Spring 2017 Genomics Exam #2
transcriptome, metagenome, epigenome

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 the first exam. You do not need to read any additional papers. 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 2:30 pm on Wednesday March 29. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE BY 2:30 pm ON WEDNESDAY MARCH 29**. You may use your notes, papers we have already read, and the internet. However, you are **not to look for source papers or abstracts** 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 2:30 pm (eastern time).

The **answers to the questions must be typed in this 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). If you capture screen images as a part of your answers, you may do so without seeking permission since your test answers will not be in the public domain. Support your answers with data from the figures in this exam.

DO NOT READ or DOWNLOAD ANY PAPERS or ABSTRACTS 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.

Name (please type):

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

How long did this exam take you to complete?

19 points

1) Here are some transcriptome data used to evaluate human and rodent datasets.

The **Figure 1.1 A2** compares data from two different datasets (Pavlopoulos aging human entorhinal cortex vs Berchtold aging human hippocampus). For both studies, the investigators performed RNA-seq analysis to find differentially expressed genes comparing younger people vs older people (don't worry how they got the brain tissue for this question) and two different brain regions. Each dot on the graph represents a single gene that was differentially expressed in both studies. The dots are placed by their $\log_2$ fold change with Berchtold data on the X-axis. **Figure 1.1 A4** compares data from two different datasets (Lockstone frontal cortex in Down syndrome patients vs Berchtold aging human hippocampus). For both studies, the investigators performed RNA-seq analysis to find differentially expressed genes comparing many younger people vs their experimental groups (don't worry how they got the brain tissue for this question) and two different brain regions. Each dot on the graph represents a single gene that was differentially expressed in both studies. **Figure 1.1 A1** compares data from two different datasets (Nagahara transgenic Alzheimer's disease model mouse hippocampus vs Berchtold aging human hippocampus). For both studies, the investigators performed RNA-seq analysis to find differentially expressed genes comparing younger people vs their experimental groups (don't worry how they got the brain tissue for this question). Each dot on the graph represents a single gene that was differentially expressed in both studies. Within all three graphs: number of genes observed (Obs), post hoc false concordance rate (FCR), percent agreement and the correlation coefficient are shown.

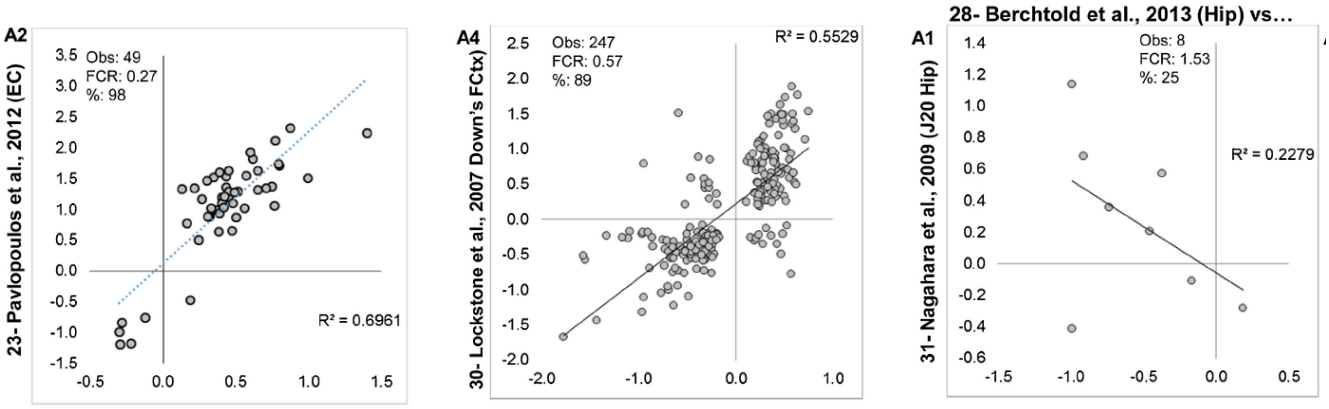


Figure 1.1 A2, A4 and A1.

a) Briefly interpret each of the three panels individually. (3 sentences maximum each)

