Bi188 Midterm Examination
Spring 2013

Due **Saturday, May 11th at 5:00 PM**
(as a PDF emailed to kfisher@caltech.edu, sgoh@caltech.edu, and woldb@caltech.edu).

The exam is closed-book and closed-notes.

You have 2 continuous hours to complete the exam.

The questions are 10 points total each.

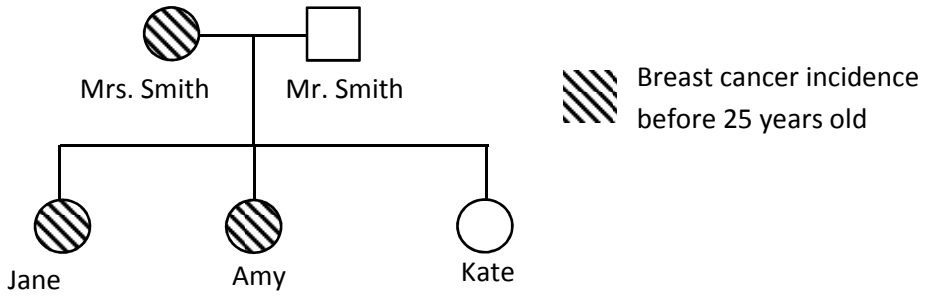
You may choose 3 questions to answer. If you answer four questions, we will grade the first three.

Express your answers concisely. When we ask for an experimental design, it is possible to give it in a few sentences so that the essence of the idea is there. Say what appropriate controls would be, but again, it is the appropriate idea we are asking for, not procedural detail.

Reminder: the exam is **CLOSED-BOOK/NOTES**.
Read no further until you are ready to take it!

Question 1

For this question, you are an investigator with a collection of primary breast tumors, normal breast tissue taken at the time of surgery, plus blood samples from each patient and their immediate family. One of the families you are studying has the following incidences of breast cancer:



You wish to identify potential tumor suppressor genes, either known or novel, whose alteration (directly in the coding sequence or otherwise) are linked to the breast cancer incidences in this family.

A) Identify two types of genomic and/or functional genomic data you would gather for this purpose. Briefly explain how you would obtain the data from currently existing or developing experiment methods. Illustrate an idealized result for each kind of data and say how you would interpret it.

B) Amy and Kate are identical twins, but obviously show different phenotypes in this incidence. Give one possible explanation to how this is possible, in the context of your identified candidate tumor suppressor genes. If your hypothesis stands, is Kate at a very high risk of developing breast cancer in the immediate future? Why?

C) You found that a very promising mutation that is X-linked. Given your explanation in (B), how likely it is that Mr. Smith carries this mutation? Explain succinctly why.

Question 2

When investigating a rare disease that is reportedly familial (and therefore likely has a heritable genetic component), you ran Western blots against several candidate proteins using samples from both blood and mouth epithelial cells. You found an unusually high amount of protein X in the patients compared to control individuals. While X is significantly expressed in these different source tissues, the relative levels (ratios) compared to a sample of 20 housekeeping gene proteins* varies by nearly 10-fold (huge!) from one family to another.

* housekeeping genes are genes expressed in all cell types, such as genes coding for ribosomal proteins, proteins in basic metabolic pathways, cytoskeleton, etc.

A) Briefly explain how structural variation in the genome could explain results described above. To get full credit, you will need to draw allelic versions.

B) Measuring protein levels via Western blot might be a study by somebody else that sets you down the path to thinking that CNV (copy number variation) is part of a possible genetic explanation. As a genome scientist, what are the most direct forms of evidence for CNV at the locus of interest? Give 2 different kinds of measurements that would be informative and complementary to each other.

Question 3

C-Myc is a long-known and famous oncogene. You recall that it is involved in positively regulating progression of cells through the cell cycle normally. Its experimentally enforced expression can be used, combined with other activated oncogenes such as a Ras family members (activated by point mutation), to convert a normal cell into a tumor phenotype. You learned that in some tumors, gene amplification of a Myc family member gene contributes to tumorgenesis, and is associated with very high level RNA and protein expression.

