Spring 2018 Genomics Exam #1
Genomic Sequences & Variations

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 February 14. **ELECTRONIC COPIES OF YOUR EXAM ANSWERS ARE DUE BY 9:30 am ON WEDNESDAY FEBRUARY 14**. 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.

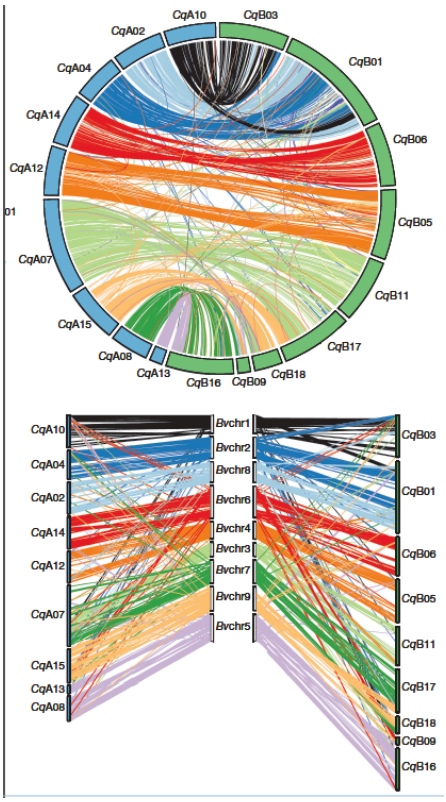
Name (please type):

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

How long did this exam take you to complete?

17 pts

1) There is a nutritious, diploid grain called quinoa (pronounced keen'-wah; *Chenopodium quinoa*) whose genome was sequenced recently. Figure 1 shows the 18 chromosomes that were sequenced. Lines connect paralogs within the genome (top panel), colors are used for clarity only. Lines in (bottom panel) connect orthologs in the diploid food crop *Beta vulgaris* that has 9 chromosomes.



a) Describe the genome-wide event that contributed to the evolution of *Chenopodium quinoa*. Support your answer with data.

whole genome duplication

paralogs on every chromosome

b) Look at chromosome CqB16 in Figure 1 top panel and describe what happened at the chromosome level to its paralogous chromosome. Support your answer with data.

Split into 8 and 13, synteny retained

15 pts

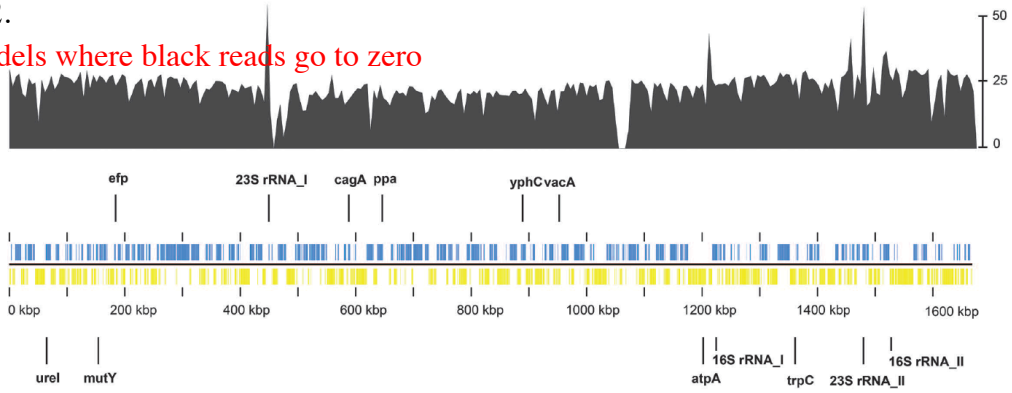
2) You may have heard about the "Iceman" whose remains were discovered by hikers in the mountains along the Austrian/Italian border after the body became exposed due to melting glaciers. Genomicists isolated ancient *H. pylori* DNA from his stomach and sequenced the ancient microbes' genomes. Figure 2 shows read depth (black graph) above the modern *H. pylori* genome (blue = CDS on + strand, yellow = CDS on – strand).

a) Explain what is indicated where the 4 rRNA genes are located by the read depth in the ancient genome.

High read depth/coverage = paralogs/CNV that got compressed in assembly

b) Summarize the two biggest differences you see between the ancient and the modern genomes based on Figure 2.

