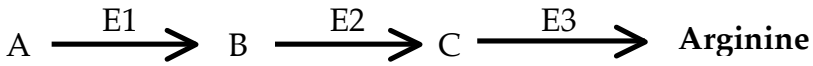
7.016- Problem Set 4

Question 1
Arginine biosynthesis is an example of **multi-step biochemical pathway** where each step is catalyzed by a specific enzyme (E1, E2 and E3) as is outlined below.



Your supervisor gives you three yeast strains, strains 1, 2 and 3, each of which fails to grow in the absence of arginine (arg-). He further explains that Strain 1 has a mutation in Gene 1 that encodes E1, Strain 2 has a mutation in Gene 2 that encodes E2 and Strain 3 has a mutation in Gene 3 that encodes E3. He asks you to clone the allele of Gene 1 that can restore the ability of Strain 1 to synthesize arginine.

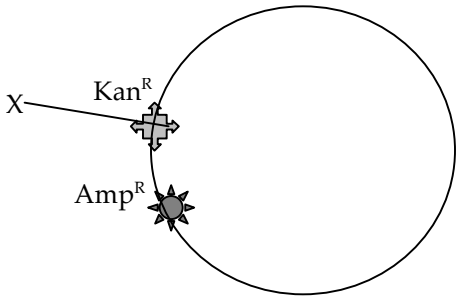
a) You begin by constructing a yeast genomic library in *E. coli* bacterial cells. From the choices below, circle **all** the yeast strain(s) from which you could potentially isolate the genomic DNA to achive your objective. **Explain** why you circled this option(s).

Wild- type **Strain 1** **Strain 2** **Strain 3**

b) You successfully isolate yeast genomic DNA from the yeast strain(s) that you chose in part (a). You need to choose a plasmid that you could use as a vector to create the yeast genomic library in *E. coli* and use this library to transform yeast Strain 1 so that it can now synthesize arginine. List the **minimum features** that your vector should have to execute this plan.

c) You generate a restriction map of the plasmid that you want to use as a cloning vector. You cut the plasmid with the following combination of the restriction enzymes and determine the size of the resulting DNA fragments by DNA gel electrophoresis. The size of the DNA fragments after digestion by restriction enzymes X, Y and Z is tabulated below (kb = kilo base pairs; 1kb= 1000bp). ***Please note*** *that* Amp^R *and* Kan^R *represent the ampicillin and kanamycin resistant genes that are a part of the plasmid vector.*

X	Y	Z	X + Y	X + Z	Y+ Z
4kb	2kb	4kb	1kb	1.5kb	0.5kb
	2kb		2kb	2.5kb	1.5kb
			1kb		2kb



Cloning vector

Complete the schematic of the cloning vector above by showing **ALL** the restriction enzyme sites and writing the distance (in kb) between them.

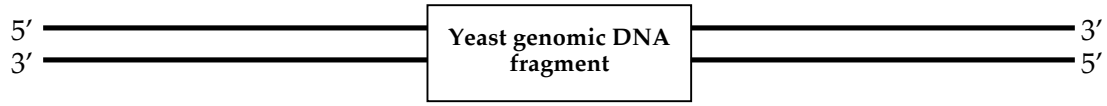
Question 1 continued

d) You then isolate the yeast genomic DNA and digest it with EcoR1 restriction enzyme. You want to clone the EcoR1 digested genomic DNA fragments into the plasmid vector that has the following recognition sites for restriction enzyme X, Y and Z. ***Please note:*** *A vertical line (|) represents the cutting site for each restriction enzyme.*

EcoR1	X	Y	Z
5′G|AATTC3′	5′C|AATTG3′	5′A|AGCTT3′	5′G|GATCC3′
3′CTTAA|G5′	3′GTTAA|C5′	3′TTCGA|A5′	3′CCTAG|G5′

i. Which enzyme (*Choose from X, Y and Z*) would you use to cut the plasmid so that it has ends that are compatible to the ends of the EcoR1 digested yeast genomic DNA fragments?

ii. Write the resulting **6- base pair sequences** at the **two points of ligation** of plasmid and the genomic DNA fragments.



e) You mix the genomic fragments with the cut vectors and add DNA ligase. You then transform *E. coli* cells with the ligation mix and select the clones transformed with the recombinant plasmid by doing replica plating. What growth medium(s) would you use to select these clones? **Explain** your choice.

f) You successfully create a yeast genomic library in *E. coli* cells. How can you use this library to **clone-by- function** the gene that can restore the ability of Strain 1 to produce arginine.

g) You successfully identify a recombinant vector that restores the ability of yeast strain 1 to produce arginine. You are curious to see if this gene can also rescue a **bacterial** cell that cannot produce arginine (arg^-). Your friend suggests that you use her yeast **cDNA** library to attempt to restore an *arg–* bacterial cell to arginine producing wild- type cell (arg^+).

