Spring 2018 Genomics Exam #1
Metagenomics, Transcriptomics & Epigenomics

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 that exams in the past. You do not need to read any additional papers. There are 4 pages, including this cover sheet, for this test. You are not allowed discuss the test with anyone until all exams are turned in at 9:30 am on Wednesday March 28. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE BY 9:30 am ON WEDNESDAY MARCH 28**. You may use a calculator, a ruler, your notes, the book, and the internet. You may work on this exam in as many blocks of time as you want. Submit your electronic version before 9:30 am (eastern time zone).

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 storage). You may 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.

DO NOT READ or DOWNLOAD ANY NEW PAPERS FOR THIS EXAM. You may search and read abstracts. RELY ON YOUR EXPERIENCE, AND YOUR SKILLS. Spell out your logic for each answer.

-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.

Name (please type):

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

How long did this exam take you to complete?

20 pts

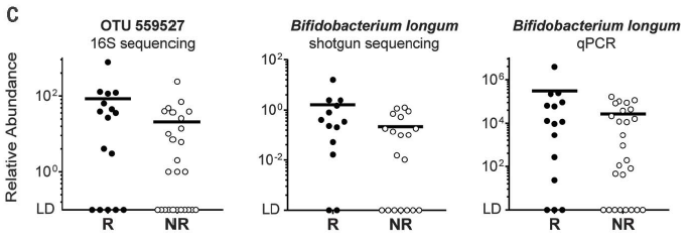
1) Below are some microbiota questions. In this study, the investigators were looking at the effect of microbiota on clinical outcomes for patients with melanoma and treated with anti-PD-L1 antibodies. Sixteen patients responded well (R) and 26 did not (NR). In panel C, each symbol represents one person in this study; LD = limit of detection. In panel D, the qPCR score represents relative abundance of 10 pooled OTUs with differential abundance between R and NR microbiota.

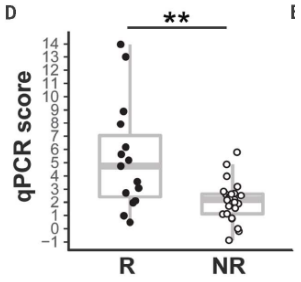
a) What can you conclude about the role of patient microbiota on melanoma clinical outcome based on the combined data in panels C and D? Support your answer with data. Limit your answer to a maximum of 50 words.

no significant differences in C, but trends

D shows significance but in a collection of 10 OTUs → combinatorial synergy

correlation, not causation





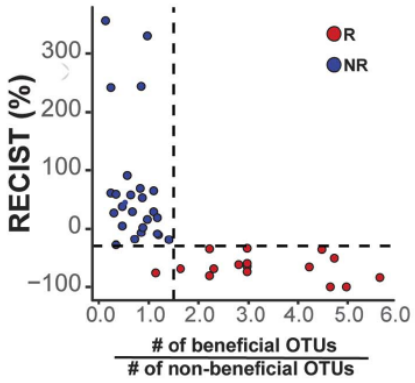
The figure to the right shows some more data from this same study. The RECIST value is a measure of change in tumor size before and after treatment for the 42 patients.

b) Interpret these data. Support your answer with data. Limit your answer to a maximum of 50 words.

most responders have ratio > 1

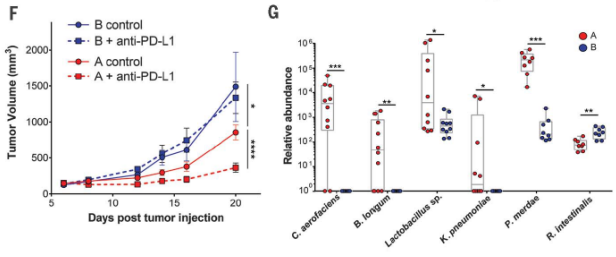
most NR have ratio < 1

ratio >1 correlates with smaller tumor size



These final figures (F & G) show what happened to germ-free mice given fecal transplants from R patients (A) or NR patients (B).

c) Interpret the two main points in panels F and one main point from G. Support your answer with data. Limit your answer to a maximum of 50 words.



