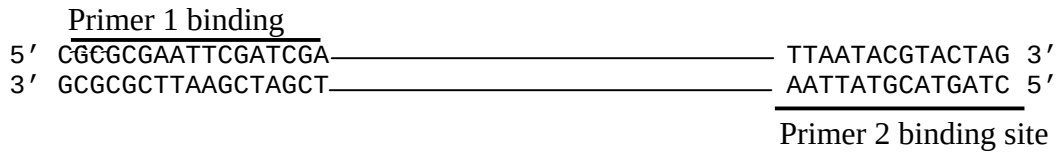
The key: 7.013 Recitation 9 – Spring 2018

1. You are interested in making many copies of a specific DNA sequence. The sequence to be amplified is flanked by regions whose sequence is given below:



	Primer 1		Primer 2
Set A:	5′TCGATCGAATTC 3′	AND	5′TAATACGTACTA 3′
Set B:	5′GCTTAAGCTAGC 3′	AND	5′GATCATGCATAA 3′
Set C:	**5′GAATTCGATCGA 3′**	**AND**	**5′CTAGTACGTATT 3′**

Select the primer pair from above (***Set A/ Set B/ Set C***) that you can use to PCR amplify the gene.

Note: You need a pair of primer to PCR amplify both strands of DNA. The strands should bind to region flanking the sequence of your interest. The binding of primer should be antiparallel to the template strand. It is the 3'-OH end of the primer that gets elongated by Taq DNA polymerase.
Set A: *won't work since although primer 1 may be the primer for the bottom strand, primer 2 is not the primer or the top strand.*
Set B: *primer 1 is binding to the top strand 3'->5' instead of 5'->3', so it will not work.*

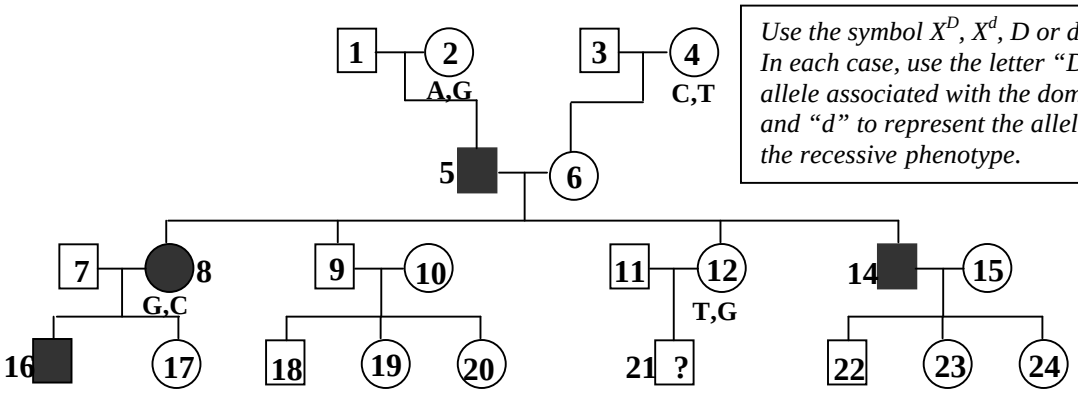
2. You decide to sequence the following piece of DNA.
5'-…GTATAGCCTCCGT…-3'
3'-…CATATCGGAGGCA…-5'

a) The primer you use for sequencing is 5'-ACG-3'. ***Note:*** real primers are usually 18-22 nts long). You set up a DNA sequencing reaction with this DNA, this primer, DNA polymerase, dATP, dTTP, dGTP, dCTP, and ddATP. How many different products would form from this reaction?
Since you are using ddATP in the reaction mixture halts whenever there is an incorporation of a ddATP instead of dATP. Since there are 3A's in the sequence (excluding the A that is a part of the primer) and therefore you may have 3 reaction products of different lengths. There is always a probability that you may have a reaction product where on dATP is incorporated and NO ddATP is incorporated at all. Combining these two together you may have 4 reaction products (5b, 10b, 12b and 13bases long)

b) How many nucleotides long would each different product of this reaction be?
They would be 5b, 10b, 12b and 13bases long.

c) Write the sequence of DNA that you read from this gel below. Label the 5' and 3' ends of the molecule.
5'-…GAGGCTATAC…-3' (the first three bases are the primers)