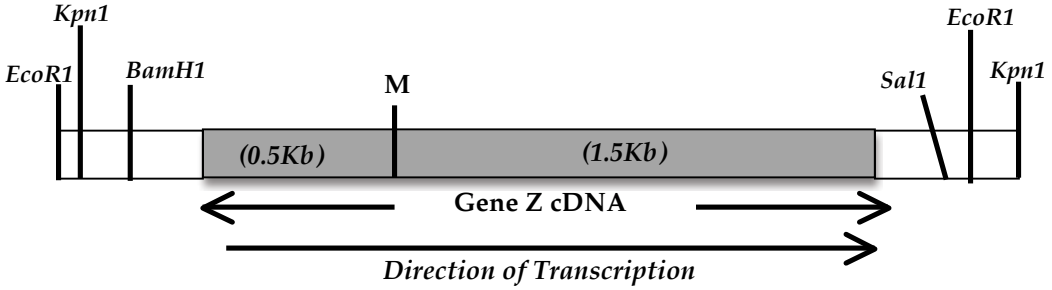
i. Why does your friend suggest that you use a yeast cDNA library instead of genomic library?

ii. List the **minimum features** (besides the features that you listed in part (b) of this question) that your plasmid vector should have to execute the plan outlined above.

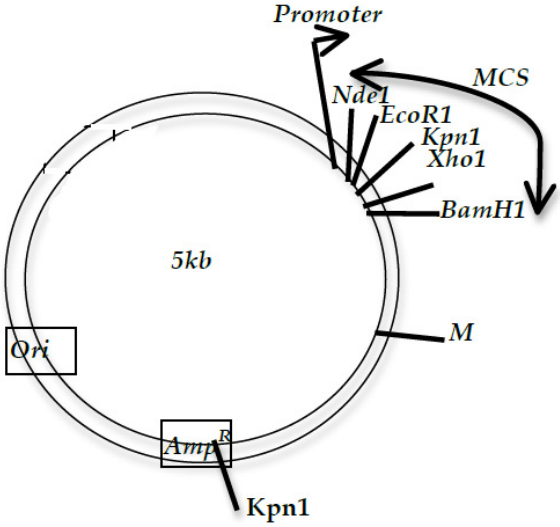
iii. Do you expect that you could use a yeast cDNA library to restore an *arg–* bacterial cell to an arg^+ cell? **Explain** why or why not.

Question 2

You want to clone and express the cDNA copy of a eukaryotic gene, namely Gene Z (2kb) in *E. coli* bacterial cells. The following is the schematic of Gene Z cDNA. ***Please note:*** *The recognition sites of different restriction enzymes (EcoR1, BamH1, Kpn1, M, and Sal1) are given shown.*



You want to use the following plasmid (5kb) as the vector for cloning Gene Z cDNA. ***Please note:*** *This plasmid contains ampicillin- resistance gene (Amp^R). It also has a multiple cloning site (MCS) that has the recognition sequence for restriction enzymes Nde1, EcoR1, Sal1, Kpn1 and BamH1. The recognition sequence for each restriction enzyme at MCS is given below. A slash ("/") represents the site at which the restriction enzyme cuts. The plasmid also has the recognition site for the restriction enzyme M, which is located at a distance of 1kb from MCS site.*



BamH1
5′G/GATCC3′
3′CCTAG/G5′

Nde1
5′CA/TATG3′
3′GTAT/AC5′

EcoR1
5′G/AATTC3′
3′CTTAA/G5′

Kpn1
5′GGTAC/C3′
3′C/CATGG5′

Sal1
5′G/TCGAC3′
3′CAGCT/G5′

Xho1
5′C/TCGAG3′
3′GAGCT/C5′

a) Give three different strategies that you could use to clone Gene Z cDNA into the plasmid.

Strategy	Restriction enzyme(s) used to cut….	
	Gene Z cDNA	Plasmid vector
1		
2		
3		

b) **Circle** the strategy (s) in the table above, that would allow directional cloning of Gene Z cDNA.

c) Put a **"√"** next to the strategy in the table above, that would allow the expression of Gene Z cDNA.

