Zoe N Rogers

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:_____Zoe N Rogers_____ SUNet ID:_znr_____

Signature: _____

Zoe N Rogers

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)*

Patient 3 is the father (the only patient sequenced with a Y-chromosome) Patient 2 is the daughter, if you look at snps related to type 2 diabetes you see that patient 3 is TT at 3020317 and patient 1 is CC and patient 2 is CT, thus patient 2 is the child. Therefore, patient 1 is the mother.

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

When you use the HGDP World or HGDP European setting on the genotation ancestry PCA page, we see that both of them have European ancestry. Specifically, the mother (patient 1) falls within the Northern European (specifically italian) group while the father (patient 3) falls within the Southern European group (specifically 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?
   According to cancer research UK, the life time incidence rates of female European populations are as follows: Western European 89.7 per 100000 people; Northern Europe 85/100000; Southern Europe 68.9/100000; Southern Europe 68.9/100000.

Source: http://www.cancerresearchuk.org/cancer-info/cancerstats/types/breast/incidence/uk-breast-cancer-incidence-statistics#world

According to 23 and Me the average risk for breast cancer in European females is around 13%.
Source: https://www.23andme.com/you/journal/breastcancer/overview/

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 rs77944974 snp is more commonly referred to as the 185delAG BRCA1 mutation. Because this is a frame shift mutation it is very likely functional and is thus a good predictor of breast cancer. The mother has genotype DI at this location and the daughter has II (D=Deletion, I = Insertion). This means that the mother is a carrier of the rs7794497 SNP. This means that the mother's risk for breast cancer is increased from 13 to 81% and her risk for ovarian caner is also increased to 50%. The daughter has the genotype II which means she is not a carrier of this snp and thus does not have a heightened risk for breast cancer (based only on this snp location).
Sources: http://snpedia.com/index.php/I4000377(D;I) ; http://www.ncbi.nlm.nih.gov/pubmed/7550349

C. Briefly outline what advice you would give to the mother about her risk for breast cancer, based on your analysis?
I would advise the mother to take extreme medical precautions. She should increase the amount of mammograms and MRIs that she gets so that if she were to develop lesions they would be discovered early and may be removed without complication. She should at least be educated on how to properly care for herself and her doctors should be informed of this finding so that they can be vigilant in preventing the growth of breast cancer tumors if they were to occur. Her and her doctors can also discuss preventative surgery.

D. Briefly outline what advice you would give to the daughter about her risk for breast cancer, based on your analysis?
Because the daughter is not a carrier for this deletion, there is no heightened risk and she need not change anything about her medical care. She should continue to be routinely screened at well visits though as should most women.

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)*
The clinical dose is 39.37 mg/week, this is what would be given in a clinic with no genetic testing. In a clinic with genetic testing the does is 24.74 mg/week. It has been found that variations in the cytochrome 5450 (CYP2C9) and the vitamin K epoxide reductase complex (VKORC1) genes change the dose requirements for patients. This has been found in numerous studies. These mutations presumably change the way that this person metabolizes this drug. The variation in these genes is then fed through an algorithm to calculate the proper dosage.
Source: Genotation; http://www.nejm.org/doi/full/10.1056/NEJMoa0809329

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! She has the genotype CC at position rs4149056 in the SLCO1B1 gene. This means that she is more likely to have myopathy as a risk of taking simvastatin and related drugs. However, 23 and me says that low doses may be OK for people with this genotype so the mother should discuss her options with her doctor.
Source: Genotation

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 prior to testing is around 23.7%.

B. What is the likelihood of type 2 diabetes following analysis of the daughter's genotype using Genotation?
   The likelihood after testing is around 44.2%.

C. How many SNPs were used to assess the risk for type 2 diabetes?
   15

D. How were the SNPs combined to give the overall score? Which SNP had the greatest influence on diabetes risk? Explain briefly.
   Each snp has a likelihood ratio associated with it, all of these likelihood ratios are multiplied together, starting with the prior likelihood ratio which is based only on the population. The one that contributes most is the one with the highest likelihood ratio, so snp rs9465871 causes the largest movement in likelihood (impact factor is 1.5) .

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

The daughter should probably be screened more frequently for diabetes. She should also be advised by a nutritionist on the best type of diet to help her avoid later onset if she does not already have it. She should also watch her weight as that is a risk factor as well.