3. The following human pedigree shows the inheritance of a specific disease. ***Please note:*** *The filled squares or circles represent the abnormal phenotype. The individuals marrying into the family do not have the disease- associated allele. Assume that no other mutation arises within the pedigree. Also assume complete penetrance.*



Use the symbol X^D, X^d, D or d where appropriate. In each case, use the letter "D" to represent the allele associated with the dominant phenotype and "d" to represent the allele associated with the recessive phenotype.

a) What is the **most likely** mode of inheritance of this disease? *X- linked recessive*

b) The disease shown by the pedigree above is caused by a mutation in Gene D that encodes Protein D. You identify a SNP that is **tightly linked** to Gene D and may be used **as a marker for the disease**. The alleles (*A, G, T, C*) of this SNP for some individuals are given in the pedigree above.

iii. Identify the SNP(s) that is/are tightly linked with the mutant allele of Gene D.
C allele from Individual 4; G allele from Individual 2.

iv. Write the SNP genotypes of the following individuals.

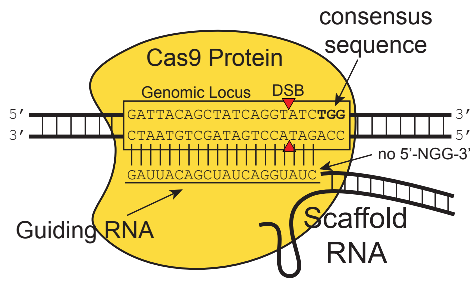
#5: *G*	**#14: *C***

4. Briefly explain how you can use expression microarrays to analyze and treat two patients both suffering from lung cancer.
You would use a microchip for each patient that has solidified DNA probes on it. You would isolate the mRNA from the cancer cells of both patients and use that as a template to make the cDNA. You would fluorescent tag the cDNAs, layer them on the microchip and analyze their hybridization with the DNA probes (under high stringency condition) by looking at the fluorescence intensity.

This would allow you to analyze the variation in the expression profile of the genes from the cancer cells of each patient. Based on this data you can subtype the type of cancer and adopt a personalized treatment regimen.

5. Recent developments in **C**lustered **R**egularly **I**nterspaced **S**hort **P**alindromic **R**epeats (CRISPR) and CRISPR associated endonuclease (Cas9) technologies in eukaryotic cells have allowed for rapid engineering of the eukaryotic genome. A 20-bp guide RNA targets the Cas9 endonuclease to a genomic locus by hybridization with unwound DNA. Following recognition of the complementary strand, Cas9 creates a double stranded break (DSB, indicated with arrowheads) that the cell can repair in an error-prone manner that introduces insertions or deletions. The guide RNA requires a 5'-NGG-3' consensus sequence at the 3' end of the genomic target strand (where N could be any nucleotide). This 5'-NGG-3' is not present in the guide RNA as shown in the schematic.

You are given a sequence below, which encodes a protein that you would like to knock out to study how it contributes to a cell:



Coding: 5′-ATGCACATCTACGCATTTATCTGG**ATGAAAACGGCC**GTTATGCAAGAATTAATTATGAG**TATGAGCACTACTCATAA**TAGTCGAG-3′
Template: 3′-TACGTGTAGATGCGTAAATAGACC**TACTTTTGCCGG**CAATACGTTCTTAATTAATACTC**ATACTCGTGATGAGTATT**ATCAGCTC-5′

5'UTR   Exon 1   intron 1   exon 2   3'UTR

a) Design **one** guide RNA to target this gene of interest and **explain** why you have selected this.

You can give multiple possible answers such as 5'CATAATTAATTCTTGCATAA3', which targets the intron.

b) You look to see whether you have knocked out your protein of interest in a population of cells and notice that not every cell has had the protein knocked out. List **one** possible explanation.
There are many possible answers some of which are below. Any reasonable answer will be accepted. For example, (1) the Cas9 gene/ guideRNA wasn't 100% effective and some cells didn't get hit (2) chromatin structure prevented Cas9 protein from degrading the gene (3) the DNA was repaired via a higher fidelity repair mechanism (4) some cells had a mutation at one or both of the targeting sites, preventing the gene from being completely knocked out.

7.013 Introductory Biology

Spring 2018