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:__Kateryna Kozyrytska_____ SUNet ID:__05798757_

Signature: __*Kateryna Kozyrytska*_____

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 – father (Y chromosome)
Patient 2 – daughter (rs3737728 is GG)
Patient 1 – mother (rs3737728 is AG), while father is GG

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

Patient 1 rs4988235 AG and rs182549 CT (heterozygous for lactose intolerance), rs17822931 CC and rs671 GG (unlikely Asian), rs1426654 AA (fair skinned) – perhaps, southern European
Patient 3 rs4988235 AA, rs17822931 CC, rs647 GG, rs1426654 AA – European
(Giardina et al., Curr Genomics 2008 Apr;9(2):110-4
Yoshiura et al, Nat Genet, 2006 Mar;28(3):324-30
SNPedia and class notes)

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?

7.21% (WHO Globocan 2008)

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.

Yes, mother is DI at rs77944974*, that is, one of her chromosomes carries a deletion in BRCA1 gene. Deletion is in exon2 and leads to frameshift in BRCA1 gene known to predispose to breast cancer (exact OR varies by the study, but Struewing published 75% chance in Ashkenazi Jewish population). The mutation, however, occurs and leads to breast cancer in other populations also.*
(Struewing et al, Nat Genet 1995 Oct;11(2):198-200
Abeliovich et al, Am J Hum Genet. 1997 March;60(3):505-514
Bruchim Bar-Sade et al, Hum Mol Genet 1998 7(5):801-805)
Daughter is II at this SNP, so she doesn't have the deletion.

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

She is at high risk of developing early-onset breast cancer and should be inspected regularly. Perhaps, consider double-mastectomy.

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

She does not appear to be at an elevated risk of breast cancer BASED ON THIS SNP ONLY, and further analysis of her genome needs to be done to say conclusively that she is not at an elevated risk. The overall risk is still quite high, so there is no reason she should not be inspected regularly. However, no apparent need for double-mastectomy based on this SNP.

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)*

Clinical: 39.37 mg/week, genetic: 24.74mg/week
*The patient is TT at rs9923231, a polymorphism upstream of VKORC1 coding for an enzyme that activates vitamin K essential for clotting. Those who carry TT at this position have lower mRNA levels of VKORC1. Warfarin inhibits VKORC1, so if a patient has low endogenous levels of VKORC1, lower warfarin dose is required for inhibition. CYP2C9 is a cytochrome P450 responsible for clearance of warfarin. Patient has *1/*2 alleles of CYP2C9, *1 being the "WT" an *2 being a less active variant, so the patient needs a lower dose of warfarin than a *1/*1 (WTWT) would.*
(Rieder et al, N Engl J Med, 2005 Jun 2; 352(22):2285-93
Genotation, SNPedia)

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)*

The mother is CC at rs4363657, which increases her risk of a myopathy ~17 fold higher on simvastatin than if she were TT. So better not choose this statin. (The SEARCH Collaborative Group, N Engl J Med, 2008 359:789-799

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?

8.3% of population (in the US) has diabetes. (American Diabetes Association)

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.

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

Cumulative risk was calculated based on genetic status at the 15 SNPs that have been shown to impact the probability of developing diabetes. Rs9465871 had the biggest impact bringing the cumulative risk up by 9.285%.

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

Healthy lifestyle and eating habits.

7. The following two SNPs were shown to be associated with risk for type 2 diabetes 15/15 ~~s.~~ *(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 has a larger effect. Its odds ratio is 1.28 (larger than 1.14), which means that assuming type 2 diabetes at this SNP is a rare condition, if you have the risk allele at this position, you have a 1.28-fold higher chance of developing diabetes.

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 be true. P-value is an estimate of the chance that the observation is random (that is, the chance that this SNP is in fact not correlated with diabetes.) The significance of the p-value will depend on the number of SNPs used in the study, but these two are likely the same for the two studies.

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 two parameters are only somewhat related. Effect on risk depends mostly on the genetic status and phenotype of the study subjects. P-value depends is a function of both the effect and the number of subjects in the study. The bigger the effect, the fewer subjects you need to have the same level of confidence.

