Bi188 Final Examination
Spring 2015

Due **Sunday, May 14th, 11:59pm**
(as a PDF emailed to sgoh@caltech.edu, phe@caltech.edu and woldb@caltech.edu).

The exam is closed-book and closed-notes, though you may use the genome browser if needed.

You are asked to draw in at least one question as well. This can be done onto a printout, or you can drop a .png into powerpoint, or other favored program and make your annotations that way. Also, you can draw while taking the exam and convert to electronic after time is up, or even hand in paper.

You have three and a half hours to complete the exam, though we expect that the exam can be completed well within 2 hours.

There are 6 parts to this exam that adds up to a total of 37 points. The final score will be scaled to 35 points.

Express your answers concisely. Most questions need only one or a few sentences.

Please read no further until you are ready to take the exam.

1. **8pts total** Consider what you have learned about how histone modification patterns, sequence specific transcription factor occupancy, DNAse hypersensitivity are related to active transcription. Assume for this part of this question that RNA levels faithfully reflect transcription from each gene. A developmental progression below consists of early progenitors, committed progenitors, and newly differentiated cells, and a 4th sample is from fully differentiated long-term mature cells. RNA levels are given in the table FPKM from polyA RNASeq.

	Gene X	Gene Y	DNAP	DNAQ	DNAR
Early (Day 3)	0	68	0	43	0
Committed (Day 5)	12	70	20	35	0
Newly Differentiated (Day 7)	26	9	16	1	20
Long-term adult differentiated (2 months)	29	0	14	0	2

For X and Y, each gene has a functionally important distal CRM located ~10 KB downstream of the polyA addition site of each gene, in addition to proximal elements within the 500NT upstream of the TSS. Sequence specific DNA Regulatory Factors P, Q and R are known to be direct regulators of X and or Y.

a) Estimate the half-life of Gene Y mRNA (in hours). **(1pt)**

b) What are the most likely regulatory functions of DNA factors P,Q, and R with respect to X and Y? **(1pt)**

c) Draw a cartoon sketch of a browser map for data from gene X at each stage with data tracks for ChIP-seq of Pol2, DNAse-seq, and P,Q,R. **(2pts)**

d). For both X and Y, show in rough cartoons, two different plausible patterns for the histone marks H3K27Ac; H3K4me1, and H3K27me3 at X and Y over the two time-points, highlighting the distinctions. **(2pts)**

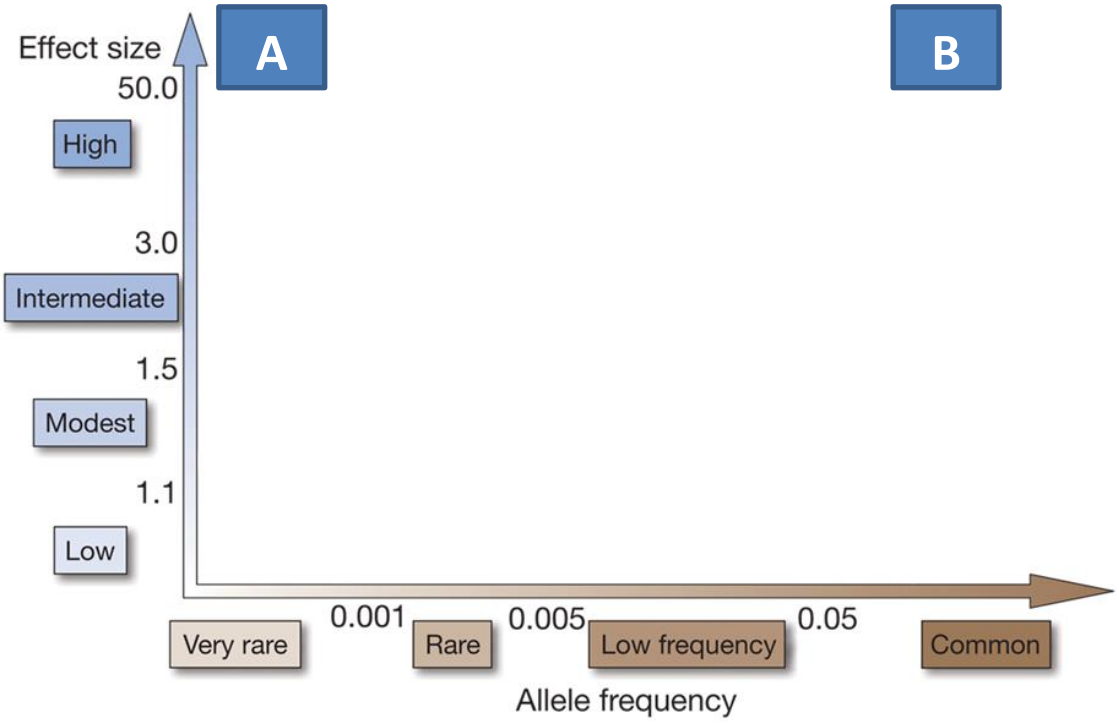
e) What is the core regulatory principle illustrated when you consider Q,P,R occupancy and DHS versus the histone mark patterns you made for part (d) in a temporal regulatory series such as this? **(2pts)**

