Spring 2016 Genomics Exam #3
Proteomics, Metabolomics & Systems Biology

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 about the same amount of time as the first three exams this year. You do not need to read any additional papers. This exam consists of 6 questions. You are not allowed discuss the test with anyone until all exams are turned in no later than 9 am on Monday May 9. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE BY 9 AM ON Monday MAY 9.** You may use your notes, papers we read, and the internet but not another person. However, you are not to look for the source paper from which the figures were taken. You may print this test to work on your drafted answers, but make sure to dispose of your scrap paper so that no one will find it. You may take this exam in as many blocks of time as you want.

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 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?

15 points

1) We have learned a lot about proteomics. Here are a couple questions to explore the area.

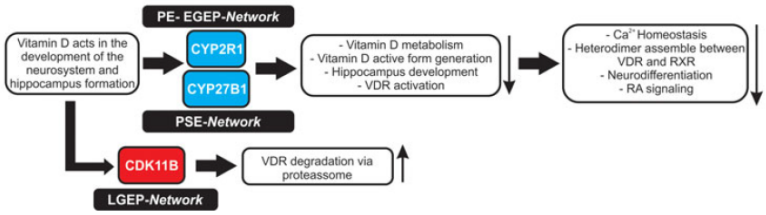
a) Insert a screenshot that shows the interactome of Bat2 in baker's yeast and the types of evidence that supports these interactions. You must include the web site's figure legend to show the types of interactions present.

b) Briefly describe the two most common types of experiments that could provide physical evidence of two proteins interacting. Limit your answers to a maximum of 40 words per method.

1.
2.

15 points

2) This figure is from a systems biology study of fetal alcohol syndrome in a rat model. Blue boxes indicate a bioprocess in which transcription of the associated genes was reduced. Red boxes indicate those which were overexpressed. Thin arrows show up or down regulation of boxed contents. (PDF version available) Abbreviations: Prenatally Exposed PE = PE Network, Early Gestation Exposed Postnatal = EGEP Network; Postnatal-Exposed = PSE Network.



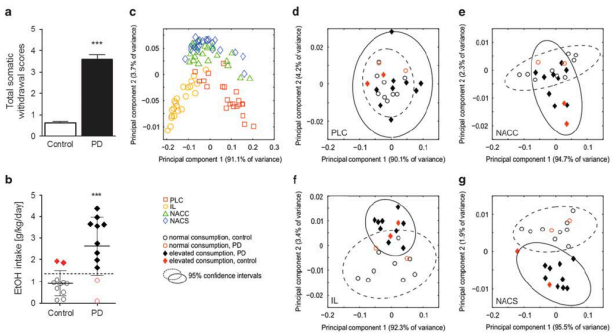
a) List the experimental (wet and/or dry lab) steps you would need to take in order to generate the biological data used to produce this figure. Limit your answers to a maximum of 40 words per numbered step.

1.

b) Draw a directed graph showing how fetal vitamin D is connected to gene regulation, calcium homeostasis and RA signaling. You must cite every source you used to develop your picture. Make your citations hyperlinks so I can view them. If you do this by hand, print neatly so I can read it.

20 points

3) This question looks at the effects of alcohol on adult brains. By this point, you are probably wondering why so much attention to alcohol. I have an extensive family history of alcoholism which is the main reason I have never consumed alcohol. Every spring frolics, I worry about student



long-term and short-term health. I wanted integrate real-world genomics into the final exam. (PDF version available) Do NOT focus on where these areas of the brain are, or what cognitive events happen in these areas. Just consider them 4 black boxes in the brain.

Figure legend: Increased withdrawal physiological responses and long-lasting excessive alcohol drinking in postdependent (PD) rats and PCA of the infralimbic (IL), prelimbic (PLC), accumbens core (NACC), and accumbens shell (NACS) based on annotated metabolite levels. (a) After prolonged alcohol exposure, withdrawal physiological measurements were scored collectively in adult mice. (b) Following 3 weeks of alcohol abstinence, mice had equal access to tap water and 8% (v/v) ethanol. Error bars show 95% confidence intervals. (c–g) The first two principal components from about 100 metabolites measured in this study. Datasets are highlighted based on brain region (c), or alcohol exposure and consumption history of the animals (d–g). Ellipses represent the 95% confidence interval for PCA figures. ***P < 0.001.

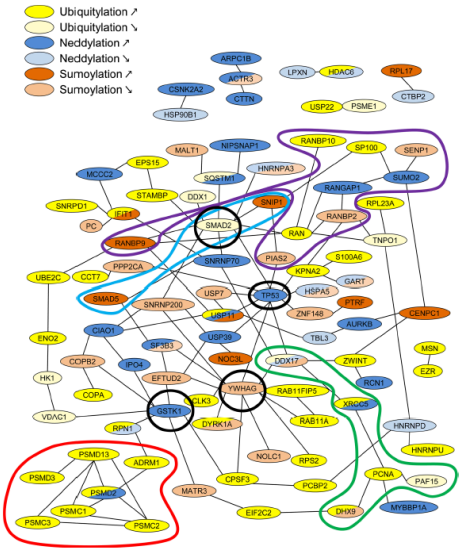
a) What can you conclude from panels a and b? Limit your answers to a maximum of 70 words.

b) Interpret panels c – g and give me the three most important conclusions that you see and connect the data to your conclusions. Limit your answers to a maximum of 80 words per conclusion.