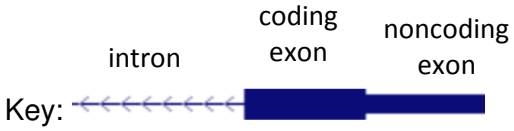
You are analyzing the differences between an archival (pre tumor) blood sample, and a sample from this patient's newly diagnosed lymphoma. After a standard RNA-Seq analysis* between the two cell types, you notice that c-Myc, one of the Myc genes, is expressed at 8 times the level in the tumor sample as in the healthy sample. Knowing that Myc genes are often amplified by copy number variation in tumors, you check to see if this is the case for c-Myc in the tumor sample – but it is not. The tumor sample also shows no mutation in the c-Myc coding sequence, based on direct DNA sequencing from EXOME capture.

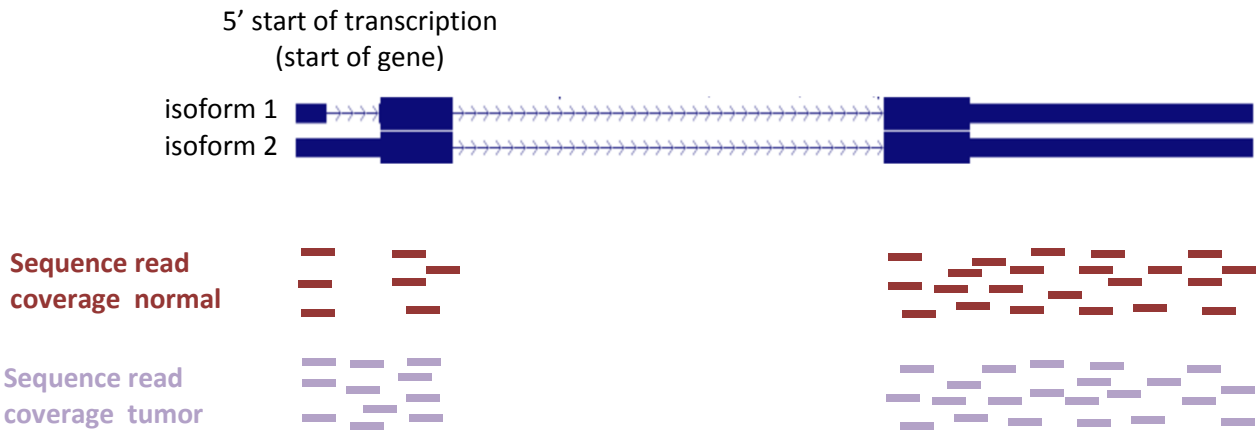
* Don't worry about the details of RNA-Seq measurements. Focus on the outcome data.

A) What kind of mutational events could have led to c-Myc being expressed at such high levels in this patient's tumor? We are asking for the nature of the mutation – how has the normal gene been altered functionally to lead to this outcome (i.e. NOT "HPV or faulty mismatch repair caused this type of mutation" but "there is X type of mutation in Y location relative to the anatomy of Myc (or other pertinent gene) and it affects the expression of *c-Myc* in Z way"). Give us two different possibilities. Note that one possibility has great historic precedent and is famous. You are free to use that, and add another plausible one, or simply give us two that should do the trick, in theory.

B) Now give an experimental out-line to test for one of your possible models. Say what resulting data would confirm your hypothesis.

Next, you analyze the differences between a second patient's healthy colon tissue and his colon tumor. After a standard RNA-Seq analysis, in which the informatics produced an aggregate level expression value for all isoform models (locus level quantification) between the two cell types, you find the following: levels of 5 other cell proliferation genes are much higher (4-8 times higher) in the tumor than the healthy tissue, but there are no detectable differences in RNA expression for any Myc family member. Frustrated, you decide to look more closely at the structure implied for RNA coming from the l-myc locus. Here you do find a difference:





C) What mutation do these RNA-Seq data suggest has occurred in the colon tumor cells? Answering in one sentence is sufficient, but specify the location of the mutation and suggest the underlying molecular mechanism.

D) How could such a mutation explain why genes regulated by a Myc transcription factor are extremely upregulated in the RNA measurements, while the overall levels of Myc RNA are similar in tumor and normal samples? Give us a very basic mechanism. One sentence would do it.