7. The following two SNPs were shown to be associated with risk for type 2 diabetes in two GWAS studies. *(15 points total)* 14/15

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 has the larger effect size, (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?
The lower the p-value, the more statistically significant, therefore, r4402960 is more statistically significant.

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. The p-value is highly dependent on the cohort size, so the larger the cohort, often the more significant the study 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
GWAS tell us nothing about functionality, so we absolutely cannot say that this is a causal mutation without many further studies. It also cannot tell us that CDKAL1 is for sure involved in type 2 diabetes, only that it associates. There may, in reality, be no connection. GWAS studies are not super helpful in finding causal genes a lot of the time.

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)* 15/15

Hemochromatosis: The mother is a carrier of the A allele at rs1800562 and the father is not a carrier, therefore, the child has a 50% of being a carrier, but no chance of having both A alleles. Therefore, there is no heightened risk that the child will have hemochromatosis since there is also no chance that they will be a carrier of the rs1799945(G) allele.

Alzheimer's disease: According to SNPedia, "The presence of both rs7412(C;C) and rs429358(C;C) indicates the highest risk ApoE-ε4/ApoE-ε4." Both parents in this case have the genotype CC at rs7412. The father has the genotype CC and the mother has genotype CT at rs429358. This means that there is a 50% that the child will be ApoE-ε4/ApoE-ε4. People who have this genotype are between ten and twenty times more likely to develop Alzheimer's disease. The father himself has this genotype which means that he has a heightened risk as well. Even if the child is only homozygous at the rs7412 position, the child will have a slightly increased risk (about 2 fold increase). Source: Lecture slides and http://snpedia.com/index.php/ApoE-%CE%B54

Breast cancer: The mother's genotype is DI and the father's genotype is II, therefore, the child has a 50% chance that any child will have the DI phenotype and thus have a heightened risk for breast cancer.

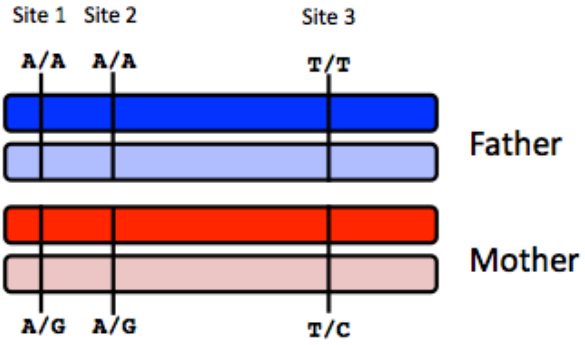
Cystic fibrosis: Both parents have the DI genotype at the rs113993960 locus which means that they are both carriers for the cystic fibrosis snp. This means that there is a 25% chance that any offspring will be homozygous for the deletion at this location and a 50% chance that any offspring will be a carrier for the disease allele like their parents. If the child is homozygous for this allele it is almost certain that they will have cystic fibrosis.

Sickle cell anemia: Both parents are homozygous AA at the rs334 locus so there is a 0% chance that they will have an offspring with sickle cell since this has been shown to be the causal mutation.

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.

Zoe N Rogers



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

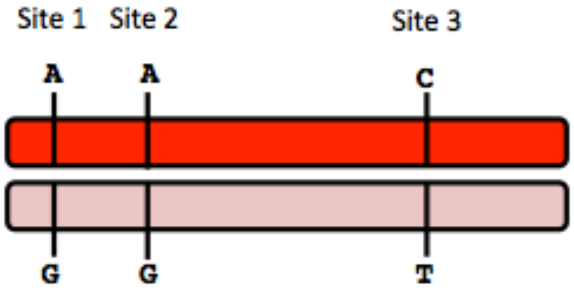
Site 3

	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/A	A/G	T/C

B) You worry that your call at site 2 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	A/A	T/C

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.

This mutation may make some form of APP that blocks the production of beta amyloid. To do this it may be a mutation in a part of APP that makes the

cleavage of this protein less efficient such that Aβ is produced less frequently and therefore decreases the amount of aggregates.

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

If we know the exact mutation, we could clone this mutated gene into an in vitro system, produce mutant protein and see if it is able to be cleaved by beta and gamma secretase enzymes the way that the wild-type protein is. This would tell us whether this was the specific mechanism of this mutant phenotype.

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

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