This examination is worth 35 points. Take your time, read the questions carefully, and answer them concisely. **Be sure to answer the question I am asking**. Many questions have more than one possible correct answer. Be sure to explain your logic in arriving at the answer you give since explaining your thinking helps me to decide whether a wrong answer still might deserve partial credit. Good luck!

1. (8 points) Liz Parrish is the CEO of BioViva Sciences, a biotechnology company developing treatments to cure the "disease" of aging in humans. In 2016, she had herself injected with two types of proprietary gene therapy agents: one of these inhibits production of myostatin, an inhibitor of muscle growth, the other is intended to increase the length of telomeres at the end of chromosomes in her T-lymphocytes. BioViva claimed in a press release that Liz's telomere length increased from an average of 6.71 kb to 7.33 kb which, if true, could represent a 20 year decrease in her apparent biological age. You have been engaged by the SEC to evaluate the validity of these claims and have samples of Liz's blood from before and after the therapy.

a) (4 points) What approach would you take to determine whether the length of Liz's telomeres has actually changed in her T-lymphocytes as a result of the treatment? What assumption(s) must you make in order for this to work? Stick to methods we have discussed rather than getting too creative.

There are a variety of relatively imprecise methods available to quantitate telomere length. Considering only the techniques we have studied, you will want to use either direct sequencing of the telomeres with 3rd gen sequencing (the only way to generate such long sequences), or develop a PCR-based method that can allow you to establish telomere length. In either case, you will first need to isolate T-lymphocytes from the blood samples, then isolate DNA from these. If you are using 3rd gen sequencing, prepare the samples appropriately (PacBio or Nanopore libraries) and sequence. One assumption you must make is that there is a way for you to sequence only the telomeres, perhaps by some form of anchored or primed sequence analysis or amplicon sequencing. If you perform PCR, electrophoresis against appropriate size standards will be your method of choice. Whole genome sequencing is not going to work because telomeres are long stretches of repetitive DNA that will not readily assemble and certainly not well enough to look for ~10% changes in the length.

b) (4 points) The gene therapy is based on a viral delivery system that presumably can target any type of cell that has a nucleus. Therefore, while only T-lymphocytes were tested in the initial experiments, Liz has recently had the viral vector infused into her blood which might allow the virus to reach nearly all cells of her body to inhibit aging; Liz is nearly 50 years old would not like to look any older than she is today. Since the immune system appears to be impaired with aging, Liz's claim that her T-lympocytes are "younger" than before the treatment could also extend to other white blood cells. **How could you test the hypothesis that the gene therapy affected which genes were expressed in all circulating white blood cells before and after the therapy? That is, what might be the best way to identify changes in gene expression that might be associated with rejuvenation in some, but maybe not all white blood cells?**

Since you are being asked to look at gene expression in WBCs before and after therapy, a single cell RNA-seq approach is indicated. Pick a method that you are happy with (e.g., 10x Genomics) isolate single cells, prepare deep sequencing libraries and conduct Illumina sequencing. Plotting the results will identify which genes are expressed in what populations of cells.

2. (12 points) After graduation from UCI, you were recruited to NASA where you will be part of a team that characterizes the biological material returned from a probe that has been secretly sent to collect samples from Europa – a moon of Jupiter that was long suspected to have liquid water under a crust of ice. Indeed, Europa has an ocean under the ice and NASA has quietly returned a sample to Earth for characterization.

a) (3 points) The first task is to determine whether there are any living organisms in the sample. **How might you go about identifying how many different organisms are in the sample with a high probability of not missing any? What is the main assumption you must make in order for this to be successful?**

This is based on a modification of the Venter paper. You will want to extract DNA from the sample and sequence it directly using nextgen sequencing. The main assumption you have to make is that these organisms will contain DNA that is sufficiently similar to that of terrestrial organisms so that you can sequence it using standard methods and assemble into genomes by computer. If you are lucky it will be and a nextgen version of whole genome shotgun will work well.