F: R (A) microbiota correlates with smaller tumor value

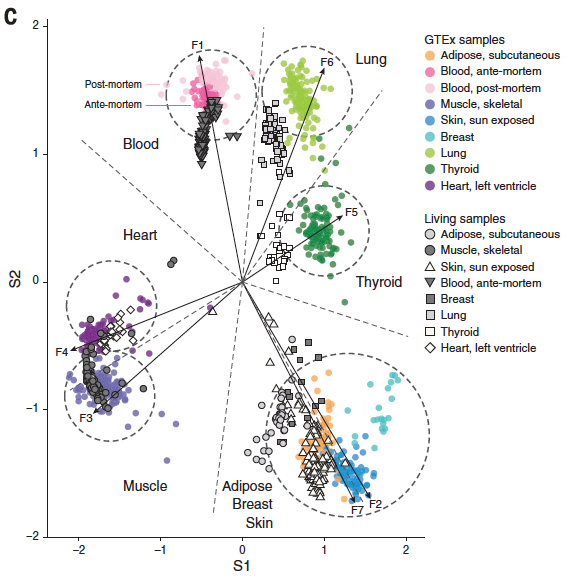
F: A mice respond better to anti-PD-L1 therapy

G: Significant difference in all 6 species, with A having more except last one (synergy but correlation)

d) What control could have been added to strengthen their conclusions? Limit your answer to a maximum of 30 words.
GF +/- therapy
add antibiotics to A and B treatments

20 pts

2) Here are some transcriptome questions based on GTEx RNAseq data. Expression values from eight GTEx tissues (colored circles) plotted radially along seven metagenes extracted from expression data. The ante-mortem samples are curated from the Gene Expression Omnibus (GEO). Each shape represents a collective transcriptome value for a single sample from a single person. The X and Y axes are the two scaling (S) factors that explain most of the variance in the data (analogous to PCA).



a) What can you conclude about RNAseq variation between tissues and between people? Support your answer with data. Limit your answer to a maximum of 40 words.
RNAseq data for tissues vary more than different people

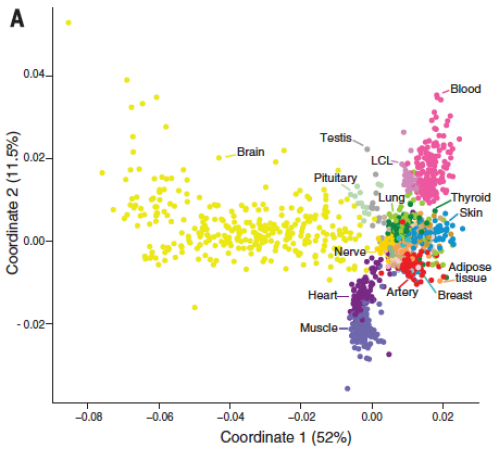
b) What can you conclude about post-mortem samples? Support your answer with data. Limit your answer to a maximum of 30 words.
post-mortem is surprisingly similar to ante-mortem, but post-cluster varies more than ante-.

c) What can you conclude about living sample sources vs. GTEx samples? Support your answer with data. Limit your answer to a maximum of 40 words.
They are similar to GTEx samples, but can vary more along S2 axis.

d) What do the arrows F1 – F7 represent? Support your answer with data. Limit your answer to a maximum of 40 words.
organ-specific averaged gene expression as a single vector (breast lacks its own vector)

20 pts

3) The same group also examined alternative splicing from the GTEx samples. In the figure to the right, you see multidimentional scaling (similar to PCA) analysis of all samples such that values show additional exons included (+ values) or excluded (- values) in the mRNA.



a) Explain the amount of variation in splicing for most tissues, excluding brain, blood and heart/muscle. Support your answer with data. Limit your answer to a maximum of 30 words.

very little alternative splicing compared to other 3 tissues

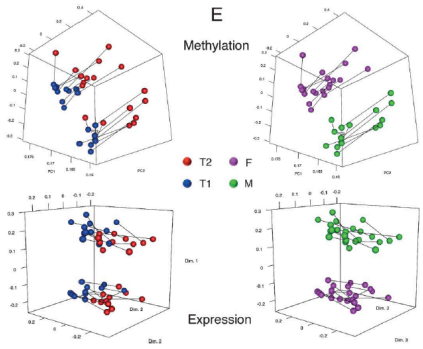
b) Explain the amount of splicing variation in heart/muscle vs. blood. Support your answer with data. Limit your answer to a maximum of 40 words.

