**Spring 2016 Genomics Exam #2**
**transcriptome, metagenome, epigenome, proteome**

There is no time limit on this test, though I don't want you to spend too much time on it. I have tried to design an exam that will take less time than exams in the past. You do not need to read any additional papers other than the ones I send to you. There are 7 pages, including this cover sheet, for this test. You are not allowed discuss the test with anyone until all exams are turned in no later than 11:30 am on Wednesday March 23. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE BY 11:30 am ON WEDNESDAY MARCH 23**. You may use your notes, papers we read, and the internet. However, you are not to look for the source paper from which the figures were taken. You may work on this exam in as many blocks of time as you want. Submit your electronic version before 11:30 am (eastern time).

The **answers to the questions must be typed in a Word file and emailed to me as an attachment**. Be sure to backup your test answers just in case (I suggest a thumb drive or other removable device). You will need to capture screen images as a part of your answers which you may do without seeking permission since your test answers will not be in the public domain. Remember to explain your thoughts in your own words and use screen shots to support your answers. Screen shots without ***your*** words are worth very few points. Support your answers with data using screen shots liberally (no permission required since your exam is a private document).

*DO NOT READ or DOWNLOAD ANY PAPERS FOR THIS EXAM. RELY ON YOUR EXPERIENCE, AND YOUR SKILLS.*

**-3 pts if you do not follow this direction.**
**Please do not write or type your name on any page other than this cover page.**
Staple all your pages (INCLUDING THE TEST PAGES) together when finished with the exam.

Name (please type):

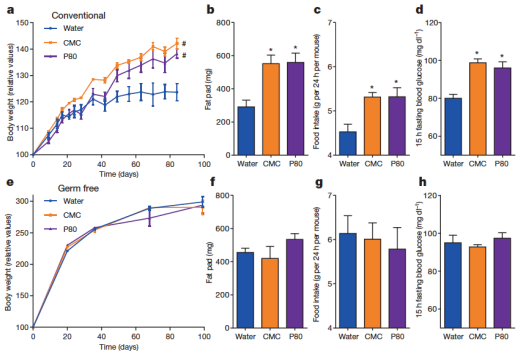
Write out the full pledge and sign (electronic signature is ideal):

How long did this exam take you to complete?

**16 points**

1) The intestine is protected from its microbiota by a multi-layered mucus which keeps the majority of gut bacteria a safe distance from epithelial cells that line the intestine. Agents that disrupt mucus–bacterial interactions might disrupt this natural balance. Detergent-like molecules that are ubiquitous in processed foods could disrupt the mucus layer and lead to harmful host responses. Investigators added to the drinking water of mice low concentrations of two commonly used emulsifiers, carboxymethylcellulose (CMC) and polysorbate-80 (P80). # and * both $p < 0.05$. Panels a – d are from one experiment and e – h another experiment.

a) Interpret the data in panels a – h. Don't go through each panel in detail. Instead, provide to me four major conclusions based on these data alone. Limit your answers to a maximum of 40 words per major conclusion.



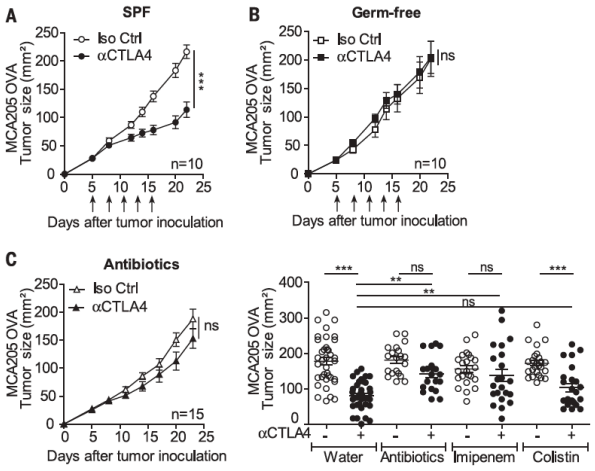

1. emulsifiers lead to increased weight, fat and appetite if microbiota is left intact.
2. emulsifiers lead to decreased glucose tolerance if microbiota left intact.
3. outcomes 1 & 2 above do not happen in GF mice.
4. GF mice have higher weight, fat and appetite than conventional mice.
5. There were other acceptable answers.

b) Design an experiment so you could determine in panel a if the composition of species changes or if the relative abundance changes in the microbiota? Limit your answers to a maximum of 50 words.

either amplify 16S RNA and then sequence or do total shotgun sequencing on microbiome extracted from feces. Relative reads could tell you change in abundance and OTUs could indicated possible changes in species.