2. 5pts total There is a noncoding GWAS snp hit located in a large second intron of the gene X from question 1. You want to find the causal variation that accounts for the high-scoring SNP. Roughly how likely is it that the causal variant will turn out to be in an exon versus noncoding sequence? **(1pt)**

How would you use the LD map in the region to judge if it is likely that the CRM located 10kb downstream is responsible? **(1pt)**

How would you use conservation information in the region of the CRM to evaluate 2 variants? **(1pt)**

Briefly, design a functional test to discriminate candidate causal variants from non-causal candidates. What would persuasive results look like? **(2pts)**



3 Six points total (plus 2 possible extra credit). The diagram above allows you to consider the phenotypic effects that alleles of varying frequency have on human health. Based on the material discussed in class (or any reading you have done), please provide an example of a:

A. A rare disease with very severe health effects whose genetic basis is explained by low frequency alleles in the population (Highlighted by Box A). Please discuss the basic inheritance pattern of this disease in 2-3 sentences**. (2pts)**

B. A common disease with severe health effects whose genetic basis is explained by alleles of relatively high frequency in the population (Highlighted by Box B). In a few sentences, please discuss whether or not the aggregate current data suggest that most common diseases fall into this category. **(2pts)**

Extra credit 2pts: Provide a theory based on what you know about natural selection to explain your answer above.

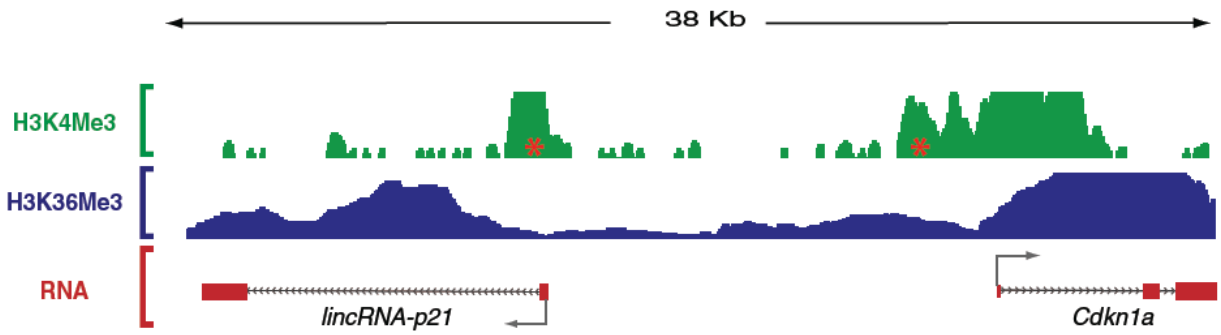
C. Briefly discuss what is known about the relative genetic contribution that high frequency (common) alleles and low frequency (rare) alleles have on the risk of developing (i) type 1 diabetes, (ii) type 2 diabetes, and (iii) mature onset diabetes in the young (MODY). **(2pts)**

