

**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.

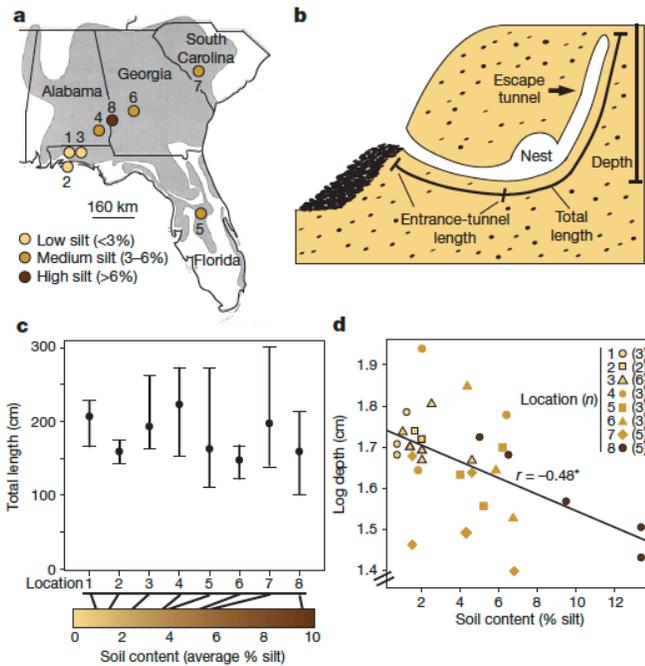


Figure 1. Natural variation in *P. polionotus* burrows. (a) Sampling of burrows at eight sites in the southeastern United States from across the range of *P. polionotus* (grey area). Average percentage of soil silt at each sampling site is provided. (b) Diagram of a typical *P. polionotus* burrow showing 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  $\pm$  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<sub>1</sub> hybrids (dark green) and progeny resulting from F<sub>1</sub> 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<sub>1</sub> hybrids and BC animals are shown (average of three trials for each mouse tested). Boxes represent interquartile ranges (median  $\pm$  s.d.). Significant t-tests, \*P = 5x10<sup>-3</sup>, \*\*P = 2x10<sup>-4</sup>. (b) The frequency of escape-tunnel construction is shown for the same individuals. Error bars represent mean  $\pm$  standard error of the mean (s.e.m.). Sample sizes are listed in parentheses below.

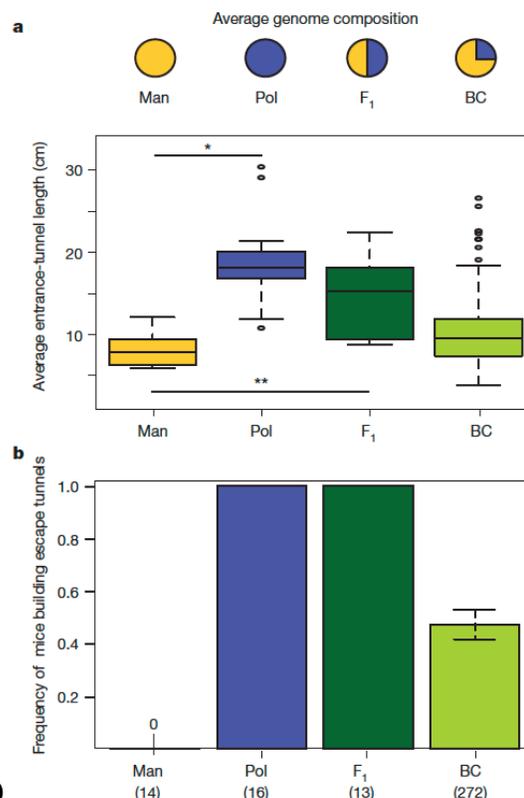
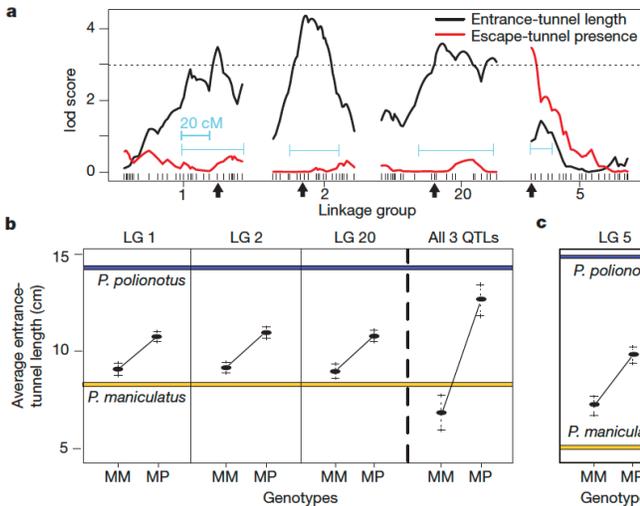


Figure 3. Genomic analysis of burrow variation.