Question 2 continued

d) You next transform the E. coli bacterial cells with a mixture of plasmid and the Gene Z cDNA that you have digested with the restriction enzyme(s) that you selected in part (c), plate them on specific medium, allow the bacterial cells to grow and form colonies and tabulate the results below.

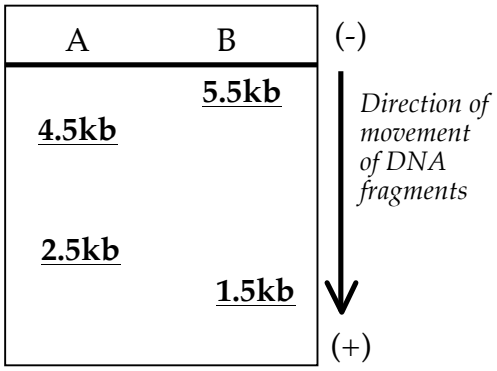
#	Treatment of E. coli cells	Tranformed colonies/ ug of DNA	
		Media	Media +Ampicillin
2	None	2×10^6	<5
3	Gene Z cDNA and Plasmid digested with restriction enzyme that you selected in part (c)	1.1×10^5	40
4	Ligation mix of Gene Z cDNA and Plasmid digested with restriction enzyme that you selected in part (c)	2.1×10^6	150

i. Based on the number of colonies obtained in Row 2 of the table above, give the phenotype of the *E. coli* cells prior to transformation.

ii. Although the *E. coli* cells in Row 3 of the table above have been transformed in the absence of ligase, you still observe a significant number of colonies. Why is this so?

iii. In Row 3 of the table above, do you expect all the transformed colonies to contain recombinant plasmid that has GeneZ cDNA insert? Why or why not?

e) You next select two transformed colonies (Colony A & Colony B) that have the Gene Z cDNA insert. You realize that the gene could have been inserted into the plasmid in two different orientations. So you isolate the recombinant plasmids from two colonies, digest them with restriction enzyme M, resolve the digested fragments by DNA gel electrophoresis, and obtain an approximate profile of the DNA as shown below. Of the two bacterial colonies, which one will express Gene Z? **Explain** your choice. ***Note:*** *You may assume that the restriction enzymes sites in MCS are only a few bases apart and the MCS is <40bp in size and does not significantly influence the resolution of the DNA fragments on the gel.*



Question 3
The following is the **wild- type allele of Gene Z** that you want to amplify using the polymerase chain reaction (PCR).

```
5'CTCGAGGTGAATATGAAAG----|          |----CATTTGGCGCGTAATCGATA3'
                         | Gene Z   |
3'GAGCTCCACTTATACTTTC----|          |----GTAAACCGCGCATTAGCTAT5'
```

a) If you amplify a DNA sequence through PCR what are the reaction components that you would **absolutely need**? Briefly state the function of each of these components.

b) Circle the **sets of primer(s)** from the options below, which you would use for PCR reaction in part (a)?

```
Set1: 5'TATACT3'  and 3'AAACCGC5'
Set2: 5'GAATAT3'  and 3'GTAAACC5'
Set3: 5'GAGTTA3'  and 3'TGGCGAG5'
```

c) In the PCR reaction, you need a three- step reaction cycle, which results in a chain reaction that produces an exponentially growing population of identical DNA molecules. Each step of a reaction cycle is performed at a specific temperature i.e. 95ºC for Step 1, 55ºC for step 2 and 70ºC for Step 3. Briefly **explain** what occurs at each of these three steps.

d) Would all the PCR amplified fragments using the primers in part (b) be of the same size (*Yes*/ *No*)?

e) You decide to determine the complete nucleotide sequence of Gene Z by DNA sequencing using fluorescent nucleotides. You derive the following sequence that includes the **coding region** corresponding to amino acids 1-7 of the protein encoded by Gene Z. ***Note:*** *You may have to determine the open reading frame in the following sequence. A codon chart is provided on the last page.*