b) (2 points) You are lucky - your sample yields 11,279 sequence assemblies, 8017 of which are complete, or nearly complete. **Are these genomes likely to be prokaryotic or eukaryotic? Why? How will you determine whether these sequences are *bona fide* residents of Europa vs. contaminants introduced on Earth by a sloppy scientist (not you, of course) ?**

If you got 8017 complete genomes, then these are probably small, making it likely that they will be prokaryotic. I will compare these sequences with those available in the GENBANK database and determine whether they are identical to any organisms found on Earth. If they are not, then they have likely originated on Europa. If some are identical, or highly similar to terrestrial organisms, then you would suspect contamination.

c) (3 points) Unfortunately, your colleagues on the project develop a strange illness. Orange fur grows all over their bodies and after two weeks, they develop high fevers, become delirious and demand to be BioSci majors at UCI. Sadly, they die shortly afterward. You are the type of person who makes lemonade out of lemons, instead of complaining that they are not sweet enough. It is tragic that your colleagues died, but the identification of Europan DNA sequences on their skin suggests that the microorganisms can be cultured. Sure enough, you are able to culture two types of bacteria from the skin of the first deceased patient. **Design a diagnostic test to determine whether someone carries these Europan bacteria on their skin, or whether it is present in any sort of random sample to be tested? Note the key features of your assay.**

After sequencing the bacteria, I would develop a RT-PCR assay using Taqman or similar technology to quickly identify these sequences in any sort of sample you receive. Such technology uses custom primers to specifically and quantitatively identify matching sequences in any number of samples. PCR makes it very sensitive.

d) (4 points) Clearly, causing a person to develop orange skin and later die is a serious problem. **Describe a relatively thorough approach to identify and track the types of changes over time induced in patients by exposure to the Europan organisms ?**

The approach I was looking for was that of the Chen paper. Since you have learned about genomic and transcriptomic analyses, it would be appropriate to perform whole genome sequencing to rule out any mutations induced by the infection and transcriptome analysis to identify changes in gene expression over time. It would be best if you had pre-infection samples to compare with those after the person was presumed to be infected

3. (8 points) The plot thickens. Patients, doctors and nurses in the local hospital begin to show signs of the disease – orange fur. Suddenly Anderson Cooper is broadcasting from in front of the hospital, bravely not wearing hazmat gear (mostly because he does not look good in a hazmat suit). Intriguingly, some of the patients do not become delirious and die, they simply have orange fur.

a) (2 points) The first thing you need to figure out is how the contamination spread from the original patient to others. You recall that your sarcastic D145 professor said that doctors are a major source of contamination in hospitals, but you don't believe him. The patients were immediately put into isolation after arriving in the hospital so it is very unlikely that the contamination spread without personal contact**. How could you test whether the Europan microorganisms in the attending ER physician, Dr. Quackenbush, are the same as in the initial patients from NASA and in the others who subsequently become affected?** Dr. Quackenbush appears to be unaffected.

Perhaps the best way to do this is to sequence microorganisms collected from the various patients and swabs from Dr. Quackenbush's body. If the sequences are identical, AND present on Dr. Quackenbush, then one appropriate conclusion would be that he is an intermediate carrier of the infection. You could also use high resolution fingerprinting to compare the strains with each other. Expression microarrays are not sufficient to discriminate closely related species as these are likely to be.

b) (3 points) Intriguingly, Dr. Quackenbush appears to have the Europan microorganisms on his hands and on his lab gloves, but remains unaffected. Nevertheless, he is placed into isolation and observed. As more data are collected, you notice that there are 3 phenotypes among people who show evidence of Europan microorganisms on their skin (detected with the test you developed in 2c). These are a) orange fur, delirious, then dead, b) orange fur but otherwise healthy, and c) unaffected. This sounds to you like a classic case of a single gene that differs when it is homozygous for one allele (sensitive, i.e. dead), or the other allele (resistant), or heterozygous (orange but alive). You perform whole genome association studies and quickly discover that one region on chromosome 20 (about 1 megabase long) is highly associated with resistance to the Europan microorganisms. **How might you determine whether this region is transcribed into mRNA and, if so, how large the transcripts are and how much of the transcript is present?**