(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  $\alpha = .05$ ,  $\text{lod} \sim 3.0$ ). 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  $\pm$  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:**

**10 pts**

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

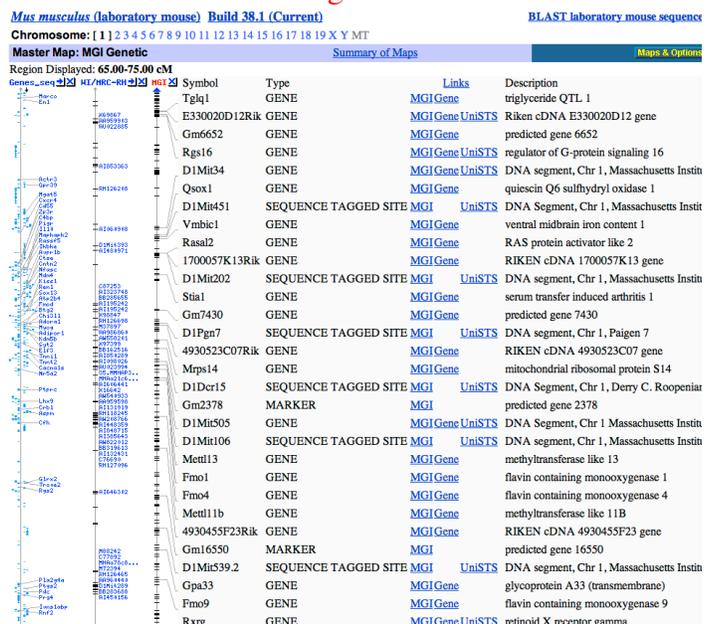
Various answers. Needed to discuss 3 QTLs for tunnel length and 1QTL for escape tunnel. Cannot distinguish if one gene or haplotypes of alleles at each QTL.

Escape tunnel locus looks dominant except the error bars don't seem right. The others are not so clear but are polygenic for sure. There must be another locus for escape tunnel too, but they did not find it.

**5 pts.**

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.

Mouse and look for syntenic regions and identify locus near markers shown in Figure 3A. (see right)



5 pts.

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.

*Mus musculus* has 19 autosomes but the species in the paper have *P. maniculatus* has 23 autosomes which means they cannot have completely syntenic genomes. Therefore mapping QTLs between these two species may be difficult.

[http://www.ncbi.nlm.nih.gov/genome/?term=txid10042\[Organism:exp\]](http://www.ncbi.nlm.nih.gov/genome/?term=txid10042[Organism:exp])

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).

Fly: NP\_573277

Human leptin: AAH69452

Normal BLAST2 alignment gave me no similarity.

The screenshot shows a BLAST2 search interface. The query is NP\_573277 (370 letters) and the subject is AAH69452 (167 letters). The description for the subject is: leptin precursor [Homo sapiens] >gi|730218|sp|P41159.1|LEP\_HUMAN RecName: Full=Leptin; AltName: Full=Obese protein; AltName: Full=Obesity factor; Flags: Precursor >gi|623332|gb|AA60470.1| ob [Homo sapiens] >gi|1226244|gb|AAC50400.1| obese protein [Homo sapiens] >gi|1407583|dbj|BAA09839.1| ob protein [Homo sapiens] >gi|1469860|dbj|BAA09787.1| leptin [Homo sapiens] >gi|38174530|gb|AAH60830.1| Leptin [Homo sapiens] >gi|46854318|gb|AAH69527.1| Leptin [Homo sapiens] >gi|46854679|gb|AAH69452.1| Leptin [Homo sapiens] >gi|51095072|gb|EAL24315.1| leptin (obesity homolog, mouse) [Homo sapiens] >gi|62361689|gb|AAH81413.1| leptin (obesity homolog, mouse) [Homo sapiens] >gi|119604046|gb|EAW83640.1| leptin (obesity homolog, mouse) [Homo sapiens] >gi|208968561|dbj|BAG74119.1| leptin [synthetic construct] Molecule type: amino acid Subject Length: 167 Program: BLASTP 2.2.28+ Citation. A message at the bottom states: "No significant similarity found. For reasons why, click here".