5′ GTAGATGGAAAACTTAGGCTATGAA 3′

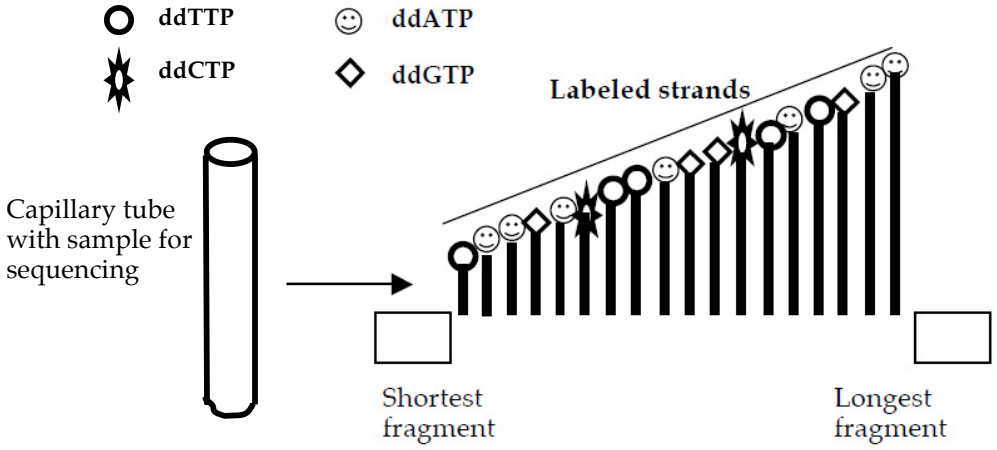
i. Write the DNA sequence for the region that corresponds to the amino acids 2-7 of the protein encoded by Gene Z and label the 5′ and 3′ ends of both strands.

ii. Write the mRNA sequence for the region that corresponds to the amino acids 2-7 of the protein encoded by Gene Z and label its 5′ and 3′ ends.

iii. Write the sequence of the amino acids 2-7 of the protein encoded by Gene Z and label its N and C termini. ***Note:*** *A codon chart is given on the last page of this problem set.*

Question 3 continued

f) You then sequence the mutant allele of Gene Z from an affected individual and observe the following pattern for the **coding/ mRNA like/ non-template region** corresponding to amino acids 2-7 of the non-functional protein.



In the schematic above show the direction of DNA strand synthesis by an arrow and label the 5′ and the 3′ ends by filling in the boxes.Write the sequence of the **non- coding/ template DNA strand** that was used as a template for DNA sequencing.

g) Based on the DNA sequence that you derived and assuming that this sequence represents the coding strand, complete the following table.

Name and position (i.e. gly^{20}) of the amino acid…		Type of point mutation (nonsense/missense/ frameshift/silent)?
in the mutant version of the protein encoded by the mutant allele of Gene Z	in the wild-type version of the protein encoded by Gene Z	

Question 4

A single nucleotide polymorphism (SNP) is a DNA sequence variation occurring when a single base pair in the genome differs between members of a species or paired chromosomes in an individual. By convention this base pair change is represented as one nucleotide — A, T, C, or G — of the base pair.

a) Circle the **best option** from the following choices. The SNPs may exist…

i. Only within the coding sequences of genes

ii. Only within the non- coding regions of genes

iii. Both in the coding and non- coding regions of the genes

iv. In the coding or non- coding regions of the genes or in the intergenic regions (regions between the genes).

Question 4 continued

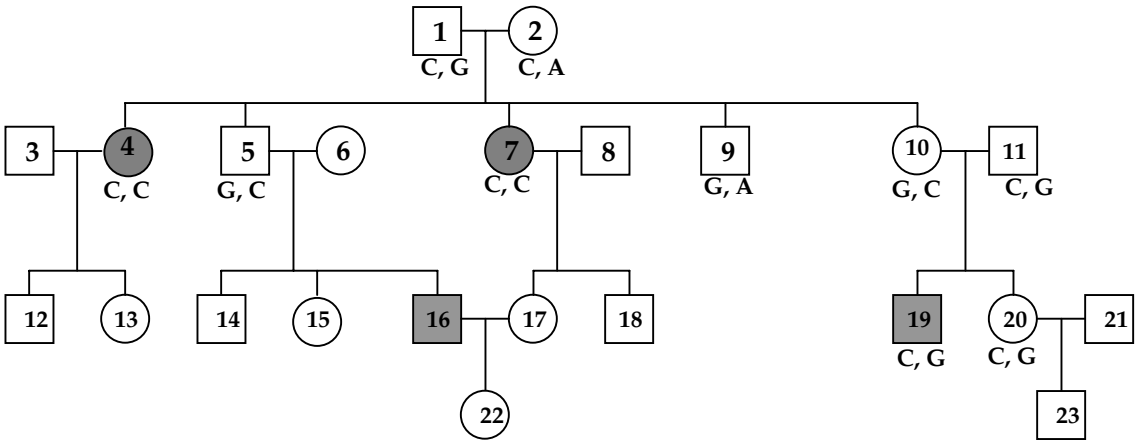
b) In which region(s) (*choose from the coding sequence, non-coding sequence, intergenic regions*) would you expect the SNP to be if...