Two big indels where black reads go to zero



17 pts

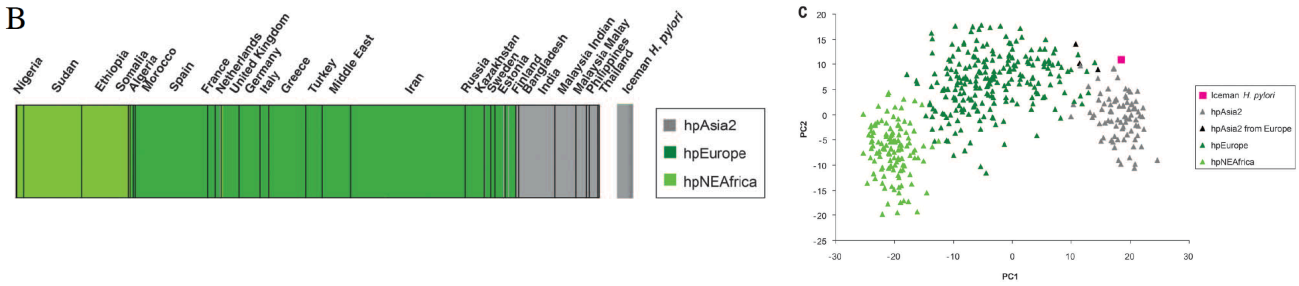
3) Figure 3 shows more information about the Iceman's *H. pylori* genomes. This time, the investigators compared the ancient genomes' SNPs to those from current *H. pylori* taken from modern people, as indicated in the figure (assume "middle east" refers to Syrian, Israel and the Arabian peninsula).

a) Interpret Figure 3B and use data to support your interpretation.

Bacteria similar to Asian isolates

b) Figure 3C shows principle component analysis of many modern *H. pylori* genome sequences. Interpret Figure 3C while being consistent with your answer about Figure 3B and the geography of where Iceman was found. Support your answer with data.

Most similar to Asian (PC1) but not far from some European isoaltes. PCA2 shows distinct from most Asian isolates.



17 pts

4) Figure 4 comes from the latest installment from a very famous cancer researcher – Bert Vogelstein. In these reports, Vogelstein and his colleagues have been measuring the correlation between the number of cell divisions by adult stem cells and the incidence of cancers for a variety of organs (17 total).

a) Figure 4A shows the number of countries around the world (y-axis) that exhibit particular correlations (x-axis) between the number of cell divisions by adult stem cells and the lifetime incidence of cancer in any organ. Interpret Figure 4A and connect your interpretation to a possible mechanism to explain why the correlation is so high.

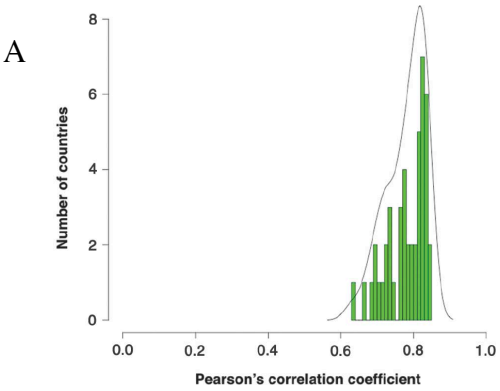
Number of divisions determines probability of cancer mutation accumulating

b) Figure 4B is a very clever infographic that summarizes the relative contribution of three possible sources of cancer driver mutations. Explain what environmental factors are causing cancers (far right).

smoking lung, UV skin, cervix viral infection, stomach H. pylori

c) What is the take-home message about what individuals can do to prevent most cancers?

You cannot prevent, only detect early to treat and survive



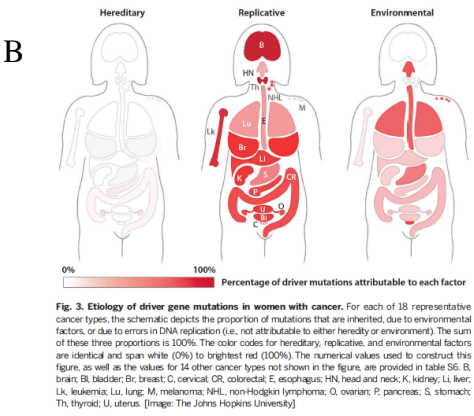


Fig. 3. Etiology of driver gene mutations in women with cancer. For each of 18 representative cancer types, the schematic depicts the proportion of mutations that are inherited, due to environmental factors, or due to errors in DNA replication (i.e., not attributable to either heredity or environment). The sum of these three proportions is 100%. The color codes for hereditary, replicative, and environmental factors are identical and span white (0%) to brightest red (100%). The numerical values used to construct this figure, as well as the values for 14 other cancer types not shown in the figure, are provided in table S6. B, brain; Bl, bladder; Br, breast; C, cervical; CR, colorectal; E, esophagus; HN, head and neck; K, kidney; Li, liver; Lk, leukemia; Lu, lung; M, melanoma; NHL, non-Hodgkin lymphoma; O, ovarian; P, pancreas; S, stomach; Th, thyroid; U, uterus. [Image: The Johns Hopkins University]

19 pts

5) Figure 5 continues on the cancer theme. In this paper, the investigators were quantifying different types of point mutations in human adult stem cells (ASC).

a) Interpret Figure 5a and support your answer with data.

For 3 cancers, mutations outside exons, more mutations for late replication, associated with H3K9me3

b) Are the results in 5a consistent with, or contradictory to, the data we have seen before? Cite your source of data to which you are comparing Figure 5a.