A2: strong correlation for aging brains, two different regions

A4: medium correlation for DS and aging brains, two different regions

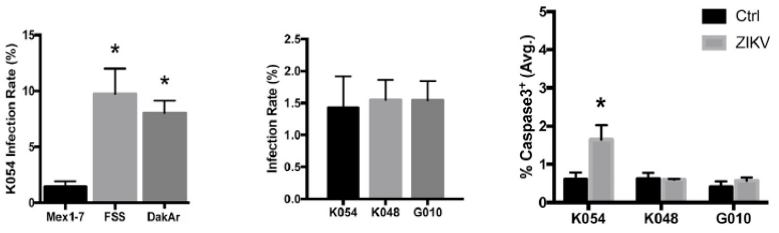
A1: negative correlation between mouse model for Alzheimer's and aging brain, both hippocampus

b) Using all three panels collectively, what conclusions can your reach? (8 sentences maximum)

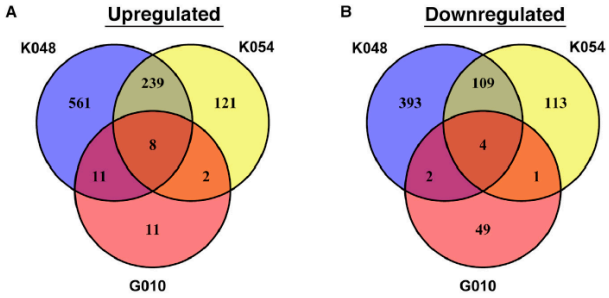
IF you want to study DS or aging, you can compare across different parts of the brain. But if you want to study Alzheimer's transcriptome, do not use this particular mouse model because they have no measurable similarities in transcriptomes.

16 points

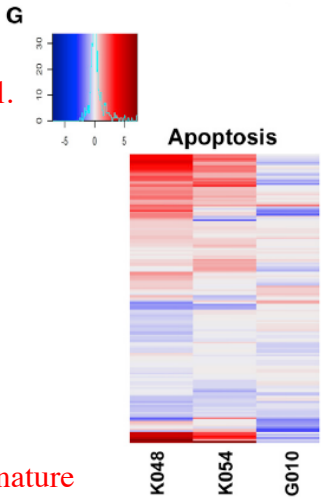
2) This examined the effect of Zika virus on fetal brains. **Figure 2.1A** (left) quantified the infection rate of K054 cells using three Zika strains, as labeled. They quantified the Mex1-7 strain of Zika infection rate of three neuronal stem cell lines (labeled) derived from three different people (right). **Figure 2.1B** shows the percentage of caspase 3 activation using Mex1-7 and the three cell lines. **Figure 2.1C** shows the RNA-seq differentially expressed genes comparing infected cells (numerator) to non-infected cells (denominator). **Figure 2.1D** shows heat map of $\log_2$ fold change (inset, x-axis) for three cell lines (columns) infect (numerator) or not infected (denominator) for apoptosis signature set genes (rows).



2.1A **2.1B**



2.1C



a) Interpret panel 2.1A by itself. (3 sentences maximum)

Different strains of Zika have different abilities to infect a particular neuronal cell. Mex1-7 is the least effective of the three viral strains to infect cell line K054. Mex1-7 infects these three neuronal cell lines equally.

b) Interpret panel 2.1B by itself. (3 sentences maximum)

Mex1-7 induces caspace-3-mediated apoptosis only in cell line K054.

c) Interpret panels 2.1C and 2.1D together. (4 sentences maximum)

Different numbers of genes are differentially regulated after Mex1-7 infection: K048 > K054 >> G010. K048 has the most genes induced from the apoptosis signature set of genes, followed by K054. G010 does not induce many genes from this set.

d) What clinical implications can you summarize from all four panels? **2.1D**
(8 sentences maximum)

Treating all Zika-infected patients the same way is inappropriate. Different genotypes of neuronal cell lines resprond with distinct transcriptomes and there appears to be two different pathways for apoptosis (caspase-3-mediated for K054 and caspace-3 independent for K048).

Furthermore, different strains of Zika are likely to affect virulence and perhaps transcriptomes too.