i. It changes the amino acid sequence of a protein? Include **all** the possible options. **Explain**.

ii. It does not change the amino acid sequence of a protein? Include **all** the possible options and give an **explanation** for each selected option.

iii. It results in a non- functional protein that has an increased amino acid length compared to the wild type protein. **Explain**.

c) Below is the pedigree of a family with a disease. *All the individuals that show the disease phenotype are shaded. The two letters identify the two alleles of the SNP that is absolutely linked to the Gene Z that is associated with this disease. For example G, A indicates that on one of the chromosomes you would find a G (a G/C base pair) and on other chromosome you would find an A (a A/T base pair). Please note that some of the individuals marrying into this family may be carriers. Assume* ***NO*** *recombination.*



i. What is the most likely mode of inheritance (*choose from autosomal dominant, autosomal recessive, X linked dominant or X linked recessive*) of this disease?

ii. Identify the SNP that is **absolutely linked** with the disease- associated allele of Gene Z.

iii. List **all** possible genotypes at the Z locus of following individuals in this pedigree? *Note: 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.*

Individual 2: **Individual 5:** **Individual 9:**

Question 4 continued

iv. Assume that individual #21 is not a carrier of the disease allele. What is the chance that individual #23 is **affected**? **Show your work.**

v. Of the different techniques in recombinant DNA technology that you have been introduced to, **list one** that you can use to identify the SNP that an individual has at a specific location in the genome.

Question 5

You have developed a mouse model for a disease that shows an autosomal recessive mode of inheritance and is associated with Gene A. You decide to adopt the following strategies to cure this disease.

Strategy 1: Using a vector, you successfully introduce one copy of the wild- type allele of Gene A in a zygote that has the male and female pro-nuclei both from affected parents (genotype: aa). You make sure that the introduced copy of Gene A, has successfully integrated into the genome and is under the regulation of tissue specific promoter. You then implant the zygote in a pseudo- pregnant female mouse and let it develop into a newborn.

Strategy 2: You isolate the cell from a developing embryo (at the blastula stage/ 8– cell stage) that is produced by the fusion of gametes from affected parents (genotype: aa). You then infect these cells with a targeting vector that has a wild type copy of Gene A under the regulation of a tissue specific promoter. You then select the cells that have undergone homologous recombination and now have a wild-type copy of Gene A. You re-introduce them into the developing embryo (genotype: aa) to obtain newborns.

a) Which of the above strategies will give you a chimeric mouse: Strategy 1 or Strategy 2? **Explain** why you selected this strategy.

b) Give **all** possible genotypes of the somatic cells of the chimeric mouse produced in Part (i)? ***Note:*** *Use the uppercase A to represent the dominant allele and lowercase a to represent the recessive allele.*

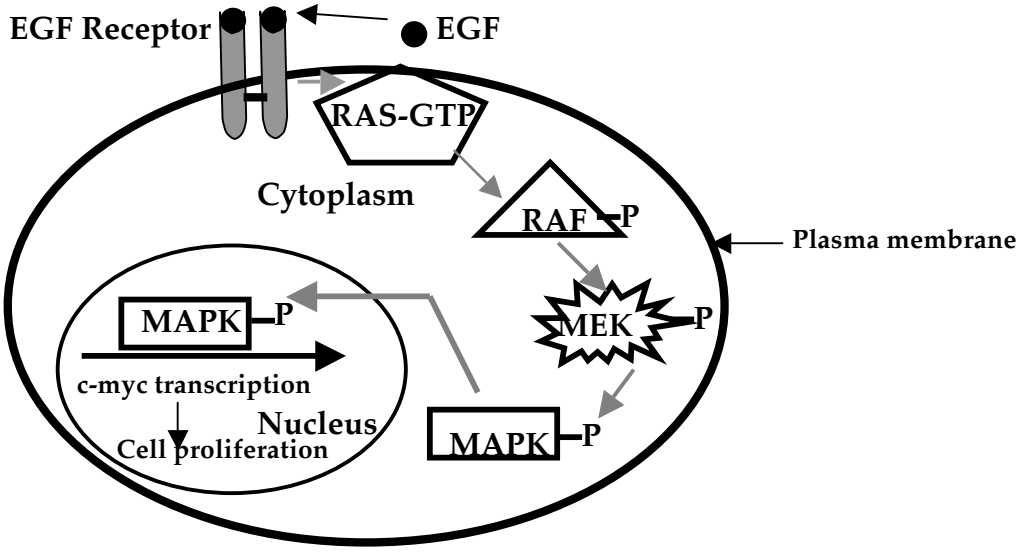
c) You allow the chimeric mouse to mate with a wild- type female mouse (genotype: AA). Would you expect all the mice from this mating experiment to have a wild- type phenotype (*Yes*/ *No*)? **Explain** why or why not.