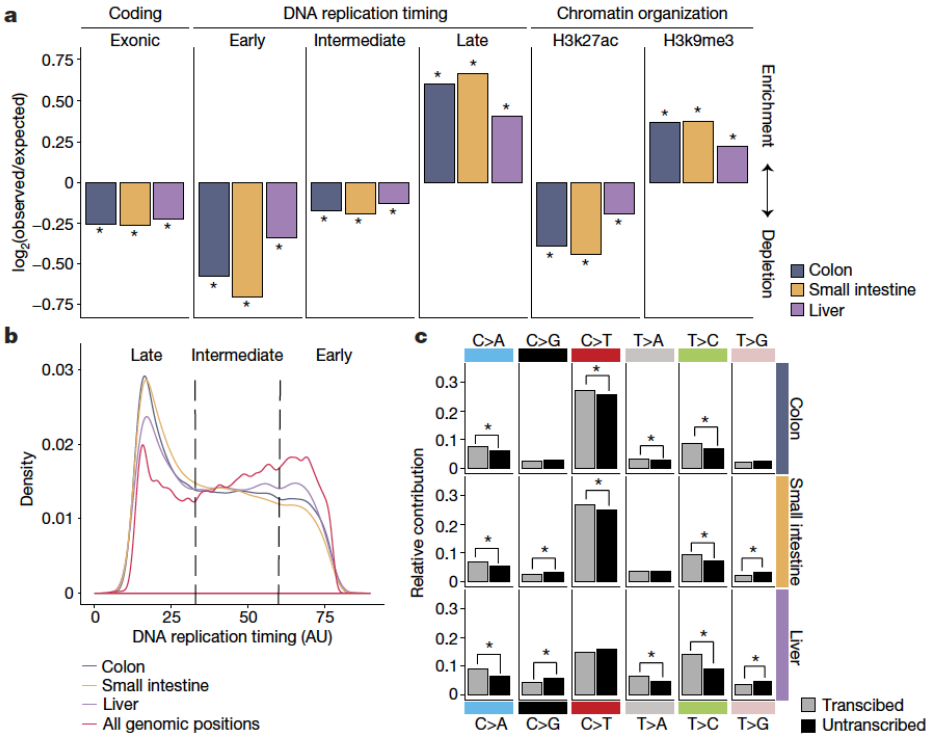
similar to paper on mutations in replicating cells

c) Figure 5b summarizes data similar to 5a. What graphical error/oddity is apparent in this figure? Explain how you would deal with this error/oddity.

All positions (red) has two lines. I assumed the flat line to be an error.

d) Without going into all the detail of each panel, summarize the main points in Figure 5c. support your answer with specific examples.

Different cancers have different types of mutations; T>C and C>A only exceptions



15 pts

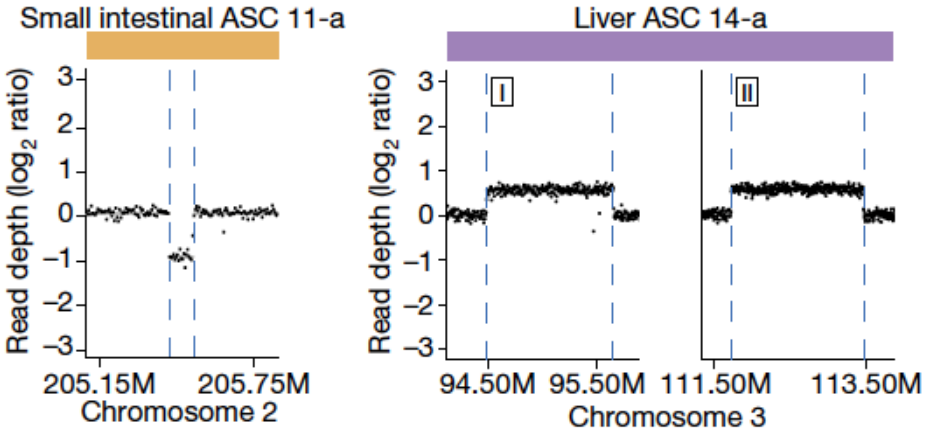
6) The final question continues to examine data from the sequencing of many individual human adult stem cells (scASC).

a) What can you deduce the about mutation status of the gene shown in Figure 6, on chromosome 2 (left). Support your answer with data.

Deletion in one of two alleles

b) Explain what has happened to the region of chromosome 3 that carries two copies (I and II) of the same gene in close proximity. Support your answer with data.

Internal partial duplication followed by whole gene duplication to produce matching pairs of mutated genes


Small intestinal ASC 11-a
Liver ASC 14-a
Read depth (log2 ratio)
3
2
1
0
−1
−2
−3
205.15M
205.75M
Chromosome 2
I
II
94.50M
95.50M
111.50M
113.50M
Chromosome 3