1.
2.
3.

20 points

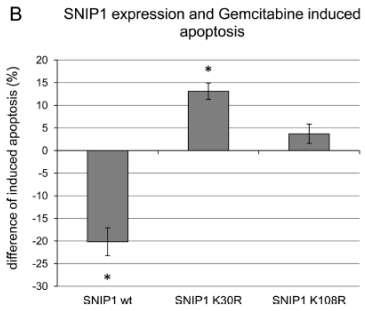
4) Pancreatic ductal adenocarcinoma (PDAC) is a very aggressive cancer that is resistant to all known anticancer therapies. This group explored the role of post-translational modifications (PTMs) mediated by members of the ubiquitin family. This paper studied the ubiquitin-, Nedd8-, and SUMO1-specific proteomes of a pancreatic cancer cell line (MiaPaCa-2) to identify changes induced by gemcitabine, the standard PDAC's chemotherapeutic drug. (PDF version available)



a) Explain how the black-circled nodes could be related to each other. Use the map shown here to guide your searching. Be sure to incorporate the oval color code in your systems-level explanation. Limit your answers to a maximum of 120 words. You may include screen shots to augment your answer. Also, include any web site URLs you used to construct your answer (not part of the word count).

Figure legend for #4a: Interacting proteins among gemcitabine-induced PTMs. Protein–protein interaction databases have been used to draw an interactions map among proteins with altered ubiquitylation, neddylation, and sumoylation following gemcitabine treatment (96 proteins out of 251 from altered modifomes).

b) Panel B shows what happened when two lysine residues were mutated to arginine in the protein called SNIP1; these lysines are sumoylated after gemcitabine treatment in *wt* SNIP1. Incorporate this figure to explain why PDAC is such a difficult cancer to cure. Limit your answers to a maximum of 60 words. (PDF version available)



15 points

5) There are three figures associated with #5. They focus on sequencing mRNAs in circulating cells and DNA in the blood. The cells and the circulating DNA are sloughed off from cancers. (PDF version available)

a) Interpret figures 5A & 5B. Limit your answers to a maximum of 80 words.

b) How could the information in 5A and 5B be used clinically? Limit your answers to a maximum of 70 words.

c) In panel 5C, patients were treated with chemotherapy to block the EGF receptor signal transduction. Circulating DNA was sequenced from their blood as indicated. What are the clinical implications for data such as these? Limit your answers to a maximum of 70 words.

Table 2. Comparison of CTCs with ctDNA.

Sample ID	Tumor type	Clinical stage	Cellular DNA (mutant fragments per 5 ml)	Plasma DNA (mutant fragments per 5 ml)
BLD 21	Bladder cancer	2	0	226
BLD 24	Bladder cancer	2	0	4
CRC 12	Colorectal cancer	4	0	79
CRC 14	Colorectal cancer	4	0	31
CRC 31	Colorectal cancer	1	0	35
CRC 32	Colorectal cancer	2	0	37
CRC 35	Colorectal cancer	2	0	5
CRC 40	Colorectal cancer	1	0	25
CRC 60	Colorectal cancer	4	680	73,000
CRC BIO 23a*	Colorectal cancer	4	370	21,000
CRC BIO 23b*	Colorectal cancer	4	400	28,000
BR 833	Breast cancer	2	0	2,500
BR 834	Breast cancer	2	0	41
BR 837	Breast cancer	2	0	3
BR 841	Breast cancer	2	0	690
BR 848	Breast cancer	2	0	9,900

Figure 5A

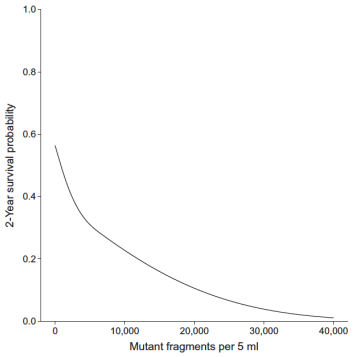


Figure 5B

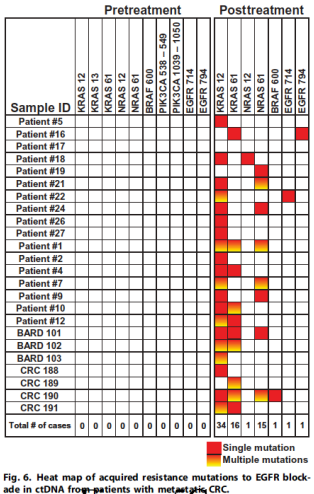


Fig. 6. Heat map of acquired resistance mutations to EGFR blockade in ctDNA from patients with metastatic CRC.

