Caleb Chan

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.

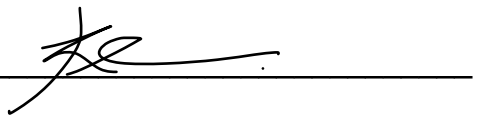
• The faculty on its part manifests its confidence in the honor of its students by refraining from proctoring examinations and from taking unusual and unreasonable precautions to prevent the forms of dishonesty mentioned above. The faculty will also avoid, as far as practicable, academic procedures that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

Signature: 

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)*

By looking at the genotype files:

- Patient 1: XX → mother or daughter
- Patient 2: XX → mother or daughter
- Patient 3: XY → father

Then, using at the GENOtation → Traits → GWAS tool:

SNP	Patient 1	Patient 2	Patient 3
rs7495174	AA	AA	AA
rs17822931	CC	CC	CC
rs4481887	GG	GG	AG
rs713598	GG	CG	CC
rs4988235	AG	AG	AA

From rs713598, patient 2 must be the daughter (with CG) since both parents are homozygous (GG and CC), and the alleles from rs4481887 and rs4988235 support this. Therefore:

- patient 1 = mother
- patient 2 = daughter
- patient 3 = father

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

Using the GENOtation → Ancestry → PCA tool (HGDP: World, resolution = 10000, PC1 & PC2), patient 1 (mother) seems to classify well within the near Eastern group, and patient 3 (father) classifies well within the European group, with patient 2 (daughter) lying midway between the two groups. There, it appears that the mother is of near Eastern descent, while the father is of European descent. Other combinations of principle components seem to support this hypothesis.

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 lifetime risks for breast cancer for the overall population of European women (aged 15 to 79 years) are:

- Central Europe = 6.6%
- Eastern Europe = 5.3%
- Western Europe = 9.7%
- **Overall Europe = 7.2%**

Source:
Forouzanfar MH, Foreman KJ, Delossantos AM, et al. Breast and cervical cancer in 187 countries between 1980 and 2010: a systematic analysis. Lancet. 378(9801):1461-84, 2011.

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 (patient 1) carries a genotype of 'DI', while the daughter (patient 2) carries an 'II', at this SNP. The rs77944974 SNP is located on chromosome 17 at position 41,276,045 (within the BRCA1 gene), and the risk alleles (the 'D's, for deletion) are introduced by the 185delAG mutation, where the bases A and G are deleted from base position 185 of the gene. This causes a frameshift mutation, and introduces a premature stop codon at residue position 39.

The 185delAG mutation is especially common among Ashkenazi Jews, where it is considered a founder mutation and is found in 1% of the population. This corroborates well with the mother's predicted near Eastern ancestry (see Q2), as she is heterozygous for the risk allele. Lifetime risk of breast cancer for women is significantly increased even for carrying a single risk allele, rising from 12% to about 60%. Therefore, the

mother's genotype confers increased risk for breast cancer, while the daughter's genotype does not.

Sources:

- http://snpedia.com/index.php/BRCA1
- http://www.23andme.com/health/BRCA-Cancer/
- Ferla R et al. (2007). "Founder mutations in BRCA1 and BRCA2 genes." Ann. Oncol. 18 Suppl 6: vi93-8.

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

Given the increased risk for breast cancer, I will advice the mother to first perform a screen using MRI or other diagnostics to make sure that breast cancer has not already developed. If the screening results come back as negative, I will suggest frequent self-examination and regular (e.g. yearly) mammograms and clinical breast exams to increase the chance for early detection. Depending on the mother's age and receptiveness I might also bring up the possibility of prophylactic mastectomy as preventive measure. If the screening shows signs of cancer growth, then treatment should commence as soon as possible.

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

Even though the daughter is not a carrier for this particular risk allele, her risk for breast cancer may still be increased due to other mutations/risk alleles related to breast cancer, such as mutations in BRCA2 (e.g. 6174delT), or another mutation on BRCA1 (e.g. 5832insC). I will first make sure from her genotyping result that she does not carry any other known risk alleles for breast cancer, and will then suggest her to perform regular self-examination as well as clinical breast exams whenever her health examinations are due. Given her mother's ancestry and the fact that not all breast cancer risk alleles are known, a heightened awareness on the possibility of breast cancer will hopefully result in early detection, should breast cancer eventually develops.