Make RNA from skin of affected people and perform Northern analysis using the region from chromosome 20 as a probe to detect whether transcripts are present and what size they are. Add a standard to determine how much is present. Northern is the only way to detect size but you could use QRT-PCR to quantitate with appropriate standards.

c) (3 points) You have identified several families where one parent is totally resistant to the infection, the other parent gets orange fur and the children are either resistant or get orange fur but do not die. Your analysis in b) above has shown that there are no differences in the coding capacity of this region between resistant, partially resistant or susceptible individuals and that the sequence of the affected region does not change significantly among patients. The data suggest that a variation in copy number of some part of the candidate region might be responsible for resistance. **Outline how you would 1) identify copy number variations in this region and 2) determine which patients have the CNV and 3) determine whether CNVs in the region on chromosome 20 are linked with the disease?**

This would use an approach similar to that in the Redon paper. 1) Use SNP or chromosome tiling microarrays to identify CNV regions and 2) test this in all sorts of patients. 3) If the CNV is linked with the disease, then the patients should segregate into groups where the number of copies is correlated with the severity of disease.

4. (7 points) As you might haveexpected, Anderson Cooper has now developed orange fur. CNN is desperate to know whether or not he will survive and the ongoing drama is generating big ratings. They need to know how long the ratings bonanza will continue so that they can project how much advertising revenue they can expect. CNN has brought Larry King out of retirement to anchor the news story since he is so old that no one cares whether he will contract the disease (plus it makes him look brave).

a) (1 point) **How could you determine whether Anderson will die, or simply remain orange (other than waiting and observing what happens)** ?

Assuming you showed that CNVs are associated with resistance in 3c, use the same method to determine which group Anderson falls into.

b) (3 points) Continuing to make lemonade out of lemons, you hypothesize that understanding how the Europan microorganisms infect skin cells and lead to the generation of orange fur might lead you to a cure for baldness and give you financial security forever. You found that the microorganisms behave somewhat like *Chlamydia* and grow as intracellular parasites in skin stem cells. Your analysis suggests that the infected skin stem cells are instead transformed into a cell type like hair follicle stem cells. **Assuming that you can isolate both cell types in pure form, outline how would you determine what transcripts are expressed in normal skin stem cells compared with infected stem cells?**

Two methods come to mind. The first would be gene expression microarray analysis. You would prepare mRNA from both cell populations, label it and probe appropriate microarrays. You would then compare the genes that are expressed in normal skin stem cell and the infected, hair follicle stem cell-like cells. Genes that are expressed differently will be the ones that you focus your attention on. A better way would be to perform RNA-seq of cDNA derived from the mRNA to identify which sequences are different between the 2 populations. This method has the benefit that it would detect microorganism encoded transcripts that would not appear in your analysis otherwise.

c) (3 points) Another possible hypothesis is that something about the infection of skin cells with the Europan microorganisms has mobilized one or more LINE elements (retrotransposon-like repeated sequences) allowing them to hop around the genome and cause the strange effects observed. **How could you determine whether LINE elements in cultured skin cells infected with Europan organisms compared with uninfected cells, move from their original location in the genome to a new location?**

First you need to culture skin cells from regions with orange fur (infected) and regions of normal skin (presumably uninfected). The most brute force method would be to conduct genome sequencing from both cells and focus on where LINE elements are located in each. If they are in different places, then you will have proven that they are moving around. Alternatively, you could use FISH to show that the chromosomal patterns of LINE localization are different between the two cell types. Credit was given for a variety of creative answers.