Figure 5C

15 points

6) This question is different than the others… Here is the abstract from a classic 2002 paper that first discovered 3- and 4-node networks in biological systems.

> Complex networks are studied across many fields of science. To uncover their structural design principles, we defined "network motifs," patterns of interconnections occurring in complex networks at numbers that are significantly higher than those in randomized networks. We found such motifs in networks from biochemistry, neurobiology, ecology, and engineering. The motifs shared by ecological food webs were distinct from the motifs shared by the genetic networks of *Escherichia coli* and *Saccharomyces cerevisiae* or from those found in the World Wide Web. Similar motifs were found in networks that perform information processing, even though they describe elements as different as biomolecules within a cell and synaptic connections between neurons in *Caenorhabditis elegans*. Motifs may thus define universal classes of networks. This approach may uncover the basic building blocks of most networks.

Below are two figures from this paper. Reading the full paper would **not** help you answer this question because the question is less data-driven than most: what functional aspect of feedforward loops make them so well suited for the three contexts in the table (below), but not foodwebs? In other words, explain why feedforward loops are adaptive in biological and human-designed networks, but not food webs. (NO PDF versions because the data are not central to this question) Limit your answers to a maximum of 100 words. Cite your sources (not part of the word count)

Concentration (C) of the feedforward loop motif in real and randomized subnetworks of the *E. coli* transcription network . C is the number of appearances of the motif divided by the total number of appearances of all connected three-node subgraphs. Subnetworks of size S were generated by choosing a node at random and adding to it nodes connected by an incoming or outgoing edge, until S nodes were obtained, and then including all of the edges between these S nodes present in the full network. Each of the subnetworks was randomized (shown are mean and SD of 400 subnetworks of each size).

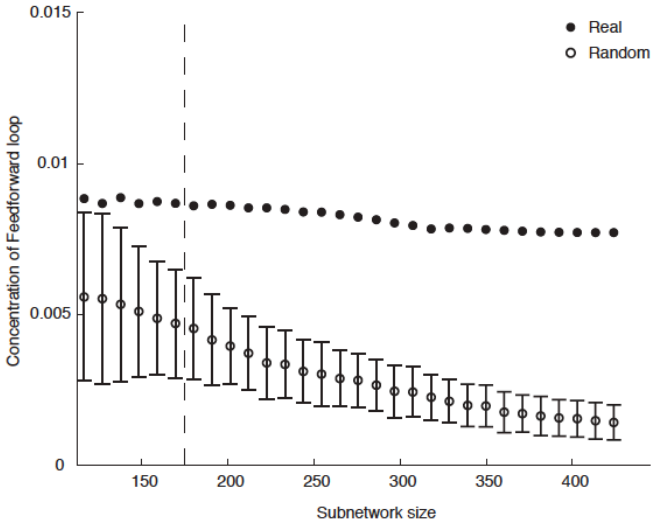
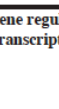
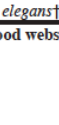
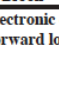


Table: Network motifs found in biological and technological networks. The numbers of nodes and edges for each network are shown. For each motif, the numbers of appearances in the real network (N_{real}) and in the randomized networks ($N_{rand} \pm SD$, all values rounded) are shown. The p value of all motifs is $p < 0.01$. As a qualitative measure of statistical significance, the Z score = $(N_{real} - N_{rand})/SD$. NS, not significant.

Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)					**Feed-forward loop**
E. coli	424	519	40	7 ± 3	10
*S. cerevisiae**	685	1,052	70	11 ± 4	14
Neurons					**Feed-forward loop**
C. elegans†	252	509	125	90 ± 10	3.7
Food webs					**Three chain**
Little Rock	92	984	3219	3120 ± 50	2.1
Ythan	83	391	1182	1020 ± 20	7.2
St. Martin	42	205	469	450 ± 10	NS
Chesapeake	31	67	80	82 ± 4	NS
Coachella	29	243	279	235 ± 12	3.6
Skipwith	25	189	184	150 ± 7	5.5
B. Brook	25	104	181	130 ± 7	7.4
Electronic circuits (forward logic chips)					**Feed-forward loop**
s15850	10,383	14,240	424	2 ± 2	285
s38584	20,717	34,204	413	10 ± 3	120
s38417	23,843	33,661	612	3 ± 2	400
s9234	5,844	8,197	211	2 ± 1	140
s13207	8,651	11,831	403	2 ± 1	225

Reminder, I am NOT asking you to interpret these data. I am giving you the data as background. There is only one question you have to answer for #6: explain why feedforward loops are adaptive in biological and human-designed networks, but not food webs. This question will require you to find good sources online and then you have to THINK of a good explanation. Limit your answers to a maximum of 100 words. Cite your sources (not part of the word count)