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)*

According to the GENOtation → Clinical → Warfarin tool, a clinic that does not perform genetic testing would have given the father a dosage of 39.37mg/week,

while a clinic that does perform genetic testing would have prescribed a dosage of 24.74mg/week (or 24.46mg/week with extension).

Warfarin is an anticoagulant frequently used for the prevention of blood clots (e.g. in the treatment for strokes). Several genes, such as VKORC, CYP2C9, and CYP4F2, whose gene products are involved in the metabolism of warfarin, are known to carry alleles that may affect the efficiency of warfarin metabolism. Depending on which part of the warfarin metabolic pathway these gene products are involved in, the alleles associated with these genes may either decrease (requiring a higher dosage) or increase (requiring a lower dosage) warfarin sensitivity. In the father's case, it appears that all of the tested alleles require a decrease in warfarin dosage (increased sensitivity), and therefore the multiplying factors associated with each of these alleles are negative, leading to a lowering of the final suggested dosage (from the clinical dosage).

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)*

According to the GENOtation → Clinical → Pharmacogenomics tool, the mother has risk alleles (genotype = CC) at SNP rs4149056 that is associated with the SLCO1B1 gene. A person with this risk allele may have a higher risk of having simvastatin-related myopathy than a person without the risk allele (CT or TT). Therefore, I will prescribe another drug that will not induce an undesired drug response to treat her high cholesterol.

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 genetic testing is 0.311 (LR), or 23.700% (probability).

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

   The likelihood following the analysis increased to 0.792 (LR), or 44.206% (probability). This gives an increased risk of (0.44206 – 0.237) / 0.237 = 0.87, or 1.87x over the average risk of developing Type 2 diabetes.

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

In the GENOtation analysis, 15 SNPs were used to assess the risk for Type 2 diabetes, with 3 of them (rs2283228, rs9460546, and rs4376068) being imputed from SNPs rs2237892, rs7754840, and rs1470579, respectively.

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

The analysis begins by assigning the patient a prior likelihood ratio/probability of developing Type 2 diabetes based on their selected background population. The analysis then continues by modifying this prior likelihood/probability by using the 15 SNPs in a sequential manner (ordered by descending study size). The likelihood ratios conferred by each SNP is multiplied to the prior probability, and a running likelihood ratio/probability is maintained until the final likelihood ratio/probability is obtained as the overall score.

For the daughter, the risk allele of rs9465871 (CC) gives the largest contribution to her diabetes risk. This risk allele has the highest likelihood ratio of the 15 SNPs (LR = 1.500), and therefore increases the probability by the largest amount (from a prior probability of 32.207% to 40.492%).

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

Even though the daughter is genetically predisposed to developing Type 2 diabetes by having a higher probability (+20%, or an increased risk of 1.87x) than the rest of the population, the genetic component for having Type 2 diabetes is estimated to be only 20 – 26%, with the remaining factors being environmental. Therefore, it would still be advisable to suggest preventive measures that can help mitigate the chance of diabetes development.

Type 2 diabetes onset is believed to be caused by chronically elevated sugar level in the blood, leading to higher sugar level tolerance and the de-sensitization in response to insulin. Therefore, my suggestions will include dietary modification (diet with high dietary fiber, low fat/carbohydrate/sugar), increase in exercise amount, body weight management (obesity is believed to be a major cause), and routine blood glucose level monitoring.

Sources:

- http://www.snpedia.com/index.php/Type_2_diabetes
- https://www.23andme.com/you/journal/type2diabetes/overview/
- Sladek et al. (2007). “A genome-wide association study identifies novel risk loci for type 2 diabetes.” Nature 445(7130): 881-5.

7. The following two SNPs were shown t[illegible] 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 a larger effect size since it has a higher odds ratio of 1.28 (over 1.14 for rs4402960). Odds ratio is a measure of effect size, describing the strength of a risk allele that is associated with the disease.

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 most statistically significant for risk since it has a lower p-value of 8.9×10^{-16}. It is most likely to have a true association because the lower the p-value, the higher the chance that this SNP is associated with the disease.

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 SNP with the biggest effect size is not necessarily always going to be the most statistically significant. Effect size and statistical significance are two separate parameters – effect size means how strong the risk allele is on the disease risk, while statistical significance determines how associated a risk allele is with the disease, and may depend on sample size. It is possible to have a statistically significant risk allele that does not have a very strong effect on the disease risk, or vice versa.

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.