**16 points**

2) Mice with a melanoma (caner cells called MCA205 OVA) were injected with an antibody that binds to CTLA-4 (antibody is called αCTLA4). "Iso Ctrl" is a negative control antibody. SPF = specifically pathogen free housing, meaning the mice would not get any harmful infections but they were NOT germ-free. Germ-free means the mice have never been exposed to any bacteria. Arrows in panels A and B show when αCTLA4 was administered to the mice. Mice in panel C also were

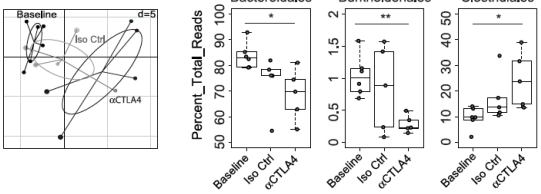


SPF; panel C has left and right parts but it is not the same mice for left and right parts. *p < 0.05; **p < 0.01; ***p < 0.001; ns, not significant. Antibiotics = ampicillin + colistin + streptomycin.

a) Interpret the data in panels A - C. Don't go through each panel in detail. Instead, provide me with four major conclusions based on these data alone. Limit your answers to a maximum of 50 words per major conclusion.

1. In SPF mice, αCTLA4 slows growth of melanoma tumor compared to negative control antibodies.
2. αCTLA4-induced slowed melanoma growth is lost GF or antibiotic (cocktail) -treated mice.
3. Only a subset of the microbiota is responsible for αCTLA4 response: antibiotic colistin did not abolish αCTLA4 effect.
4. Treating with colistin is no more effective than treating with water, so the antibiotic does not enhance αCTLA4 effect.
5. there are other acceptable answers

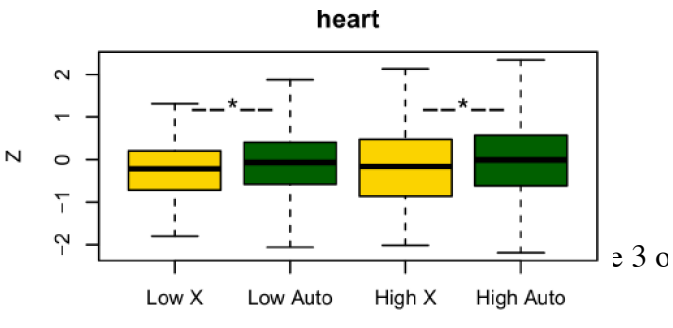
b) 16S RNA PCR primers amplified bands using extracted microbiomes from mice prior to treatment (baseline), and 48 hours after one injection of αCTLA4 or Iso Ctrl. 16S amplicons were sequenced and PCA (first panel) was performed on relative abundances of the OTUs. The next three panels quantified the indicated orders of bacteria. *$p < 0.05$; **$p < 0.01$; *** $p < 0.001$; ns, not significant. Use these data to propose an experimental treatment that might improve the outcome of melanoma cancer patients treated with αCTLA4. Support your answer with data. Limit your answers to a maximum of 50 words.

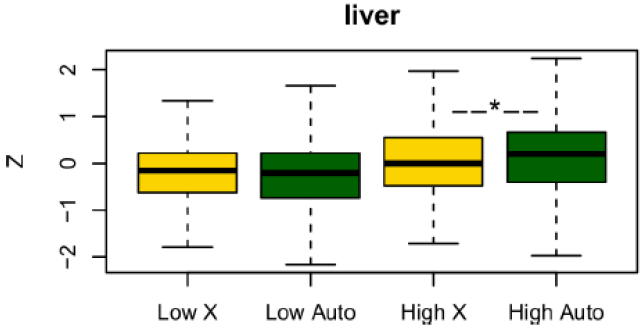



treat mice with massive antibiotic to clear out microbiota and then supplement with Clostridiales before administering αCTLA4. The converse was also acceptable with the argument that Bacteroides and Burkholderiales population decline is part of the beneficial outcome for αCTLA4.

**10 points**

3) Here are some transcriptome data from adult humans. RNA from many tissues and organs were sampled, two of which are shown. Genes were subdivided into those on the X chromosome *vs*. autosomes, and then further separated into low or high levels of transcription. * indicates significant differences ($p < 0.01$).