CLUSTALW alignment:

```

CLUSTAL 2.1 multiple sequence alignment

Human_Leptin
Fly_Upd2
MILSVVMPFTKARHLRDLQVALDYAYDEDNSGRSSSSSSSSSSSGDILPAFDVNPYRGLA 60

Human_Leptin
Fly_Upd2
DSFGEGSYDSDMSLSESSYEEIAQAAIRAARRRMRQRTRHARSHNLRFLSFKSEIPEWE 120
                                     *

Human_Leptin
Fly_Upd2
TLCGLFWLWPLYFYVQAVPIQKVQD--DTKTLIKTIVTRINDISHTQSVSS--KQKVTG 59
NPCGGTYTPDESLSDSASQGRVVKRKHLENLNRNITMSEYQDITHRATLEYGNLQHWQRE 180
. ** :      : .:. * . . . * : :. : ** * : :. :

Human_Leptin
Fly_Upd2
LDFIPGLHPILTLRMDQTLAVYQQILTMPSRN-----VIQISNDL 101
YKFLPNMTRPTSAVKLKTWYRHMQTFFVGSFYLGRAQYKFRKDQQKSLNEAVTKELHDDL 240
. * : : : : * : . * : * : . . . : : : *

Human_Leptin
Fly_Upd2
ENLRDLLHVLAIFS KSCHLPWASG-----LETLDLSLGGVLEASG----- 139
VSARYMLCEIELTINASYPNSNGAKLSRVSRQAMEERLNFHTPANGSMEDQRDLKVSKE 300
. * * : : : . . * : * . * : . . * : * :

Human_Leptin
Fly_Upd2
-----YSTEVVALSRLQGS-----LQDMLWQLDLSPGC--- 167
LYFYQLNNVVKTLHRVLRPRRNSAERRHLAAIHGTGASGLRAGSSEMLDVGSNSNGSGGS 360
* * * * : : : : . . : : : * *

Human_Leptin
Fly_Upd2
SGSDGGSNES 370
    
```

What is really interesting is that these two proteins have very little sequence conservation and yet functionally, the human leptin protein can functionally rescue *udp2*<sup>-/-</sup> flies. There are some conserved amino acids so perhaps this degree of conservation is sufficient to replace the missing Udp2 protein so similarity must be qualified by function or primary protein structure.

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. Autosomal dominant allele.

From OMIM:

.0001 ADERMATOGLYPHIA  
SMARCAD1, IVS1, G-T, +1

In affected members of a 4-generation Swiss kindred with adermatoglyphia (136000), originally studied by Burger et al. (2011), Noursbeck et al. (2011) identified heterozygosity for a G-T transversion in intron 1 of the SMARCAD1 gene (378+1G-T), predicted to abolish the donor splice site adjacent to the 3-prime end of the first unique exon of the skin-specific short SMARCAD1 isoform. The mutation segregated with disease in the family and was not found in 100 Swiss controls or in 100 Jewish controls. Transfection into HeLa cells and into primary human fibroblasts followed by RT-PCR revealed that the mutant yields 2 aberrant splice variants, both of which were only marginally detectable compared to wildtype.