Blood has more variance for both coordinates in added exons than heart/muscle which has loss of exons in coordinate 2.

c) What can you conclude about the brain samples? Give an example we covered in class that is consistent with your conclusion. Support your answer with data. Limit your answer to a maximum of 50 words.

Brain has more loss of exons (coordinate 1 and mostly added exons in coordinate 2). This is consistent with *Dscam* gene in flies that produces > 38,000 mRNAs in neurons

20 pts



4) Now for some epigenetic questions… Some investigators wanted to understand the role epigenetics plays in physiological responses to regular exercise. They got volunteers to exercise one leg but not the other (:-o) every day for 3 months (T1 = before exercise; T2 = after 3 months exercise; M = male; F = female). Three dimensional PCA was employed using only autosomal DMPs (differentially methylated positions; top), or gene expression for the top 1,000 genes with largest biological variation, and the biological coefficient of variation used to produce the PCA graphs.

a) Why do the two PCA plots on top look so "wonky"? Critique this way of displaying the data in this figure. What would you have done instead? Support your answer with data. Limit your answer to a maximum of 50 words.

The top plots are rotated 180 degrees. Flip to match below and alter axes is same direction trend is to be highlighted.

b) Interpret these results for methylation comparing T1 and T2. Support your answer with data. Limit your answer to a maximum of 30 words.

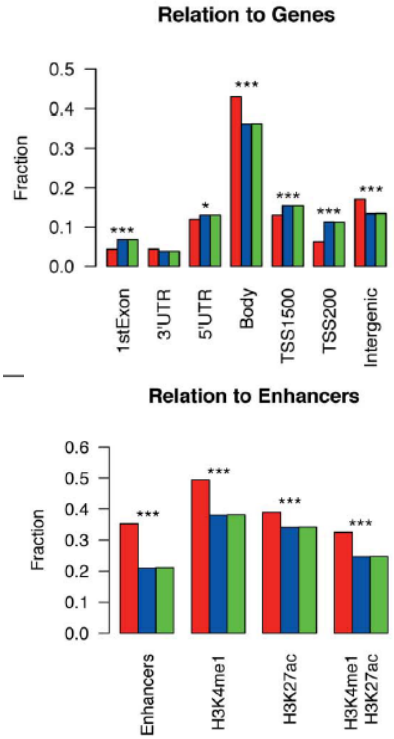
3 months of exercise altered DMP in males and females

c) Interpret these results for transcription comparing T1 and T2. Support your answer with data. Limit your answer to a maximum of 30 words.
3 months of exercise altered transcription in males and females

d) Interpret these results for methylation and transcription comparing M and F. Support your answer with data. Limit your answer to a maximum of 30 words.
Before exercise, M and F already have differences in DMP and transcription, but both sexes respond in similar ways to exercise with altered DMP and transcription.

20 pts

5) In the same study, they annotated where the DMP were located with regards to gene features (right figure, top panel). For each annotation category, the relative fraction of positions located within each feature type is calculated for DMPs (red bars), non-DMPs (blue bars) and the entire genome (green bars). TSS1500 is the 1.5 kb region upstream of the TSS. Significance codes: *p < 0.05; ***p < 0.001; Fisher's exact test. In the bottom panel, they annotated known enhancers as well as histone modifications.



a) Summarize the top panel. Support your answer with data. Limit your answer to a maximum of 30 words.
gene bodies and intergenic regions have increased DMPs, other parts have reduced DMPs in response to exercise.

b) Summarize the bottom panel. Support your answer with data. Limit your answer to a maximum of 30 words.
Enhancers and histone 3 have increased methylation, indicating this epigenetic change may have increased transcription due to exercise.

c) Integrate all the data in question 4 and 5 to produce a cohesive model of how transcription was altered in response to 3 months of exercise. Support your answer with data. Limit your answer to a maximum of 50 words.
M & F differ prior to exercise with respect to methylation of DNA (top, + enhancers) as well as epigenetic changes (bottom histones). However, both sexes respond in similar ways after exercise with changes in DMP and transcription.