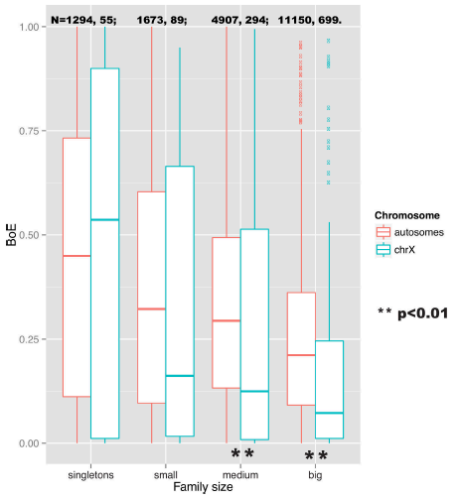


a) Categorize the level of transcription for genes on autosomes vs chromosome X based only on the data above. Limit your answers to a maximum of 30 words.

Genes on the X chromosome are transcribed significantly less than genes on autosomes. This is true for highly transcribed genes, and in some tissues, genes that are not transcribed as high (low).

b) These data examine human genes for which there are no paralogs (singletons), those with one paralog (*small*), those with a *medium* number of paralogs (3-5) and those with a *big* number of paralogs (>5). They quantified for each paralog size category the relative "breath of expression" (BoE) as an indication of the dynamic range of transcription in all tissues over all time points. What is the general trend exhibited in these data? Limit your answers to a maximum of 40 words.

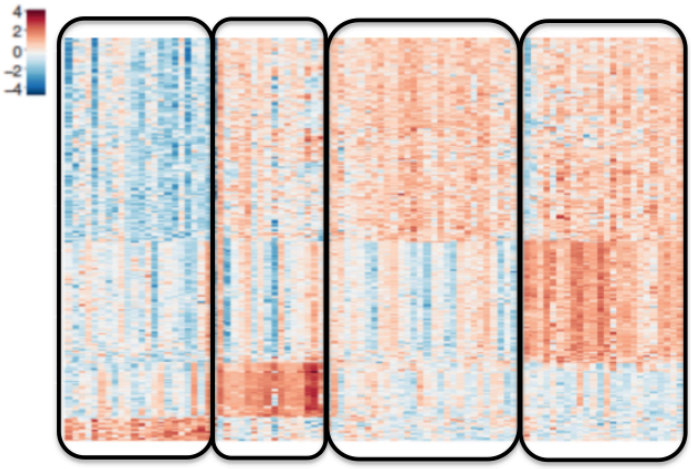



The more paralogs, the lower the dynamic range of transcription. With 3 or more paralogs, then sex-linked genes have lower dynamic range than paralog families on autosomes.

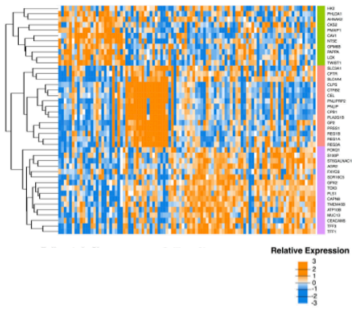
**12 points**

4) Here are some RNAseq data for a type of cancer.

a) Use the PDF version to inspect these data closely. Implement a visual *k*-means clustering and draw *k* boxes around the types of cancers you see here. Genes are listed on the Y-axis and biopsies are listed along the X-axis. Genes were clustered based on correlation of expression profile using reads normalized by RPKM. Samples were clustered based on similarities of gene expression profiles. Print the PDF data and draw your boxes on it, then scan or photograph your answer.



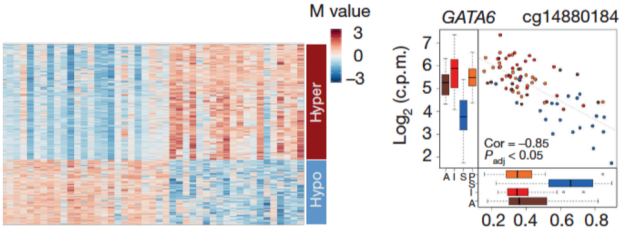

b) Here are similar data from different samples of the same cancer as in 4a. Genes are listed on the Y-axis and biopsies are listed along the X-axis. Genes were clustered based on correlation of expression profile using reads normalized by RPKM. Samples were first clustered by category and then by similarity of gene expression profiles. The 3 colored boxes on the right



denote the 3 categories used to subdivided the cancer into 3 categories using histological characteristics instead of RNAseq data. Evaluate the validity of each of the 3 categories based on the data in question #4. Limit your answers to a maximum of 40 words per color.

Two key points: k = 4 so one of the three categories (rows) had to be split up; justify which genes show a category needs splitting either by eye or better yet, by dendrogram on the left side.