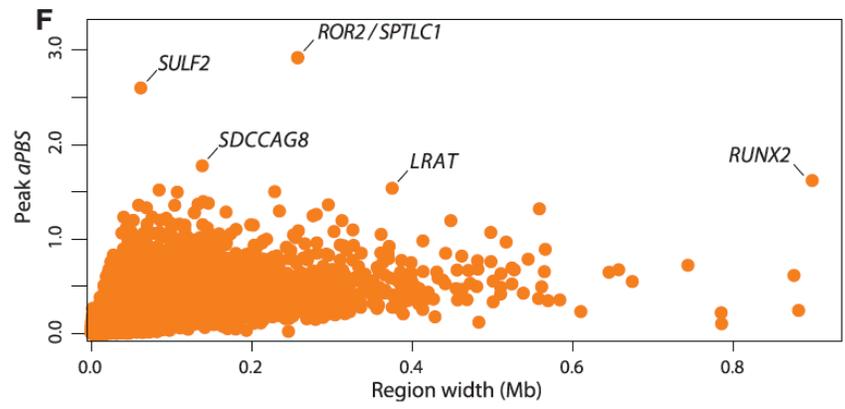
Experiments with cycloheximide showed that the aberrantly spliced variants of the SMARCAD1 short isoform undergo degradation.

>gi|363807327|ref|NM\_001128429.2| Homo sapiens SWI/SNF-related, matrix-associated actin-dependent regulator of chromatin, subfamily a, containing DEAD/H box 1 (SMARCAD1), transcript variant 1, mRNA

```
CTCTCAGCTGGGATCGCGCCGCTCAACTTCCGGGCGGATGCCCGCCAGCAGGCCTCCGCCGCTCCCT
TCTTTGGCCCTTTGTGTCCCCGAGTGTGAGGCGCGGGCCCTGGCAGGTCCTTTCTCGAGGCAGGGGG
CACGGTAGCACAGGGAGCTTCTTTTGGGGCGGCGGAGGCCGCTAGGGGTTTACTGGGAACTG
GCCTGCGCGTGTCTGGATAGTTCAATTTAAAGCCCCATCCCTGCAAGGTGGTCTTTCTACCAATATGAA
TCTTTTCAACCTGGACGCTTTTCGCTTTGAGAAAAGGAATAAGATTGAGGAAGCGCCGAAGCAACCCT
CAACCTTCCCAGCCTGGCCCTTCTTACCAATTTCTCTTAGTGTCTGAAGAGGAGAATGCTGAAGGGGAA
TTAGCAGGGCAAACACTCTGATTCAGATATACTGAAAAAACAGAAGATTCTAGTGTCCAGAACTCC
AGATAATGAAAGAAAAGCAAGTATATCATATTTCAAAAATCAAAGAGGAATACAGTATATTGATTTGTCT
TCTGATAGTGAAGATGTCGTTTCCCAAATTGCTCCAATACAGTTCAGAGAAAACATTCACAAAGATA
CAGTGATTATAGTTTCTGAGCCATCTGAAGATGAAGAGTCCCAAGGCCTTCTTACCATGGCACGTAGAAA
TGATGATATTTCCAGAACTGGAAAGACTTTCGGAATTGGAAGACCTTAAAGATGCTAAACTTCAGACTTTG
AAGAACTTTTCCCAAAGAAGTGACAATGATTTACTTAAAGTTGATTGAATCAACAAGCACATATGGATG
GAGCAATTGCTGCTGCCCTGCTGATGTTTGGTGTATGCAGGTGGTGGGCCCAGGAAAAGAAAATATCTTC
TTCTTCAGAGCCATATGAGGAAGATGAATTTAATGATGATCAATCTATAAAAAAGACAAGACTGGATCAT