4. 7 points total (plus one point extra credit optional). TP53 is a pivotal regulator you have studied in multiple contexts.

a) Name at least two pathway(s)/process(es) that p53 is known to be involved in. **(1pt)**

b) What is the expected phenotype of individuals with a germline loss-of-function mutation in one allele of p53? (**1pt)**

c) In a recent study, a large intergenic non-coding RNA, called lincRNA-p21 was identified as a regulatory target of p53. The model is that lincRNA-p21 mediates p53-induced repression at target genes. Given the diagram below, what is a pertinent shared feature of the histone marks over lincRNA-p21 and its neighbor gene Cdkn1a? **(1pt)**



d) It is hypothesized that p53 directly binds the lincRNA-p21 promoter (labeled with a red star). How could you test this possibility bioinformatically without new experimental data? (**1pt)**

Now assume you have good new ChIP-seq data using a p53 antibody. Some of the processed data are shown below. The genomic coordinates of lincRNA-p21 and Cdkn1a are also attached. Do you still believe that p53 binds at the promoter of lincRNA-p21? Why or why not? (**1pt)**

(Extra credit point) What additional discovery can you make from the data? **(1 pt)**

Gene name	Chromosome	Strand	Start	End
lincRNA-p21	Chr17	+	29227924	29236472
Cdkn1a	Chr17	-	29194419	29215906

Chromosome	ChIP-seq Position	Enrichment
chr17	28345215	524.63
chr17	28345215	524.63
chr17	28360410	161.00
chr17	28360410	161.00

chr17	28367581	418.06
chr17	28367581	418.06
chr17	29215997	23724.00
chr17	29227883	49203.80
chr17	29228540	64.21
chr17	29230807	6.05
chr17	29231448	389.91
chr17	29277807	232.906
chr17	29297233	102.044

e) Design an experiment to test whether lincRNA-p21 mediates some, all, or none of the p53's repressive role in the mouse system. What result would support the hypothesis that lincRNA-p21 is playing an essential role and what would not? Keep it simple. **(2pts)**

5. (5 points total) Based on the properties of common allele risk factors for the common diseases, briefly discuss the pros and cons of direct-to-consumer genetic testing for determining genetic disease risks. In your answer, discuss specifics using one of the following conditions to make key points: Alzheimer disease, age-related macular degeneration, or diabetes.

There is no single 'correct' answer in the balance of pros and cons, and the issues are not identical for all genes. Full credit will be given for a well-reasoned assessment of the pros and cons, including human genetics based on information from GWAS; professional medical community considerations; data types (this is not meant to be about specific companies and their tests, but about the range of data offered).

This is a free form question, but please keep the answer as fact-based as possible with an upper size of about 10-12 sentences. It can be done in less, and being concise is desirable. Please budget your time accordingly.

6. (6 points total, 1 extra optional) You have fresh biopsies from a triple-negative breast cancer tumor – it turned up negative on the better-known classifications for breast cancer (ER, her2, PR). Imagine you are charged with doing custom analysis to advise the tumor board and physician on treatment in addition to surgery. You will have the primary surgical specimen to work with.

a) How does triple-negative status affect treatment options? Specifically, identify two major molecularly targeted treatments that are ruled out for this case **(1 pt)**

b) Would you test for BRCA mutations? Why or why not with respect to treatment of this individual? If you tested and found the patient to be germline BRCA2 positive, what are the implications for her future care? What specific actions would she and her doctor could consider to alter her risk profile? **(1 pt)**

You want to determine if any potential oncogenes are overexpressed.

c) Describe two methods of finding out which oncogenes are overexpressed. Include the material needed and the actual type of information you can get from both assays. Be sure to describe controls should you use. **(2 pts)**

d) The assay you described above identified a transcription factor X that is highly overexpressed relative to its expression in normal breast tissue. Design experiments and show sample data to (i) screen globally for candidate regulatory targets of X, and (ii) to test definitively which of your candidate targets is a regulatory target. **(2 pts)**

bonus point: how could you discriminate a direct molecular target of X from a strong indirect target?