~~The~~ -1. vague. was identified to have a statistically significant association with Type 2 diabetes and that it lies within the CDKAL1 gene does not necessarily mean that it is a causal mutation. This SNP is located within the intronic region of CDKAL1, and the exact function of this gene is not yet known. Therefore, we can only conclude that there is a

correlation between this SNP and Type 2 diabetes, but we cannot say that this is a causal mutation.

Sources:

- http://www.snpedia.com/index.php/Rs7754840
- http://www.ncbi.nlm.nih.gov/gene?cmd=retrieve&list_uids=54901
- Steinthorsdottir, V.; Thorleifsson, G.; Reynisdottir, I.; Benediktsson, R.; Jonsdottir, T.; Walters, G. B.; Styrkarsdottir, U.; Gretarsdottir, S. et al. (2007). "A variant in CDKAL1 influences insulin response and risk of type 2 diabetes". *Nature Genetics* **39** (6): 770–775.

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).

15/15

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

SNP	Father	Mother	Risk Allele
Hemochromatosis (rs1800562)	GG	AG	A
Alzheimer's disease (APOE4 – rs429358/rs7412)	CC/CC	CT/CC	C/C
Breast cancer (BRCA1 – rs77944974)	II	DI	D
Cystic fibrosis (rs113993960)	DI	DI	D
Sickle cell anemia (rs334)	AA	AA	T

- **Hemochromatosis**: the risk allele is A (in the C282Y gene), and with the parent's genotype being GG/AG, the chance of the child getting the A allele is 50% (genotype = AG). However, the chance of developing hemochromatosis is very low for someone who carries only a single A risk allele. Mutations in other associated SNPS (in the S65C and HFE genes) may increase the relative risk, so the genotypes in those other SNPs must be taken in account as well.

  Sources:
  - http://www.snpedia.com/index.php/Rs1800562
  - https://www.23andme.com/you/journal/hemochromatosis/overview/

- **Alzheimer's disease**: an APOE4 allele occurs when both rs429358 and rs7412 contains the C risk allele. Since both parents carry a CC genotype at rs7412, the child will have a 50% chance of carrying one APOE4 allele (when rs429358 = CT and rs7412 = CC), with the remaining 50% chance of carrying two APOE4 alleles. From the lecture slides it was shown that with 1 copy of APOE4 allele the risk of AD will increase by ~2 times, while with 2 copies of

APOE4 alleles the risk will increase by ~11 times. Therefore the risk of having Alzheimer's for the child is high, with 50% chance having a 2x increased risk, and the other 50% having an 11x increased risk.

Sources:
- http://www.snpedia.com/index.php/Rs429358

- **Breast cancer**: the risk allele in BRCA1-associated breast cancer is a 'D' allele, which stands for a 185delAG frameshift mutation (an AG deletion) in the BRCA1 gene. Since the parents have genotypes of 'II' and 'DI', the child will have a 50% change of carrying a 'DI' allele. Lifetime risk of breast cancer for women is significantly increased even for carrying a single risk allele, rising from 12% to about 60%. Lifetime risk for males with a single risk allele is much lower, estimated to be around 6.9%.

  Sources:
  - http://www.snpedia.com/index.php/I4000377
  - https://www.23andme.com/health/BRCA-Cancer/

- **Cystic fibrosis**: the risk allele for cystic fibrosis is a 'D', which indicates an F508 deletion in the CFTR gene. This disease is inherited in a recessive manner, and since the parents each have one risk allele (D), there is a 25% chance of the child having cystic fibrosis (and 50% chance that he or she will be a carrier).

  Sources:
  - http://www.snpedia.com/index.php/Rs113993960
  - https://www.23andme.com/health/Cystic-Fibrosis/

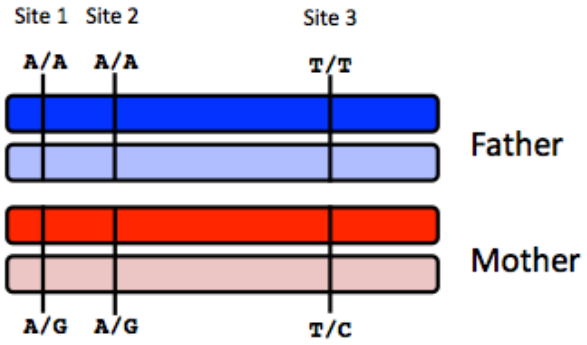
- **Sickle cell anemia**: the risk allele for sickle cell anemia is T, which encodes for the sickling form of hemoglobin, Hb S. However, since both parents have genotypes of AA, no risk allele will be passed on to the child, and the chance of the child developing sickle cell anemia is extremely low.