GGAGAGGAATCAAATGAGTCTGCAGAACTAGCAGTAATTGGGAAAAGCAGGAAAGTATTGTACTGAAAT
TGCAAAAGGAATTTCCAATTTTGATAAACAGGAACTTAGAGAAGTACTCAAGGAACATGAATGGATGTA
CACAGAAGCTTTAGAATCTCTAAAAGTGTTCGAGAAGACCAAGATATGCAATATGTATCACAAAGTGAG
GTTCCAAATGAAAAGAAGTTTCTTCAAGAAGTCAAAAATACCTAAAAATGCAACTAAAACAAAACATA
AACAGAAATTTTCAATGAAAGCACAAAATGGCTTTAACAAGAAACGTAATAAAAAATGTTTTAATCCAAA
GAGAGTTGTTGAAGACTCTGAATATGATTCAGGTTCTGATGTCGGTAGTTCACTAGATGAGGACTATAGT
AGTGGTGAAGAAGTATGAGGAGTGGCTATAAAGGTAAAATTTCTCACTTCTTCAAGATGCTTCAATTG
GTGAACCTACTTTGATCTCTCAGTGTCTCAGAAAAGGCTCAGAAGATAACAGAACCTCCGGCCCTTAA
TAGTTGGGAGGCTCTGTTCCAAAGATGTCAAAACCTAATGGCTTATCAGAAGATTTGATATGGCACTGT
AAAACACTGATCCAGAAAGAGATGTAGTTATAAGGCTTATGAACAAATGTGAAGACATTTCAAATAAAT
TGACCAACAAGTTACATGCTTACTGGAAATGGAGGTGGATGGAACATAGAACAACCTTCCATTCTAAA
CCAAAGTTTGTCACTCAAGCCCTATCAGAAGTTGGTTTGAATTTGGCTGGCATTTGGTACATAAACATGGA
CTTAATGGCATTTTGGCAGATGAAATGGGCCTAGGAAAACTATTCAAGCCATTGCATTTCTGGCATAAC
TCTATCAGGAGGTAATAATGGTCCCTCATTGATCGTTGTTCCAGCTTCAACTATAGATAACTGGTTAAG
GGAAGTTAATTTATGGTCCCTACTTTGAAGGTCCTCTGTTACTATGGTCTCAAGAAGAACGTAACAA
ATTAGATTAAACATTTAGCCCTATCAGAAGTTGATGAAGATTAACATGTAATTTGACCACATATAACTGTGCGATCA
GCAGTTCTGATGACCGTAGTCTGTTTCGACGGCTGAACTTAATTACGCAATTTTGTATGAGGGCCATAT
GCTGAAGAAATATGGGCTCCATTGCTACAGCACCTTATGACAATTAATGCAATAAACCGTTGCTGCTC
ACAGGCACACCTGTACAGAACAATCTGTTAGAATCATGTGCTGTTGAATTTGTTATGCCACACATGT
```