d) Retroviruses are very often used as the vector to introduce genes into a cell or organism. **Explain** why retroviruses may serve as good vectors to introduce a copy of a gene into a host cell or organism.

Question 6

Following is the schematic of a signal transduction pathway that is activated by the binding of a Epidermal Growth factor (EGF), produced by one cell type, to its specific membrane receptor on a target cell. The major steps involved in this pathway are outlined below:

- *EGF ligand binds to the EGF receptor.*
- *Ligand bound EGF receptors become active through phosphorylation and homodimerization.*
- *Active EGF receptor converts Ras from its GDP bound inactive form to its GTP bound active form.*
- *Active Ras activates the kinase cascade (RAF, MEK and MAPK) through phosphorylation.*
- *This increases the expression of c-myc gene, which **results** in cell proliferation.*



a) Would a cell that has a constitutively active Ras protein show an *increased/ decreased / no change* in the proliferation compared to a wild- type control cell that has been treated with EGF? **Explain** your choice.

b) You are looking at a cell type that shows the deletion of the sequence that corresponds to the kinase domain of MAPK. Would this cell type show an *increased/ decreased / no change* in the proliferation compared to a wild- type control cell that has been treated with EGF? **Explain** your choice.

c) You decide to engineer mammalian cell lines, each expressing a specific mutant variant of either the EGF ligand or the EGF receptor (EGFR).

- **Cell line-1** has a mutation that results in the deletion of **only** the signal sequence of **EGF ligand**.
- **Cell line-2** has a mutation that results in the deletion of **only** the transmembrane domain of **EGFR**.
- **Cell line-3** has a mutation that results in the deletion of **both** the signal sequence and transmembrane domain of **EGFR**.

You incubate each of these mutant cell lines with fluorescent antibodies that specifically bind either to EGF or the EGFR. You then observe these cell lines under the fluorescent microscope to study the localization of EGF ligand or EGFR.

Question 6 continued

i. In cell line-1 where do you expect to find the EGF ligand *(choose from cell membrane, cytosol, cell culture medium)?* **Explain** your choice.

ii. In cell line -1, for the EGF ligand…

- In the DNA strand that is used as a template for transcription, where do you expect to see the base sequence that corresponds to signal sequence of EGF (*close to the 5′ end or the 3′ end*)?
- In the mRNA transcript of EGF, where do you expect to see the base sequence that corresponds to signal sequence of EGF (*close to the 5′ end or the 3′ end*)?
- In the EGF ligand, where do you expect to see the signal sequence (*close to the N- or C- terminus*)?

iii. If cell line-2 is incubated with EGF ligand, do you expect these cells to proliferate? Answer as **Yes/No** and **explain** your choice.

iv. If cell line-3 is incubated with EGF ligand, do you expect these cells to proliferate? Answer as **Yes/No** and **explain** your choice.

Codon Chart

	U	C	A	G
U	UUU Phe UUC Phe UUA Leu UUG Leu	UCU Ser UCC Ser UCA Ser UCG Ser	UAU Tyr UAC Tyr UAA *Stop* UAG *Stop*	UGU Cys UGC Cys UGA *Stop* UGG Trp
C	CUU Leu CUC Leu CUA Leu CUG Leu	CCU Pro CCC Pro CCA Pro CCG Pro	CAU His CAC His CAA Gln CAG Gln	CGU Arg CGC Arg CGA Arg CGG Arg
A	AUU Ile AUC Ile AUA Ile AUG Met	ACU Thr ACC Thr ACA Thr ACG Thr	AAU Asn AAC Asn AAA Lys AAG Lys	AGU Ser AGC Ser AGA Arg AGG Arg
G	GUU Val GUC Val GUA Val GUG Val	GCU Ala GCC Ala GCA Ala GCG Ala	GAU Asp GAC Asp GAA Glu GAG Glu	GGU Gly GGC Gly GGA Gly GGG Gly

7.016 Introductory Biology
Fall 2014