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 and no. Nothing suggests that rs7754840 is the causal mutation. In fact, the SNP may be a silent mutation. It is likely that CDKAL1 is involved in type 2 diabetes. However, it depends on how large the segment is that co-segregates with rs7754840. It may be that the SNP is in linkage disequilibrium with many other SNPs and many other genes. Then, the SNP will only have come up because it is linked to something that plays a role in type 2 diabetes. To account for that, the researchers should check if the SNPs near rs7754840 also show association with diabetes.

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 chan 15/15 have that disease? Briefly explain your answer. *(15 points total)*

Hemochromatosis: rs1800562 AG x GG. A is the risk allele. There is a ½ probability of having a child who is a carrier. Carriers, however, will not be affected since they will be CC at rs1799945. No need to worry until planning grandchildren!

Alzheimer's: rs7412 CC x CC, rs 429358 CT x CC. The mother has two APOE-e4 alleles that are considered high-risk Alzheimer's, while the father has one APOE-e4 and one APOE-e3. The child has a high risk of developing Alzheimer's: 50% probability of being a homozygous APOE-e4 and 50% probability of being heterozygous. While some carriers and even some e4 homozygotes never develop the disease, 1 copy of e4 increases the risk of developing the disease 2-fold, while having 2 copies of e4 increases the risk 20-fold.

Breast cancer: rs77944974 DI x II. The child has 50% chance of inheriting the del185AG (discussed earlier in the final). Given the child inherits the deletion, the chance of developing breast cancer is ~75%. The overall probability, then, is .5.75=.375. It is up to the parents to figure out if that is a high probability.*

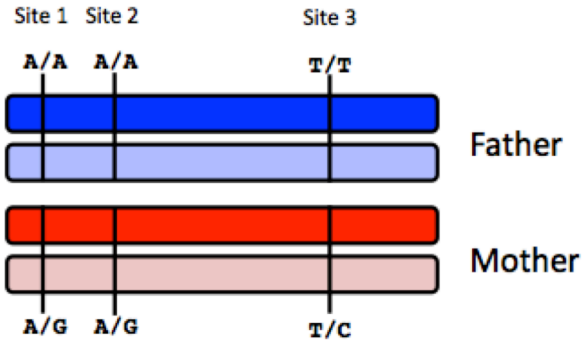
Cystic fibrosis: rs113993960 DI x DI. Deletion here indicates susceptibility. Homozygotes have a very high chance of developing cystic fibrosis. The parents have a 25% chance of having a child who develops cystic fibrosis. 50% chance that the child will be a carrier, which may predispose him or her to the disease (compound heterozygote).

Sickle cell anemia: rs334 AA x AA. There is no chance of developing the disease according to the parents' genetic status.

(SNPedia)

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*

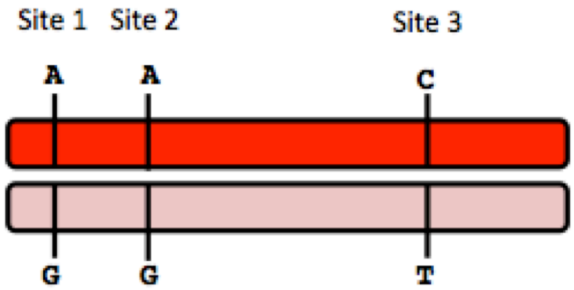
Site 3

	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	***AG***	***TC***

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.

If Ab do, in fact, cause Alzheimer's by aggregation, perhaps A673T is less likely to aggregate. (One can imagine that a hydroxyl at position 673, if exposed, may enhance solubility by hydrogen-bonding with water.)

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

First, we want to confirm that A673T is inside the region that normally aggregates. Next, we can measure (very crudely) the solubility of the two peptides containing either A673 or T673 in water.

Of course, it can be that peptide containing T673 is more soluble than A673 not because the threonine directly interacts with water, but because it interacts with an accessory protein that makes the processed peptide more soluble or through some other mechanism. In that case, it may be more useful to assay propensity of APP A673T to aggregate in a cellular assay with, for example, some of the previously established protocols for aggregation of the native peptide.

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

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