TTAGTAGTAGCACCAGTGAAATCGAAGAATGTTTTCTCTAAGACAAAATCAGCAGATGAGCAAAGCAT  
 ATATGAAAAGGAGAGAATAGCACATGCAAAACAAATATAAAGCCATTATTCTCAGAAGAGTAAAAGAA  
 GAGGTTCTCAAGCAGTTACCCCCCAAGAAAGATCGAATTGAGTTGTTGCAATGTCCGAGAAGCAGGAGC  
 AACTCTATTTGGGCTTTTCAACAGATTGAAAAATCTATCAATAACTTGGTCACAGAAAAAACACAGA  
 AATGTGCAATGTCATGATGCAGTTGAGGAAAATGGCCAAATCATCCTTTATTACATCGCCAATATTACACA  
 GCTGAAAACTCAAGGAAATGCTCAGCTTATGCTAAAGGAACCTACACATTTGTGAGGCTAACCTGACC  
 TGATCTTTGAAGATATGGAAGTTATGACAGACTTCGAACTACATGTACTTTGTAAACAGTACCGACACAT  
 TAATAACTTTCAGTTAGACATGGACTTGATTTTAGATTCTGGAAAATTTGAGTTTATAGGATGCATCTTG  
 TCTGAATTGAAACAGAAGGGTGTAGAGTTGTGTTATTAGCCAAATTTACCATGATGTGGATATCTTAG  
 AGGTTCTATTAACATCATCAGCATAGGTACCTCAGATTAGATGGAAGACTCAGATTTCTGAAAGGAT  
 TCATCTAATTGATGAGTTTAAATACCGATATGGATATCTTTGTGTTTCTGCTATCAACAAAAGCTGGTGA  
 TTAGGAATAAATCTGACTTCAGCAAATGTTGTTATACCTCAGCATATTGACTGTAATCCTTATAATGACA  
 AACAAAGCAGAAGATAGATGCCATAGAGTAGGCCAGACTAAAGAAGTACTAGTTATAAAAACATAAAGCCA  
 AGGGACGATTGAAGAATGCAATGCTAAAAATTAACCAACAGAAATGAAACTAGAACAGGATATGACTACA  
 GTAGATGAAGGTGATGAAGGGATATGCCAGCAGATATAGCCACATTTACTAAAAACATCAATGGGCCGT  
 GAAATAAGAACTGTGAATCTCAATTGATGAGGAAATATCAACTTGGTGCACCAAGGACATTTACATTA  
 TGATGACCATGGGGTTTATGAACATTTATAACTTTTTATAATTTCCATATTACATTTCTCATAGTATGGA  
 CAACTTTTGCCTAATGTGATGTTTCCAAATCTCACAAGTAAATTTCAAAAAGAAAGCCAAAATAT  
 GTAGTTCTGAAGATGTTGAATAATCATTTTACAAGCAGTTTTCTGAATGGGGATTAGTTGGTATTGTT  
 TGTAAACAAATATGCTAATGCTTTAGAAATGTCAGTATTTTGTAAATTTTCTACCTCCAAATATATATA  
 TATTGTCTTCTACTGGATAATGTGTGATGATTTTACATGTGCCTTATTTGACAATGCTTATGTCTTGT  
 TTTGCTTGTCTCATTGAGTTCTTTTTATTATGTTAAAGAATGCAGCTGTATAGATTATATAGCTTTC  
 ATTTTATGCTATTTGAAGCAGATGTTCCAAATGTCAGCAAGAACTCAACCTGAATTTAAAGGTGGCAT  
 TCCATATACTAACATCCCCAGGTCTCTCAAGTACTCTGCTGAAACAAATTTATTTGGCTAGGCACATA  
 AGTTGTTTTCCAGTGAATAGTAACTAAAGAAGCCCTACCTTGCTCCATGGATTAATTCCTTCTGTTCAT  
 TTTCCAACTGCCTAATGTGCAATTTACTCTGCTAATCTTGTGCAATTTTTCATTGATTTCCCTCTCC  
 CGGCTTTTGTCTTCTTGAACCTGTTGCCAGTCACTTCTGCTCAATTTCTTCTCTCTAAATAGTAG  
 TTTATTACTGCCACATCTCCATGCATCAGCAAAATGTTGGTGCATTTTCTAGCCTGGCAGAACAGATT  
 ACTTAAAGCTATTTCAATTTCAAGCAGACTGAATGTGACTTCATCTAAAGGCAGCATTAGGTACTGCATG  
 GAAATAGGTCATTAACTTGAAACTCTTATCAAAATATATTTTACCAGTTTCCAGAATTTCCAGTACAGGA  
 CCGCTGAAGAGAGGCCATTGTTCAATTTCAATTCAGTGTGAGTGACAAAGTGAATTTAGAAGTGAAG  
 TTGCTATTTGATATTTAACTCTTTATAAATCTTCTTTAAATTTCTGCCTGTCAGTCTATATTGCTGT  
 TTTTATTATACATCAGTTCTTTGTATAAECTTGTGAGTTCATGTTGTTTATTTATATGTAATATC  
 ATTATAAATAAACTTATTATAAATCAAAGATTTGTTAATTTTGGAAATCATGCTTTTCAAAGCATCT  
 AACTTGTCTAAGATGCTAGGTAGTACGCCCTCTGGATTGGAAGGCAATAAACTCTTACAGTGATTAT  
 TTAGATATTAAGACTGAGAACTCAGGCTTAACCCAGTCTTGATGGTATATTGAACAGACTGAATATA  
 TTTTACCATTACAGGGCTAAAAGGAGTCTCATGTGTTAATACTACCCTCTTGTACATCACATACCCAA  
 ATCAGTTATTCAAATTTAGGAATTTTACCTTTTAAATCTCATAATGGATATCTCATGTTTCTGTAT  
 TAATGTATTTTCAATGATAGGCTGTTTCTTTTTTGTGTTATTTGTTGTTGTTATATCCATACCTTT  
 ATCTCTAATGAAATGTAGTTGGTCTTCTCTGTAATGCGCTATTATGCTTGGGCTTAATAAAAAATATT  
 GTGATCATAAAAAAAAAAAAAAAAAA

