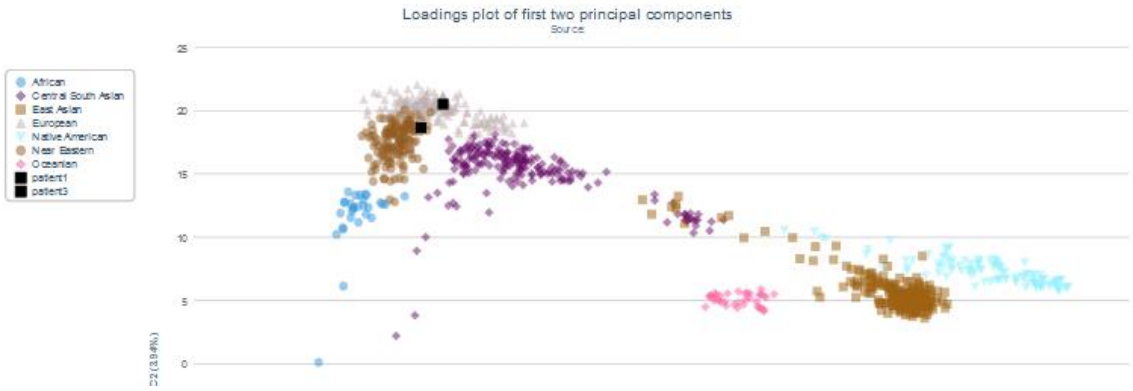
A family of 3 (mother/father/daughter) has come to you to find out what they can learn from their genotypes. The parents were both adopted, so they do not know any of their family history. You have sent their DNA to LabCorp, which ran their genotypes on a custom 1M OmniQuad array, and they've returned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

I looked at several SNPs, for which patients 1 and 3 are homozygous, and patient 2 is heterozygous: (like rs2180311 and rs7520996). Patient 3 has SNPs on the Y chromosome. Thus, Patient 1 is the mother, patient 2 is the daughter, and patient 3 is the father.

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

By mapping the mother and father onto the World reference panel of principal component analysis on Genotation, the mother falls into the boundary between Near Eastern and European part of the scatter plot, and the father – in the European:



By looking at the European reference panel, the mother seems to be of Italian/Portugal ancestry, and the father - French.

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?

13.5 out of 100 Women will develop breast cancer between the age of 29 and 79.

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.

Rs77944974 is a two base deletion mutation, 185delAG, in the BRCA1 gene, which is implicated in breast cancer. Mutations in BRCA1, as well as in BRCA2 gene account for most of cases of inherited breast cancer in women. The risk conferred by these mutations varies, as those in BRCA1 have bigger effect than those in BRCA2. Some of the most common mutations are 185delAG (DD or DI at Rs77944974) in BRCA1, 5382insC in BRCA1 (II or DI at i4000378), and 6174delT in BRCA2 (DD or DI at i4000379). Rs77944974 leads to 12 to 60% increased risk of developing breast cancer, and 2 to 40% increased risk of developing ovarian cancer. By the age of 70, 50-60% of women who have BRCA1 mutation will develop breast cancer and 20-40% will develop ovarian cancer. The mother is heterozygous at that position (genotype DI) and thus have 60% increased risk of developing breast cancer. The daughter is homozygous (genotype II) and thus does not have the BRCA1 mutation.

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

The mother is at higher risk for developing breast cancer, so she should have screening tests more often so that the cancer can be caught early if it develops. Another option is preventive surgery.

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

The daughter does not have the 185delAG mutation. Despite that, regular screening tests are recommended. She should test herself for other breast cancer mutations and estimate her risk.

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)*

If the clinic doesn't perform genetic testing, he would be given a very low dose (~4mg/ml) and then he would have to come in once a week to be checked how he responds to the drug dosage. Then, that dose would be gradually increased.

Alternative to that would be to take into account his age, weight, and height and calculate what dose would be appropriate for him: 39mg per week.