20 points

3) Investigators wanted to know if physically active mouse mothers produced any effects in their F1 offspring (**Figure 3.1** →). They had two populations of female mice: one with a locked exercise wheel (L) and one with unlocked wheel (U) that animals could voluntarily run on if they wanted. Preliminary tests chose only the 18 most active mice which were randomly assigned to U or L treatments. The females had 1 week to acclimate before mating with a male. Ten days after the pups were born on day zero (P0), the wheels were locked to prevent F1 generation from running on the wheel. CLAMS indicates careful metabolic measurements taken on the indicated days. QMR is a 3 week period when a subset of the females from both treatments were placed in cages with unlocked wheels. In <u>all figures</u>, * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

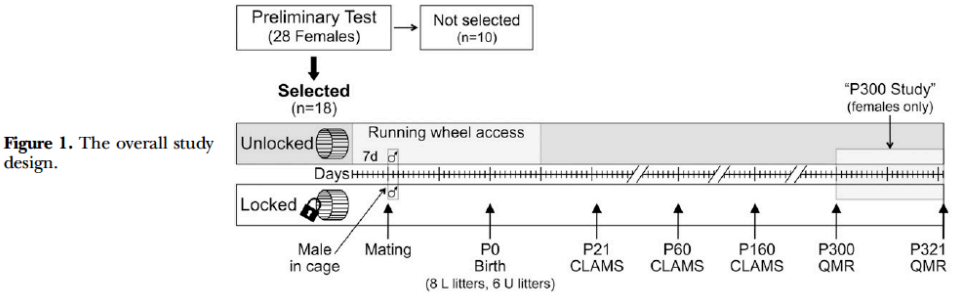


Figure 1. The overall study design.

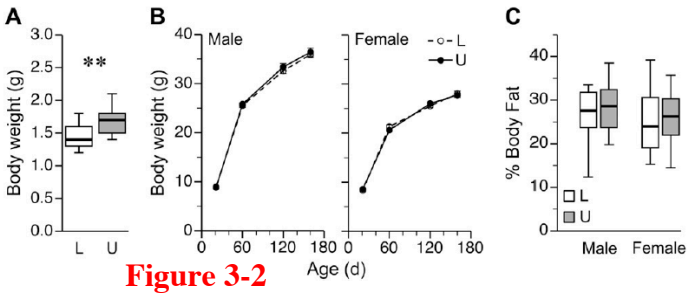


Figure 3-2

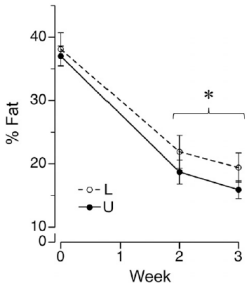


Figure 3-4

Figure 3-2 A: pups on day 1 weighed. B: growing litters weighed (averages graphed +/- stdev). C: Pups on day 160 measured.

Figure 3-4: F1 mice during QMR period**.**

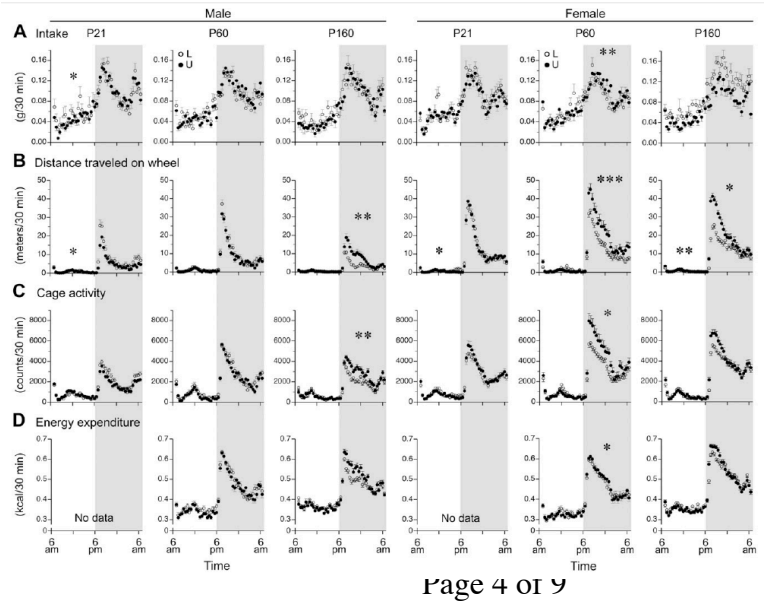


Figure 3-3: Food intake, physical activity, and energy expenditure of L and U offspring as indicated.

