Spring 2013 Genomics Exam #2
Sequence Variations and Microarrays

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 5 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 10:30 am on Wednesday March 27. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE BY 10:30 am ON WEDNESDAY MARCH 27**. You may use a calculator, a ruler, your notes, the book, and the internet. You may take this exam in as many blocks of time as you want. Submit your electronic version before 10:30 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 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?

20 points

1) To understand what these investigators did, watch this video. Do NOT look up the paper or the abstract. It is important that you ONLY use the data in this exam to answer the questions below. If the video does not load properly via Word, copy the URL and paste it into your browser. It should work that way.

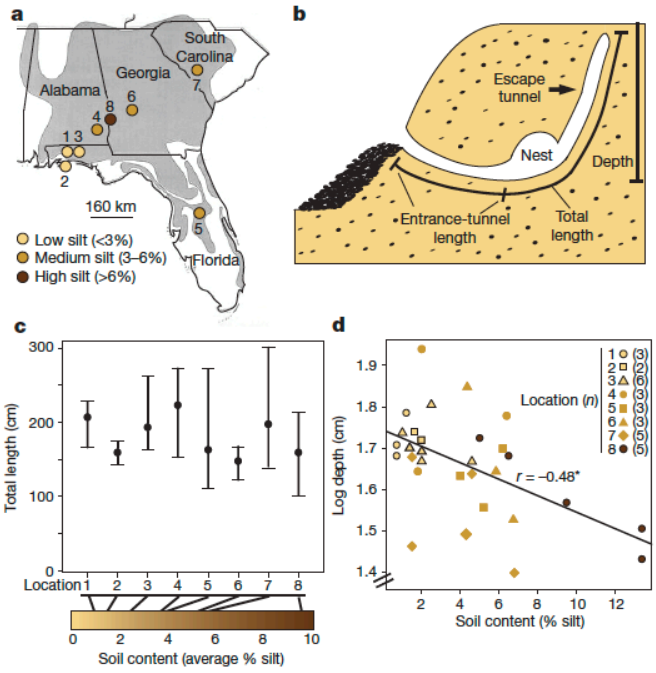


Figure 1. Natural variation in *P. polionotus* burrows. (a) Sampling of burrows at eight sites in the southeastern United States from across the range of *P. polionotus* (grey area). Average percentage of soil silt at each sampling site is provided. (b) Diagram of a typical *P. polionotus* burrow showing the measures for entrance-tunnel length, total length and burrow depth, as well as a typical escape tunnel. (c) Variation in total burrow length among sites (mean ± range), which are ordered by increasing percentage of silt (left to right). (d) Correlation between silt composition of soil and burrow depth (asterisk indicates Spearman correlation). Each point represents a burrow, and shapes represent the eight different sampling sites. The number of burrows measured at each site is shown in parentheses.

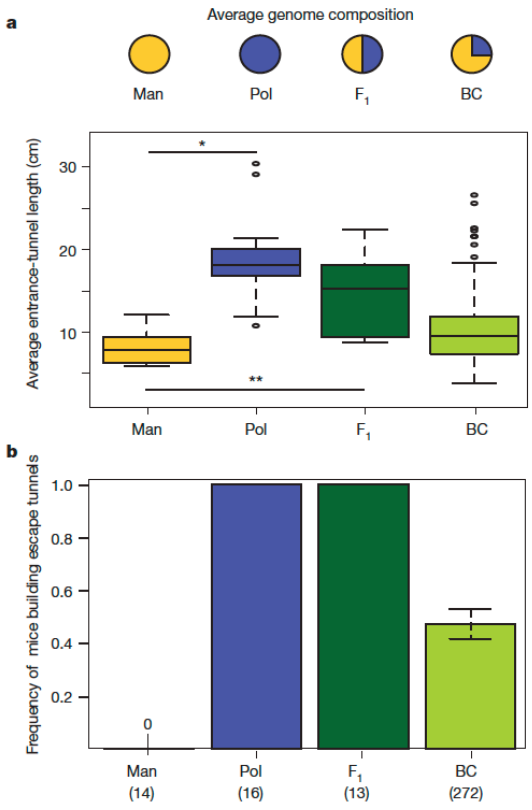
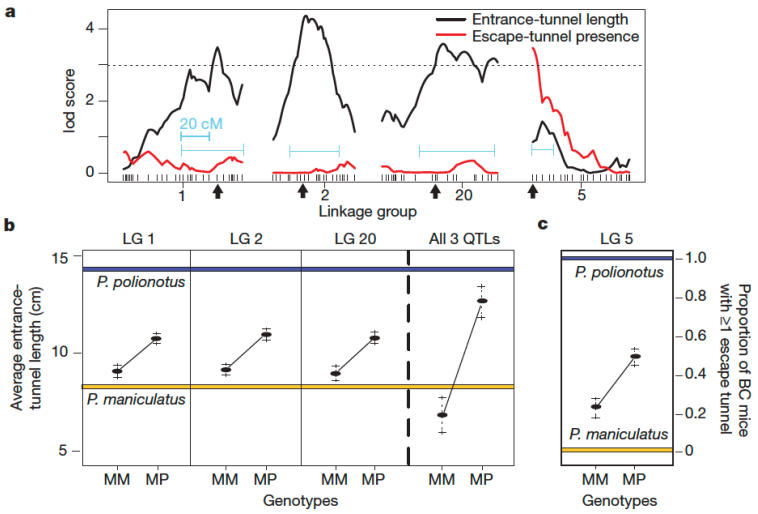