If the clinic performs genetic testing, they can look for mutations in the VKORC and the CYP2C9 and CYP4F2 (CT) genes. Based on his genotype, they would prescribe him 25mg/week.

Gene	*SNP*	*Risk allele*	*Genotype*
VKORC	*rs9923231*	*T*	*TT*
*CYP2C9 *2*	*rs1799853*	*T*	*CT*
*CYP2C9 *3*	*rs1057910*	*C*	*AA*

CYP2C9 metabolizes warfarin, so more risk alleles (C) will cause the drug concentration in the bloodstream to decrease more slowly and the person would need lower dose. The patient has one C, so that would increase his sensitivity to warfarin.

VKORC1 is the inhibited gene upon the action of warfarin. Thus, if the patient has a mutation in this gene, he would already need a lower dose of warfarin. This patient is homozygous for the risk allele, which increases his warfarin sensitivity.

In conclusion, according to his genotype, the patient is very sensitive to warfarin, so he would be prescribed 25mg/week, which is almost twice as low as the clinically predicted dose. I used the Genotation website to calculate the genetic dose of warfarin.

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)*

I would not prescribe her simvastatin, because she is homozygous for rs4149056 (CC), which genotype is associated with higher risk of simvastatin-related myopathy than the genotypes CT or TT.

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?

The likelihood ratio is 0.311 and the probability that she will develop type 2 diabetes is 23.7%, according to Genotation. In 23andme, the average risk for women with European ancestry is 20.7%.

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

The likelihood ratio of type 2 diabetes following analysis of the daughter's genotype is 0.97 and the adjusted probability is 44.2%.

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

15 SNPs we used to assess the risk for type 2 diabetes.

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

Each likelihood ratio (LR) for each SNP is multiplied by a probability ratio based on the background population selected. The current likelihood is converted to a probability. This is done for each SNP in the dataset and with each addition of a SNP the probability is being adjusted to obtain a final probability.

rs9465871 had the greatest influence on the diabetes risk, because the likelihood ratio of type 2 diabetes given the CC genotype was 1.5, which was the highest LR of all the SNPs. Adding this SNP to the risk calculation, resulted in increase of the total probability from 31.2% to 40.4%, which was also the highest increase of the adjusted probability.

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

I would advise them to keep their daughter's weight in check and do diabetes screening more regularly.

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.

As the odds ration reflects on the effect size of the SNP on the risk for a given disease, the rs7754840 SNP has a larger effect.

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 more likely to have a true association with type 2 diabetes, because the p-value is lower than the p-value of rs7754840 . ***Bonferonni correction?***

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?

The SNP with the biggest effect size on risk for type 2 diabetes will not always be the most statistically significant SNP, because the Odds ratio and the p-values reflect on different features of the obtained data, which are independent of each other. The odds ratio describes the strength of the association between the SNP and the disease, whereas the p-value is a measure of how true this association is.

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.

-1.

The fact that a SNP is located in a gene doesn't necessarily mean that this mutation is the causal one and that this gene is implicated in the disease. GWAS give us information only about the association, but not about the causality. Thus, further experiments should be done to demonstrate causality.

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) C (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)*

-3

Hemochromatosis:
Mom is AG, dad is GG. G is the allele causing C282Y mutation in the HFE gene. The child has a 50/50% chance of having AG or GG genotype, and thus 50% chance of developing the disease, as hemochromatosis is inherited recessive disorder.

BRCA1:
Mom has DI and dad has II genotype. As mom is a carrier, there is 50% chance that the child will be carrier of BRCA1 mutation, and if that is the case, there is 20-60% chance that the child will develop breast cancer if it is a female and 6.9% if it is a male.