a) Interpret Figure 3-2 by itself. (3 sentences maximum)
Pups from mother who had access to unlocked wheel are born with more mass, but this difference goes away quickly. There is not sustained weight or fat difference between L and U pups.

b) Interpret Figure 3-3 by itself (general findings, not panel by panel details; 4 sentences maximum).
Pups from active mothers are more active, run further and burn more energy than pups from inactive mothers. Female pups seem to be more affected than male pups, but this trend is not evaluated statistically.

c) Interpret Figure 3-4 by itself. (3 sentences maximum)
Pups from active mothers lose more fat in weeks 2 & 3 than control pups even though they all have equal access to functional wheel.

d) Hypothesize a biological mechanism for all these outcomes.
Epigenomic alteration of DNA or histones.

e) Propose an experiment to determine if your hypothesis is correct or not. Use the exact same animals from this study and focus your answer on only the genomic-level of the experimental design.
Could do whole genome analysis for methylation (see question #4 below) or ChIP for histone modification.

20 points

4) A nutritionally-reduced, early-life environment can influence lifelong phenotypes in the offspring. Epigenetic factors are thought to be key mediators of these effects. A research group showed that protein restriction in mice from conception until weaning induces a linear correlation between growth restriction and DNA methylation at ribosomal DNA (rDNA). This group used reduced representation bisulfite sequencing (RRBS) to generate genome-scale, single-base resolution DNA methylomes. After mating, genetically identical mothers were randomly assigned to either protein restricted (PR) or control (C) diet treatments until the F1 pups were weaned. Only F1 males were studied.

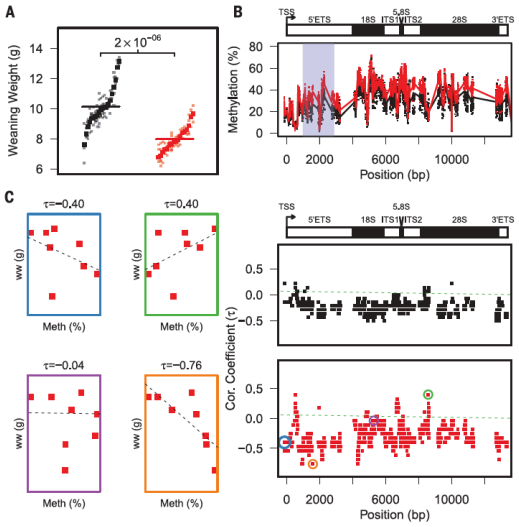


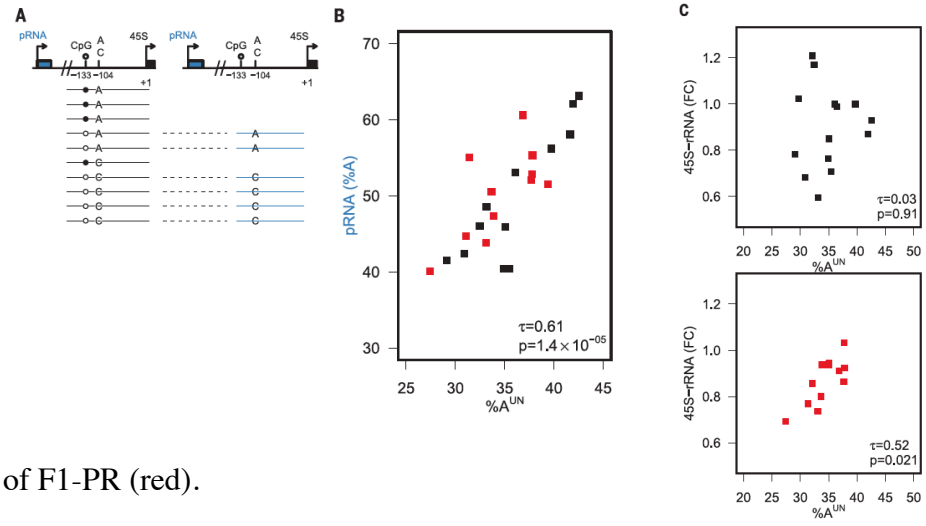
Figure 4.1. A: Weaning weight of F1-PR males (red) was compared with F1-C (black; $p = 2 \times 10^{-6}$ using

litter means). Small points represent individual mice; larger squares represent the mean of a given F1 litter. B: RRBS analysis of rDNA in F1 sperm compared to controls. The lines (upper) represents mean methylation, and points (lower 2) represent individual mice. The rDNA schematic above the graphs shows the rRNA in black and transcriptional start site (TSS); ETS and ITS are spliced off the final RNA. C: The correlation coefficient (t) between weaning weight (ww) and DNA methylation across the rDNA. Three specific bases are highlighted green, purple, orange in panel B. Base -133 is circled in blue.