Figure 2. Burrow variation across generations. (a) Burrow dimensions of *P. maniculatus* (Man; yellow), *P. polionotus* (Pol; blue), F_1 hybrids (dark green) and progeny resulting from F_1 X *P. maniculatus* backcross (BC; light green). Pie charts depict average genome composition in each generation. Distributions of entrance-tunnel length in the parental species, F_1 hybrids and BC animals are shown (average of three trials for each mouse tested). Boxes represent interquartile ranges (median ± s.d.). Significant t-tests,*P = $5x10^{-3}$, **P = $2x10^{-4}$. (b) The frequency of escape-tunnel construction is shown for the same individuals. Error bars represent mean ± standard error of the mean (s.e.m.) Sample sizes are listed in parentheses below.

Figure 3. Genomic analysis of burrow variation.
(a) Linkage groups (LGs) 1, 2 and 20 with log-transformed average entrance-tunnel length (black line). Linkage group 5 with escape-tunnel presence (red line). Dotted line represents log odds ratio (lod) significance threshold (genome-wide $\alpha = .05$, lod ~ 3.0). 1.5-lod confidence intervals and scale in centimorgans (cM) are shown in light blue. Dashes indicated genetic markers, and black arrows indicate markers used to define each peak (used in b).
(b) Phenotypic effect of individual and combined genotypes (linkage groups 1, 2 and 20) on entrance-tunnel length in 272 BC mice.
(c) Proportion of BC animals that construct escape tunnels for each of the two genotypes. All error bars represent mean ± s.e.m. Blue and yellow lines represent average phenotype of the parents (pure species) used to found the cross. Genotypes are either homozygous *P. maniculatus* (MM) or heterozygous *P. maniculatus/polionotus* (MP).



Questions:

a) Please write an abstract limited to 200 words maximum. Your goal is to summarize the research in Figures 1 – 3.

b) Choose a model genome that is closely related to the species in this study. Use this model genome to design a method to identify the alleles responsible for the behaviors in the original species once the two genomes from these species in the paper are fully sequenced.

c) I want you to evaluate the model species you chose. How close to the target species is your model species? Use genome-scale evidence to support your answer, not individual genes.

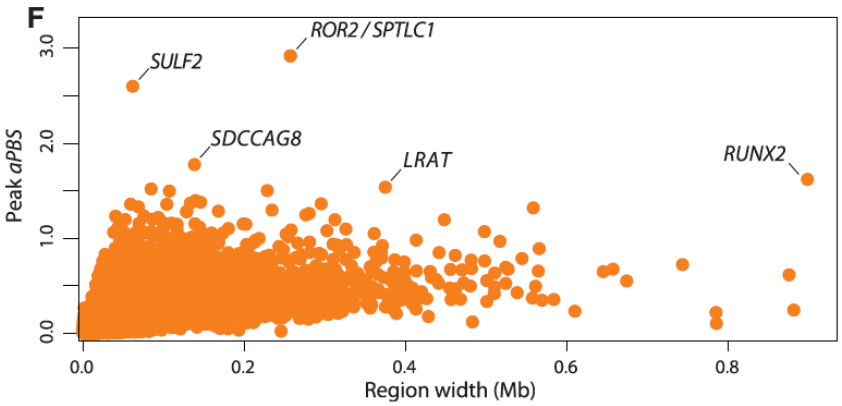
20 points

2) Number two contains a series of unrelated but interesting tidbits that I'd like you to ponder. Don't look for these papers or their abstracts. Use sequence databases to answer these questions.