**10 points**




5) Here are some data connecting the transcriptome and the epigenome. The left panel shows the methylation states of some genes (Y-axis) in two (of four) types of cancer from 96 patients (columns). The right panel shows the level of one gene (*GATA6*) and its normalized level of transcription (c.p.m.) in all four types of cancer (A, I, S or P) and *GATA6's* level of methylation (0 – 1 scale with 1 fully methylated). *GATA6* is located on the Y-axis of the left panel at the same height as the letter y in Hyper.

a) Label (A, I, S or P) the two types of cancers in the left panel best you can based on the data in this one figure. Support your answer with data. Limit your answers to a maximum of 40 words per each of the two cancers.
left cancer = __P, I, and/or A____ (type)
support: decreased methylation

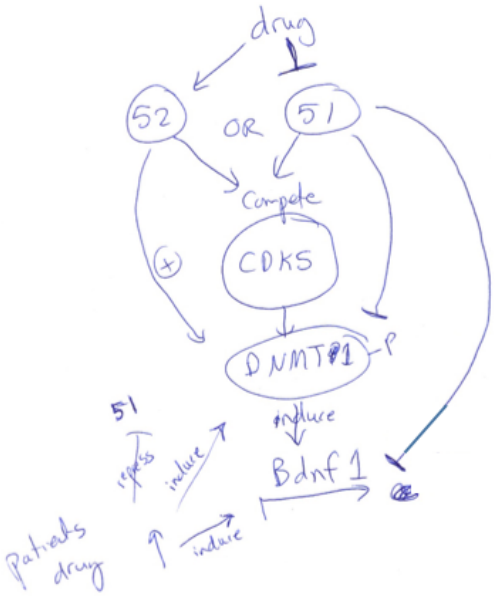
right cancer = ____S_____ (type)
support: increased methylation

b) What is being methylated is this figure? Explain your logic behind your answer given ONLY the data in this panel and what we have learned in class. Limit your answers to a maximum of 30 words.
Either DNA or histone (H3K27me3) assocaited with repressed transcription.

**12 points**

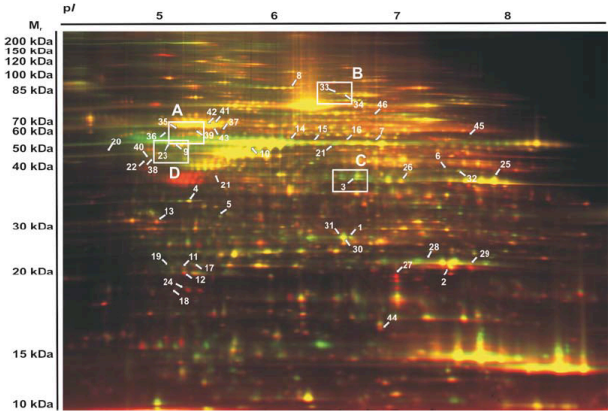
6) Use the information from this paragraph to draw a picture that illustrates what is described here. Be sure to include all the genes and proteins described as well as the clinical outcome. A good answer is one that a biology major who has not had genomics could look at and figure out what is going on using only your diagram and any labels. DO NOT write supplemental text. Communicate all of this in your figure with labels. Draw your answer on a blank piece of paper, either by hand or electronically. If you draw by hand, scan or photograph your diagram.

DNA methylation and chaperone proteins have been implicated in mental disorders and the action of antidepressant drugs. Chaperone FKBP51 associates with cyclin-dependent kinase 5 (CDK5), which can covalently modulate and activate DNA methyltransferase 1 (DNMT1). FKBP51 competes with its paralog FKBP52 for binding to CDK5. Human calls can produce FKBP51 which displaces FKBP52 from CDK5 and reduces the interaction of CDK5 with DNMT1. In tissue culture, FKBP51 facilitates several effects of the drug paroxetine such as protein-protein interactions of DNMT1 with CDK5 and FKBP52 which leads to induction of brain-derived neurotrophic factor (*Bdnf* ) in the cerebrum. In human blood cells, FKBP51 expression is inversely correlated with genome-wide and *Bdnf* methylation. Blood cells isolated from depressed patients were treated *ex vivo* with paroxetine. BDNF protein levels increased as did phosphorylated DNMT1, but FKBP51 went down in cells from patients who responded favorable to paroxetine in the clinic.




**12 points** proteome

7) Here are some colorful proteomic data using the 2D-DIGE method (look it up) that used microdissection (look it up) to isolate cancerous cells (red) and adjacent healthy cells (green).




a) Identify one spot that could be used as a "healthy cell" marker and another one that could be used for a "cancer marker". Explain why you chose the spot you did using data from this figure. Print your .png file and scan or photo your answer.