  Sources:
  - http://www.snpedia.com/index.php/Rs334

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	**55A**
If mother transmits G	59	50A

Mother transmits A: 45A / 45G + 5A / 5A = 55A
Mother transmits G: 45A / 45G + 5A / 5G = 50A

Site 2

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

Same as above

Site 3

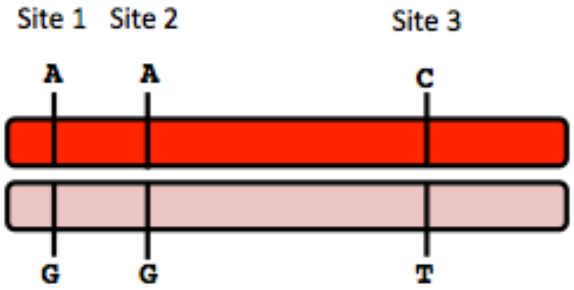
	T reads observed	T reads expected
If mother transmits T	49	55T
If mother transmits C	49	**50T**

Mother transmits T: 45T / 45C + 5T / 5T = 55T
Mother transmits C: 45T / 45C + 5T / 5C = 50T

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	TC

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.

I hypothesize that the Ala673Thr mutation on APP may be close to the proteolytic sites where cleavage by beta- or gamma-secretase occurs. The mutation affects APP's structure in such a way that it reduces the efficiency by which it is cleaved by the secretases, leading to the reduction of Ab peptide production and therefore a protective effect against Alzheimer's disease.

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

We can perform an in vivo assay by expressing beta- and gamma-secretase constructs in mammalian cells (e.g. HeLa or 293T) along with a Ala673Thr mutant APP construct (the control will express the wild-type APP construct). We can then measure the amount of Ab proteins produced over time by lysing the cells at different time points, perform immunoprecipitation of the Ab proteins on the lysates, and run a Western blot to detect for mutant/wild-type Ab (mutant Ab signal should increase at a slower rate than wild-type Ab).

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

Trait	SNP	Genotype
Ear Wax Type	rs17822931	TT = dry, CT/CC = wet
Hair Color	rs1667394	CC = brown, CT = usually brown, TT = blond
Eye Color	rs7495174 rs12913832	AA = blue/green, AG/GG = non-blue/green GG = blue/green eyes, AG = 50/50, AA = brown
Hair Type	rs17646946	GG = curly, AG/AA = straight
Asparagus Smeller	rs4481887	AA = smeller, AG = moderate smeller, GG = non-smeller
Bitter Taster	rs713598	CG/GG = can taste, CC = inability to taste
Milk Drinker	rs4988235	AA/AG = can drink milk, GG = lactose intolerant
Sprinter or LD	rs1815739	CC = sprinter, CT = moderate, TT = long distance
Height	rs6060371	TT = taller, GT = medium, GG = shorter
Has Pitch	rs3057	CC = has pitch, CT = moderate, TT = no pitch?

Person	Guess	Ancestry	Ear Wax Type	Hair Color	Eye Color	Hair Type	Asparagus Smeller	Bitter Taster	Milk Drinker	Sprinter or Long Distance	Height	Has Pitch
1	C	E Eu, Hungary	CC	CT	AA, AA	GG	AG	GG	AG	CT	GT	TT
2	H	Germany, Romania	CC	CT	AA, AG	GG	GG	GG	GG	CC	TT	CT
3	A	Asian	TT	TT	AA, AA	GG	AG	CG	GG	CT	??	CT
4	G	S Eu, mid E, Iraq Jews	CC	CC	AA, AA	AG	GG	GG	AG	CT	??	CT
5	D	Italy, Armenian, Jews	CC	CT	AG, AA	AG	AG	CC	GG	TT	TG	CT
6	F	UK	CC	TT	AA, GG	GG	AG	CG	AG	CT	GT	TT
7	E	N Eu, mid E, Turks, Romania	CC	TT	AA, GG	GG	GG	CC	AA	TT	TT	CC
8	B	Italy, Jews	CT	CC	AA, AA	GG	GG	CC	GG	CT	GT	TT

My guess is:

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