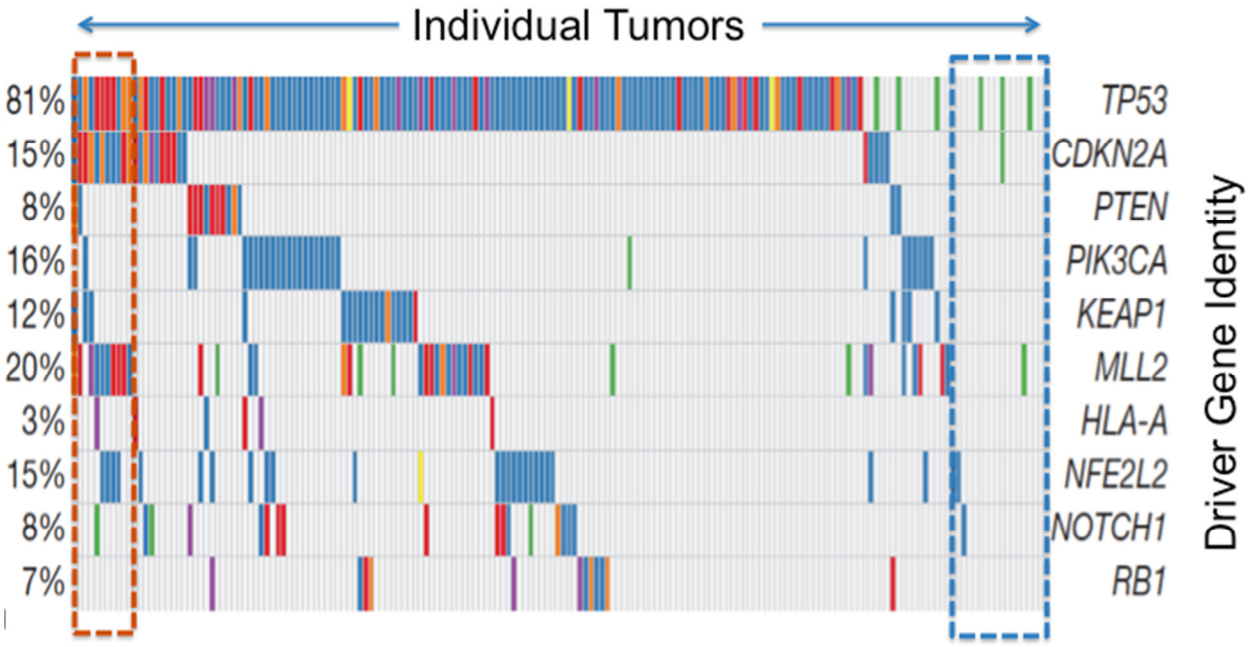
Question 4

A) Chromothripsis – What is it? How would you postulate that it could contribute to tumor development?

B) Tumors evolve and become resistant to treatments. If you were going to study this in a tumor of white blood cells, how would you design a study to ask if two sequential relapses were descended from a single founder clone of cells or arose from different founder clones?

C) From the data in the figure below showing results of genomic characterization of lung tumors:

Give your conclusion about the role of p53; your expectation about what explains the sparsity of highlighted candidate driver genes in the group boxed at the right; and how you would, in theory. use this information to suggest that an MD colleague of yours treat some cases differently from others. How would you expect this picture to change if 10,000 more tumor/normal pairs were added to these data?



D) Mutations in the CFTR gene give rise to cystic fibrosis (CF). It is a classic Mendelian recessive, yet the severity of the disease and its course varies tremendously. If you make the simplifying assumption that all variation between CF patients is genetic in origin and that all variation maps to the CFTR locus, what can explain the differences between patients? How does this affect what a physician should prescribe from among multiple drugs available (include newly developed experimental drugs discussed, even though they would not be available for routine clinical use, in real life)? Should the physician simply try each one?

E) Huntingtin disease (HD) is due to so-called triplet repeat instability in the huntingtin locus How do different huntingtin alleles – normal versus disease causing -- differ from each other (answer in one sentence)? If your father died of HD and your daughter decided to be tested, what does a positive result (she has an HD disease allele) or a negative result (she has only normal huntingtin alleles) tell you about your own HD status (you have not been tested yourself)?