1a) $\Delta G0 = RTlnKeq$ = −8.314 ∗ 310 ∗ ln13000 = − 30.3$kJ\ mol$ = −5.04 ∗ 10−20J **6pts**

1b.) ΔG = $\Delta G0 + RTln$ ($[ADP][Pi]$ $[ATP][H2O]$) = − 30.3$kJ\ mol$ + 8.314 ∗ 310 ∗ ln (10−4 ∗10−3 10−2∗1) = 60.0$kJ\ mol$ = −9.97 ∗ 10−20J = −99.7$pN * nm$**6pts**

1c.)ΔG = $\Delta G0 + RTln$ ($[ADP][Pi]$ $[ATP][H2O]$) = − 30.3$kJ\ mol$ + 8.314 ∗ 310 ∗ ln (10−3 ∗2∗10−3 2∗10−3∗1) = − 48.1$kJ\ mol$ = −8.00 ∗ 10−20J = −80.0$pN * n$m **6pts**

1d.) d) 100$pN * nm$ = 10−19J = 60.2$kJ\ mol$, this value is close to the value in b. And 80-100 pN*nm here is consistent with the value reported for the free energy of ATP. **5pts**

2a.) We need the forward and the reverse primer, obtained from the initial sequence and reverse complement of the ending sequence of the DNA **6pts**

2b.) 2^20-2*20. The 2^20 refers to the number of DNAs after 20 rounds of replication. The 2*20 term comes from the fact that for every cycle the original unlabeled DNA strands are also replicating. **6pts**

2c.) The simplest method would be gel electrophoresis, where the short primers would travel through the gel much faster than the long genes. The genes can be recovered through agarose digestion. **6pts**

3. Gene Chips (see diagram)

a) Explain why a gene chip (i.e. a DNA Microarray) would be ideal to use when determining which genes are being turned on (i.e. proteins expressed) and which genes are being turned off during cell division (or any other cell process).

There are many, MANY genes, and a priori we do not know which ones are up or down regulated in any given time. Thus we need some sort of technology that 2 can deal with a large number of genes, and preferably can be automated so no poor graduate student needs to move around small volumes of liquid for every single combination. Indeed, even from the very earliest days of the technology, robots did all the experimental preparation. **6pts**

b) You have isolated the mRNA from cancerous tissue and labeled it with a red fluorescent dye. You also isolate the mRNA from a healthy version of the same tissue and label it with a green fluorescent dye. After mixing the mRNA and hybridizing it to a human gene chip you see the results below. Each Gene (genes 1→15) has its own spot and the color of each spot is given by the letters G→Green, R→Red, B→Blank (i.e. no color at all) and Y→ Yellow. Qualitatively, what does

yellow correspond to in terms of gene expression in the cancerous and healthy tissue? If you were to design a drug that could *decrease* any particular gene expression, which gene(s) from the diagram might you choose to target and why?

This might not be what you learned via the color wheel, but if both green and red dyes are fluorescing, it looks yellow. Thus a yellow color would indicate that the gene is being expressed both in healthy and cancerous tissue. For downregulation, yellow genes would be poor targets. If it is expressed in both tissues, it is liable to have many side effects in healthy cells and not actually change the phenotype. Instead, we would use our magic chemistry wand to downregulate genes only expressed in cancerous cells (red), as they have a better chance of being oncogenes. **12pts**

c) If you could design a drug to *increase* the expression of any particular gene product, which gene(s) from the diagram might you choose as a target and why?

We want to increase the expression of tumor suppressor genes to try and combat the cancer. Clearly, they are not already expressed in the cancerous cells, and are probably active to some degree in normal cells, so you would be looking at a blank or a green gene on the chip. **6pts**
NOTE: Cancers can be caused by overexpression of certain genes, called oncogenes, or suppression of other genes, caused tumor suppressor genes.

4.) DNA Sequencing.
A general video on DNA sequencing is:

How to sequence the human genome

https://www.youtube.com/watch?v=MvuYATh7Y74

Read it!

a. A genome, she says, are genes "plus some extra that make up an organism". How much of your DNA is "this extra DNA"? What is it useful for, if anything?

Extra DNA is useful for regulating transcription and translation. It also codes for tRNA, rRNA, and telomeres. **6pts**

b. Once we have all of the genes read, we can tell about the genes—your eye color, for example? Or can we? What does the "extra" DNA have to do with how the genes are turned on or turned off?

Eye color is controlled by many genes. As stated in part a, extra DNA helps regulate transcription and translation. It makes up promoter regions and regulatory sites. **6pts**

c. Later she says: "Scientists then take a picture of these fluorescently label DNA, capturing their sequence and then recombined."

In fact, there are many methods for taking this picture and then reassembling it into a complete "book" of your DNA.

We discussed the PacBio method of sequencing DNA. It was very nice in that it could potentially make very long reads, i.e. the sequence of bases were determined by how long the DNA stays on the polymerase, which can be like 10,000 bases. A more standard method is based on the chain termination method, also called the Sanger method. It has much shorter reads.

https://www.youtube.com/watch?v=SRWvn1mUNMA

d. Explain in one paragraph how the Sanger method works.

The method works but is limited to fairly short stretches of DNA, about 400 base pairs long because the gel electrophoresis is not sensitive beyond this range (i.e. it cannot separate out the 400th long piece of DNA from 401.) So to sequence an entire genome, e.g. 3 billion base pairs, what you do is a "shotgun" approach where you add in a primer of random sequence (at least 16 long), sequence the 400 base sequence, then do it again with another random sequence primer. You then have a whole bunch of 400 long sequences, which means that you will have a lot of sequences which overlap each other. By using a computer (and the overlap is at least 16 bases long), you can stitch together the entire sequence.

In the Sanger method, 3 your initial sample and primer are split up into four pools. Each pool will have a high concentration of the 4 standard deoxynuclotides and a lower concentration of a unique type of dideoxynuclotide, in each pool. Dideoxynuclotides will terminate DNA synthesis, so as a result each pool will now contain fragments that are all different lengths (and thus sizes). On a electrophoresis gel, these will form distinct bands from which you can read off the sequence based on how far the bands travelled down the gel. **6pts**

e. A more detailed look at Next Generation Sequencing, the most recent technique(s), can be found:

https://www.youtube.com/watch?v=jFCD8Q6qSTM

Explain the Sequencing by Synthesis, the Sequencing by Ligation, and the Pyrosequencing. Illumina is the company that has largely taken over the DNA sequencing field. Which technique do they use?

Sequencing by Synthesis Here synthesis is directly controlled by reversible fluorescent terminators. After each base, the reversible terminator is cleaved and the next base is imaged. This is what Illumia uses.

Sequencing by Ligation Rather than using the DNA Polymerase base selectivity, this uses DNA ligase to pick out the identity of each base.

Pyrosequencing The concentration pyrophosphate, a byproduct of DNA synthesis, is monitored by the activity of a luminescent enzyme after the addition of different deoxynucleotides. 5 By continuing to rotate through each of the different possibilities, you can identify each base 1 or 2 at a time. 17pts