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 ( $\leq 100,000$  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.

### Bone, skull and vision

### 30 points (5 pts each)

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.

missense in almost half except basal-like

CNV amplified in basal for at least half

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.

Luminal A and B have many diverse mutations and CNVs

HER2-enriched looks intermediate with high degree of TP53 and fewer other mutations.

Basal-like looks the most homogeneous (TP53 and few others).

Clinical implications are the two luminals may not respond the same way to treatment while Basal-like might perform more uniformly.

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.

HER2-enriched has the highest mutation rate overall and more samples that exceed the indicated threshold at 50%. Basal-like is second with fewer extremely mutated samples followed by luminal B (and a few extremely mutated samples) with luminal A samples having the fewest mutations overall, though a few samples reach the 50% mark.

**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.

With the exception of BRCA2, ovarian and basal-like breast cancers have many similar genes differentially expressed (induced red and repressed blue). This might lead to common treatment for these two cancer types.

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. **You can see that basal-like breast cancers are the most similar to each other (lots of yellows) and the opposite of the two luminal cancer types (intense blues). HER2 and luminal B are the next most internally similar but luminal A is diverse even when compared to other luminal A samples (see inconsistent yellows)**

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.

**Normal-like breast cancers are more like kidney cancer, basal is similar to lung and ovarian while luminal A is most like prostate cancer which is unexpected. This could lead to trying already approved drugs for one cancer on the other similar cancer.**

**30 points (5 pts each)**

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.

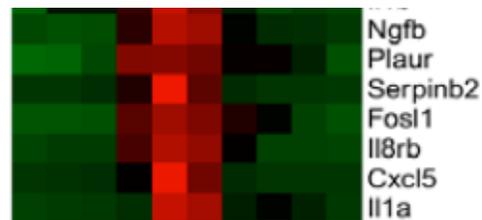
**It was hard to determine the order, so I was pretty flexible in grading. I moved halfway down the list of genes and determined what time point was intense red. This gave me a fairly objective**

way to determine order. The further to the right induction was at midway down the list of genes, the faster that response had happened. The slope of induction was determined by the number of early genes which not a good measure, but it would be easy to get fooled this way by the visual perception of rapid induction. You cannot use the middle column because that is always the same time point. However, you could use the middle column and ask how far down the list of genes are activated by the mid time point.

1. immune response & axon guidance & axon regeneration (4 days)
2. proliferation & apoptosis (1 day)
3. inflammation & migration (9 hr)
4. response to stimulus (4hr)
5. myelination (1 hr)

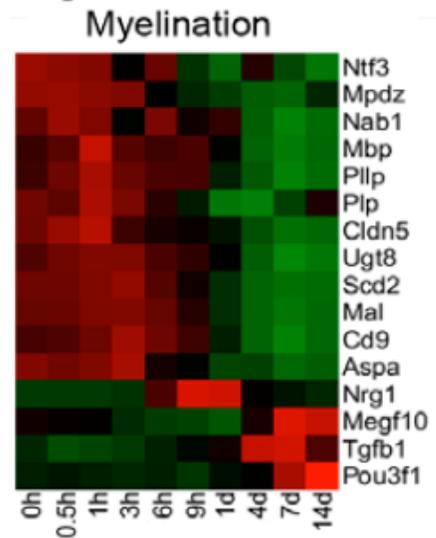
b) Can you find a gene that appears to return to its basal state at day 14? Explain your answer in no more than two sentences.

Many of the genes in the middle of pathways start repressed and return to repressed since their time has come and gone. See screenshot of several examples, though there are many more in several of the clusters.



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.

It seems impossible to sample the neuron without damaging it so time point zero is really time point close to zero but not instantaneously zero or prior to tissue damage. You can see several myelination genes already activated at time point zero which suggests to me that they had already begun to respond to the cut neurons. However, if they used a pooled source of control mRNA, then some genes could be up or down regulated compared to a generic pool of mRNA.



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.

The inflammatory and immune responses may not be involved in regeneration, so I would focus my attention on the next three classes of genes: regeneration, proliferation and migration. I would look for compounds or physical stimuli that could induce these three gene clusters in hopes that I could stimulate nerve regeneration anytime they were damaged.

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.

The first column is a positive control so we can be certain all four samples contained good quality mRNA and therefore blank lanes for the Olf genes indicates no detectable transcription. The sciatic nerve and Schwann cells both transcribe these 10 olfactory receptor genes while the dorsal root ganglion and fibroblasts do not (at least to the detectable limits of RT-PCR).

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.

If these Olf genes are translated, then we would expect them to be located in the plasma membranes of neuron and Schwann cells (which form myelin covering of neurons). Therefore it might be possible to determine what odorants bind to these receptors and use them as potential therapeutic agents to stimulate regeneration or repair.