a) Interpret panel A by itself. (3 sentences maximum)
Pups from PR mothers have lower birth weight at weaning.

b) Interpret panels B and C collectively. (5 sentences maximum)
Pups from PR mothers have more methylation of their DNA than control pups. The correlation between methylation and rRNA production is lower for PR pups. Panel C shows base-specific correlation with methylation: blue and orange = negative correlation; purple = no correlation; green = positive correlation.

Figure 4.2 A: The rRNA promoter (pRNA) is transcribed when rDNA is unmethylated at base -133. Therefore, the percentage of pRNA RNA-seq reads that encode an A at base –104 [transcribed pRNA (%A), indicated in blue, right] should reflect the proportion of A-variant rDNA copies that are unmethylated at base -133 (%A^{UN}). Panel A is only a schematic to explain the methodology – not real data. B: Transcribed pRNA (%A) in F1-C (black) and F1-PR (red) liver. C: Correlation of %A^{UN} with the abundance (as a fraction of control; FC) of transcribed 45S rRNA in liver of F1-C (black) and liver of F1-PR (red).



c) Interpret panel B by itself. (3 sentences maximum)
positive correlation for base -104 C and A alleles where the level of transcription increases when base -130 is unmethylated.

d) Interpret panel C in light of panel B. (5 sentences maximum)

The fold change (FC) in liver rRNA is not correlated with base -130 is methylation status for pups born from control mothers, but pups from PR mothers have a positive correlation between rRNA change in transcription and base -130 is methylation status.

e) How does nutrition, epigenomics and genotype intersect in this system?
Protein restriction during pregnancy and nursing have long-term epigenomic consequecnes for F1 progeny by altering the amount of rRNA transcribed. Methylation of base -130 is likely to reduce protein production levels in the liver of F1 mice. There is no clear connection to genotype in the data presented here.

13 points

5) A group of research dentists wanted to know the role of microbiota in periodontal disease. Samples were taken form patients and WGS was performed on microbiota total DNA extracts.

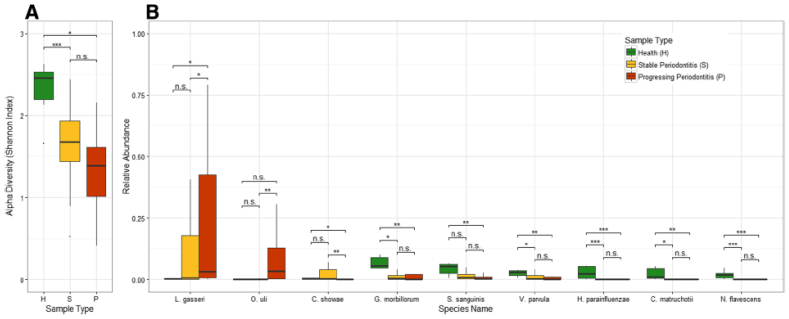


Figure 5.1: Microbial diversity and abundance difference between healthy and periodontitis samples. A: Box plot and the test results of alpha-diversity measure. B: Data for differentially abundant species. Green represents healthy samples, yellow represents stable periodontitis samples and red represents progressing samples. Statistical significance is coded as: n.s. ($p > 0.05$), *($p <= 0.05$), **($p < 0.01$), ***($p < 0.001$), and labeled above the corresponding boxes.