a) "Strikingly, we find that human Leptin can rescue the *upd2* mutant phenotypes, suggesting that fly Upd2 is the functional ortholog of Leptin." How similar are these two proteins? Support your answer with data. Limit your answer to a maximum of 2 sentences plus screen shot(s).

b) Find the SNP responsible for "immigration delay disease." You have to show me the wt DNA sequence to get full credit. You also have to explain why this SNP causes a change in phenotype.

c) On the next page is a figure taken from a paper published in 2012. Do not look for this paper, but instead, I want to see if you can assemble a reasonable hypothesis based on the data. The



context was a massive study of early human evolution. The investigators sampled hundreds of Africans with a special emphasis on tribes based in the southern half of Africa. They used a statistical measure (aPBS) which you don't need to understand. The aPBS measure was designed to identify regions of recent selection (≤100,000 years ago) that may have distinguished modern *Homo sapiens* from other hominids. Investigate information about the named points in the figure and speculate what anatomical features were being selected for that lead to the modern human physique. I want you to focus on no more than two general traits in your speculation and summarize your ideas in a maximum of 5 sentences. Be sure to use all the named points when formulating your answer.

30 points

3) These questions focus on breast cancer sub-types. Please do not look for the *Nature* paper or any commentaries on it (other than what we already covered in class – you can go back to those readings but I doubt they will prove insightful). **Look at the attached Figure4.pdf file.** ER = estrogen receptor. PR = progesterone receptor.

a) Summarize what you notice about PIK3CA in breast cancers. Support your answer with data. Limit your answer to no more than two sentences.

b) Which cancer type(s) look the most homogeneous and which one(s) look the most heterogeneous? What are the clinical implications of your interpretation? Support your answer with data. Limit your answer to no more than four sentences total.

c) Comment on the mutations per Mb for each breast cancer subtype. Support your answer with data. Limit your answer to no more than two sentences.

Look at Figure5.pdf for the remaining questions.

d) Summarize the significance of panel a. Support your answer with data. Limit your answer to no more than two sentences.

e) It might seem odd to have breast cancers on the Y-axis and breast cancer subtypes also on the X-axis. What can you conclude from just this portion of the data in panel b? NL = normal-like breast cancers. Support your answer with data. Limit your answer to no more than two sentences.

f) What is the utility for comparing different breast cancer types against other non-breast cancers? Cite specific examples from panel b to illustrate our answer. Limit your answer to no more than two sentences.

30 points

4) The final question is based on another paper but I don't want you to look for this one either. The educational goal is for you to demonstrate your ability to interpret the information I provide on this exam. You can look up genes or terms online, but please don't go looking for this paper, its abstract, or any commentaries on this paper.

ABSTRACT

After traumatic injury, peripheral nerves can spontaneously regenerate through highly sophisticated and dynamic processes that are regulated by multiple cellular elements and molecular factors. Despite evidence of morphological changes and of expression changes of a few regulatory genes, global knowledge of gene expression changes and related biological processes during peripheral nerve injury and regeneration is still lacking. Here we profile global mRNA expression changes in proximal nerve segments of adult rats after sciatic nerve transection. According to DNA microarray analysis, a huge number of genes was differentially expressed at different time points (0.5 h–14 d) post nerve transection, exhibiting multiple distinct temporal expression patterns. The expression changes of several genes were further validated by quantitative qPCR analysis. The gene ontology enrichment analysis was performed to decipher the biological processes involving the differentially expressed genes. Collectively, our results highlight the dynamic change of the important biological processes and the time-dependent expression of key regulatory genes after peripheral nerve injury. Hopefully, this study may provide a useful platform for deeply studying peripheral nerve injury and regeneration from a molecular-level perspective.

Open the file Question6A.pdf

a) Produce a numbered list the cluster of genes from fastest (#1) to slowest (#last) to response to the transection. You may preface your list with no more than one sentence.

b) Can you find a gene that appears to return to its basal state at day 14? Explain your answer in no more than two sentences.

c) What do you think of the "time point 0"? What would you want to know about the experimental design if you knew these were two-color microarray data? Support your answer with data. Limit your answer to no more than two sentences.

d) Design a research program that uses these data to develop a treatment to stimulate human nerve regeneration in the case of accidents. Support your answer with data. Limit your answer to no more than four sentences.

Open the file Question6B.pdf from the same paper. RT = reverse transcriptase.

e) Summarize the results from this figure. Support your answer with data. Limit your answer to no more than two sentences.

f) Speculate what the therapeutic implications might be for these data, assuming these proteins maintain their traditional molecular function. Limit your answer to no more than two sentences.