Alzheimer's disease:
Mom has CC in rs7412, and CT in rs429358

Dad has CC in rs7412, and CC in rs429358
The child will have CC in rs7412 and CT or CC in rs429358 with equal probability. As a result, the child's APOE status will be either ε3/ε4 or ε4/ε4 with 50% chance for each. If the child has ε3/ε4 APOE status, it will have two times increased odds of developing Alzheimer's, and if it has ε4/ε4 status, it will have 11 times increased odds of developing the disease. The average odds of people with European ancestry are 7.1-7.2%.

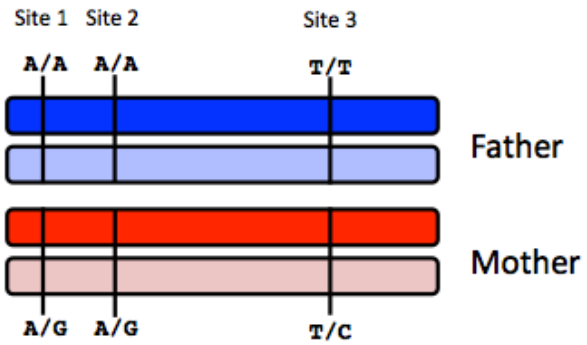
CF:
Mom is DI, Dad is DI – both are carriers of the CF mutation, so there is 25% chance that their child will inherit the two bad copies and if it inherits them, it will develop CF with 100% chance.

Sickle-ce -3
Mom is AA and dad is AA and thus the child will have AA genotype. Thus, the child will develop the disease with 100% certainty.

The above disease risk odds are calculated based only on the SNPs mentioned in the Question. The risk of getting the disease might increase/decrease if other SNPs are considered.

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*

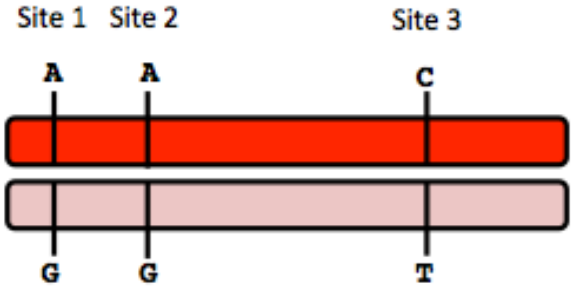
Site 3

	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
A	***G***	***C***

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
A	***A***	***C***

10. Neurodegenerative disease genetics (15 points total)

A) Mutations in several genes connected to production of amyloid-beta (A□) peptides are associated with early onset Alzheimer disease. These include mutations in APP (amyloid□□ precursor protein), and presenilin 1 (PSN1) and

presenilin 2 (PSN2). APP is the protein from which A□ peptides are derived and PSN1 and PSN2 are components of gamma-secretase, the enzymatic complex that cleaves APP to generate A□ peptides. So far, all Alzheimer disease-linked APP mutations lead to increased production of A□□peptides as does Down Syndrome (trisomy 21), since the *APP* gene is located on chromosome 21. Thus, it appears that increased levels of A□ 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 A673T mutation is located at position 2 in the amyloid-β peptide, and thus is very close to the β-cleavage site. An explanation why this mutation is protective against Alzheimer's disease might be because due to this substitution the β-cleavage efficiency is reduced, which would result in the production of less amyloid-β peptides.

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

I would compare the efficiency of β-cleavage in A673T mutants and wild type APP cells by comparing the amount of sAPP β (product of the β-cleavage) and $Aβ_{x–40}$ and $Aβ_{x–42}$ (products of the γ-cleavage, which follows the β-cleavage) by performing immunoprecipitation and western blot analyses on cell supernatants. Decreased concentrations of sAPP β and Aβ fragments and unchanged yields of the sAPPα (product of the α-cleavage) would support my hypothesis.

11. Extra credit question available in the May 23 slot at http://stanford.edu/class/gene210/web/html/schedule.html (*13 pts*).

Person 1	*H*
Person 2	*C*
Person 3	*A*
Person 4	*D*
Person 5	*G*
Person 6	*F*
Person 7	*E*
Person 8	*B*