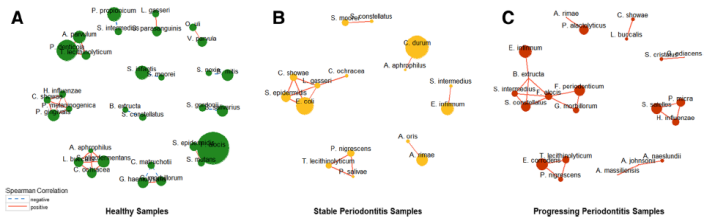


Figure 5.2: The co-occurrence correlation networks of subgingival samples under different peridodontitis states. Spearman correlations of relative abundances for all pairs of microbial species were calculated under different states of periodontitis respectively, with p-values adjusted by Benjamini-Hochberg correction, and selected those species pairs whose correlation coefficients were over 0.9 and adjusted p-values were less than 0.05 as the edges of networks. The size of spots represents the average abundance of the species in samples.

a) Interpret both figures collectively. Don't focus on all the details – just point out two major insights from these data. (3 sentences maximum each)

1) greater diversity in healthy gums
2) particular species are associated with healthy gums; different species with diseased gums

b) What possible treatment could you hypothesize to prevent periodontitis? (3 sentences maximum)
You could try antibiotics followed by microbiota transplant.

c) What is the major unanswered question from this research that will determine the effectiveness of your proposed treatment? (3 sentences maximum)
We do not know if the change in microbiota is cause or consequence of the disease. It is possible the disease comes first, followed by altered microbiota. In this case, transplanting microbiota may not have any effect.

12 points

6) A group wanted to know how effective fecal microbiota transfer (FMT) really was and they chose two strains of mice as their model system (B6J and B6Hsd).

Figure 6.1: They performed metagenomic analysis of 16S DNA before and after 5 days of a broad spectrum antibiotic (abx). Colored ovals represent 95% confidence intervals for enclosed samples.

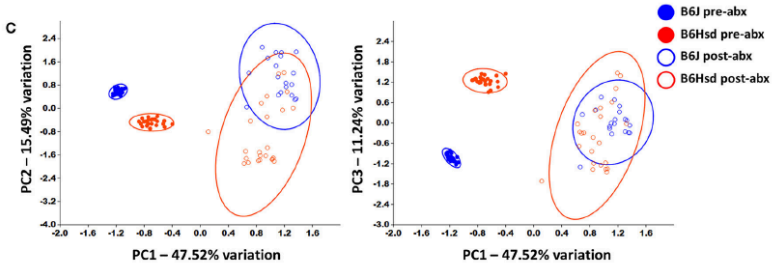
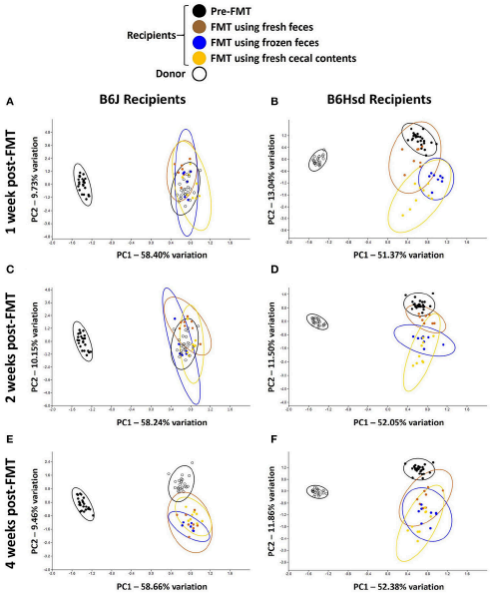


Figure 6.2: The same mice from the previous figure were given new microbiota from the opposite strain, taken before antibiotic treatments. Fecal samples 1, 2 or 3 weeks post FMT were used to prepare 16S DNA to identify operational taxonomic units (OTUs).



a) Interpret Figure 6.1 by itself. (3 sentences maximum)
Pre-abx populations are distinctive and explain most of the variation in the data. After abx, the microbiota are very similar to each other.

b) Interpret Figure 6.2 but you may include data from 6.1 if you feel it is helpful. (3 sentences maximum)
The first column shows that one strain of mice can accept and sustain the microbiota from the donor mice. The second column shows that some recipents are unable to sustain a donated microbiota and they revert back to their original microbiota. The only apparent differences in the

two columns are genotype of donor, genotype of the recipient, and the microbiome of their two microbiota.

c) What is the overall FMT lesson from this experiment? (5 sentences maximum)
Fecal transplants of microbiota may not work every time. One or more of the three variables listed above determine the outcome of FMT which means we do not understand the complex biology that determines whether FMT works or fails.