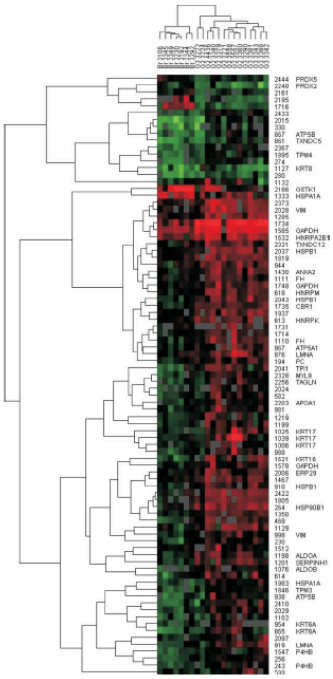
any red spot separated from other spots (cancer) or green spot (healthy)

b) How could you identify which proteins you have chosen from this gel? Because the samples are human biopsies, you cannot redo the experiment or use fresh cells. Limit your answers to a maximum of 40 words.

had to use MS/MS to sequence protein

**12 points** proteome

8) Here are some proteomic data showing proteins abundant in cancerous bronchial cells or healthy bronchial cells. Red squares in the heat map represent high spot intensities, whereas green squares represent low spot intensities. The proteins were clustered first based on their abundance (Y-axis) then the samples were clustered (X-axis). Healthy sample names start with Br in their names; G2 and G3 refer to cancer grades 2 and 3, respectively. High grade numbers indicate more advanced cancer.



a) Choose one protein number (on Y-axis) in the figure to use as a potential biomarker for bronchial cancer and another biomarker for healthy cells. Support your answer with data in the red/green figure only and you cannot use any proteins appearing in the figure below. Limit your answers to a maximum of 40 words for each marker.

cancer biomarker # : several good ones such as 2422 or 1735

healthy biomarker # : only good one was 1716

b) Now look at the data in the figure that used the method of immunohistochemistry. How could you use this panel of 5 antibodies to classify which form of bronchial cancer a patient has? Don't worry about the immunohistochemistry method details, just focus on designating which antibody/antibodies you would use for each of the three types of bronchial cancer shown in this figure. Display your answer in the form of a chart that is easy to read. You should not need any additional text to explain your answer.

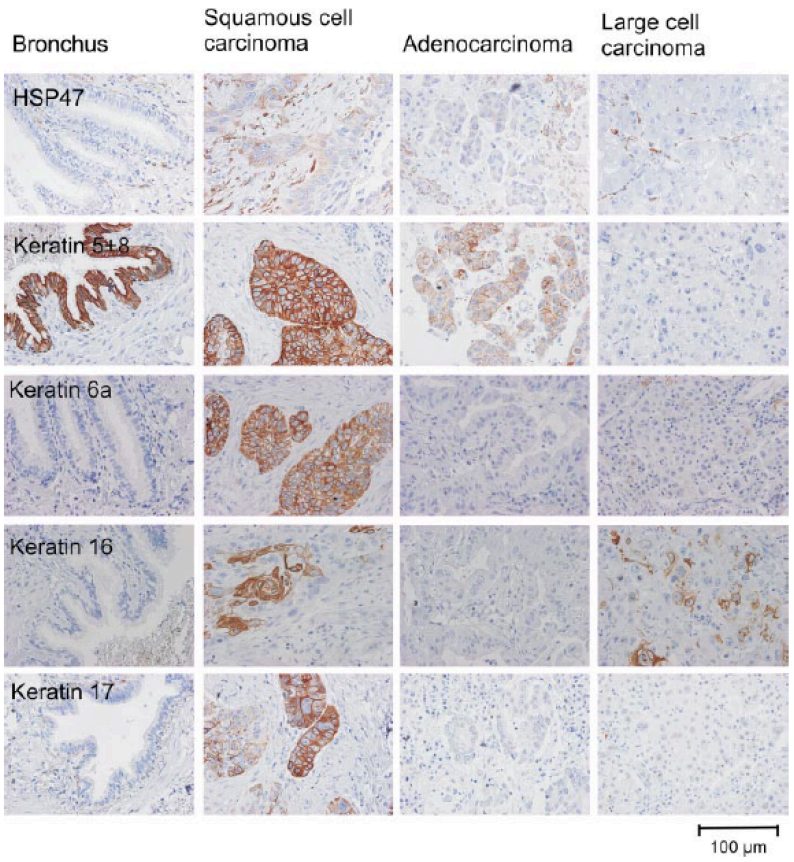


Chart Here:
squamous: + with K17 (uniquely so)
adenocarcinoma: + with K 5 + 8 but not 17, 6a, or 16
large: + with K16 but negative with K 5 + 8