

Solution Key - 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 achieve your objective. **Explain** why you circled this option(s).

Wild- type

Strain 1

Strain 2

Strain 3

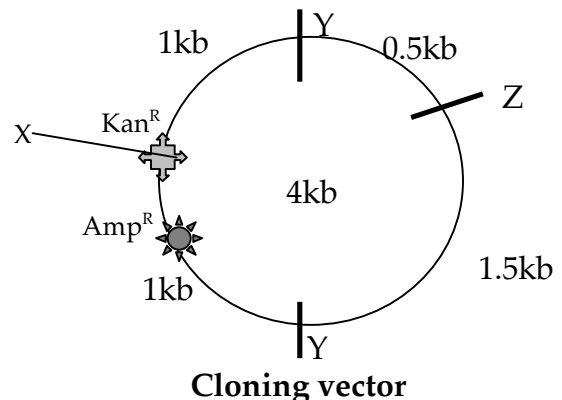
You can isolate the genomic DNA from any strain that has a wild- type copy of Gene 1. Thus you can use the genomic DNA from the the wild- type yeast cells. Based on the information provided you can also use the genomic DNA isolated from Strain 2 (that has mutation in Gene 2 only) and strain 3 (that has mutation in Gene 3 only) since they both have alleles of Gene 1 that encode the wild-type copy of E1 and should therefore be able to complement the mutation in Strain 1 (arg-) and restore its ability to produce arginine (arg+).

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.

Your plasmid should have both the bacterial and yeast origins of replication so that it can replicate in the bacterial and yeast cells. It should also have the recognition site(s) for restriction enzyme(s) so that the plasmid can be cut open and used as a vector to clone the yeast genomic DNA fragments that have either been digested with the same restriction enzyme(s) or a different one(s) that generates complementary ends. The plasmid should also have a selection marker such as the tetracycline resistance gene (*tet^r*) that can be used to distinguish the host cells that have been transformed with the recombinant plasmid that has the yeast genomic DNA insert, from the remaining host cell population. Note: There is no need for your plasmid to have a promoter since the gene by definition is the promoter + the transcribed region or in other words the gene comes with its own inherent promoter.

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.**

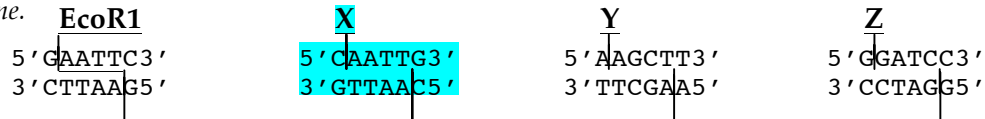
X	Y	Z	X + Y	X + Z	Y + Z
4kb	2kb	4kb	1kb	1.5kb	0.5kb
	2kb		2kb	2.5kb	1.5kb
			1kb		2kb



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

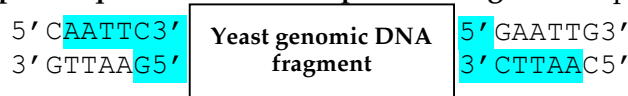
d) You then isolate the yeast genomic DNA and digest it with EcoR1 restriction enzyme. You want to clone the EcoR1 digested genomic DNA fragments into the plasmid vector that has the following recognition sites for restriction enzyme X, Y and Z. **Please note:** A vertical line (|) represents the cutting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



- 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?

You will select Enzyme X since it generates ends that are compatible to the ends generated by EcoR1.

- 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. You will first plate / grow the bacterial cells (*Kan^r* & *amp^r*) in minimal media that contains neither ampicillin nor kanamycin and let them form colonies (Plate 1). You will then replica plate the colonies from Plate 1 on plate 2 that contains minimal media with ampicillin. Only those bacterial cells that have been transformed either with the recombinant plasmid (that has the yeast genomic insert) or the self-ligated plasmid will grow and form colonies on Plate 2. You will then replica plate the colonies from Plate 2 (or Plate 1) on Plate 3 that contains minimal media with ampicillin and kanamycin. Only those bacterial cells that have been transformed with the self-ligated plasmid will grow and form colonies on Plate 3. So the colonies that grew on plate 2 but not on plate 3 will be the colonies of your interest.

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.

You isolate the recombinant plasmid from the transformed bacterial cells, transform the yeast Strain 1 with the recombinant plasmid and see if they grow and form colonies on the minimal media that lacks arginine. The recombinant plasmid that has a wild- type copy of Gene 1 can rescue the arginine producing ability of Strain 1.

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?

Unlike the eukaryotic cells the bacterial cells do not splice their mRNA. Since the cDNA is a copy of mature processed/ spliced mRNA (unlike the genomic DNA), the bacterial cells will have a higher chance of expressing it.

- 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.

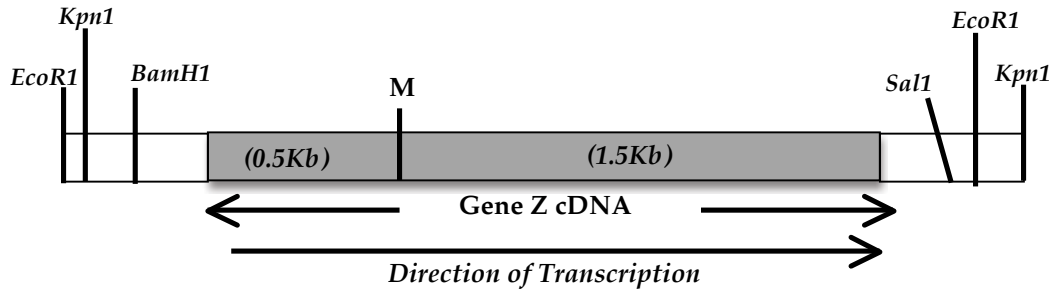
Your plasmid should also have a promoter in the right orientation to transcribe Gene 1 cDNA. It should also have the ribosome binding site so that the mRNA can be translated.

- 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.

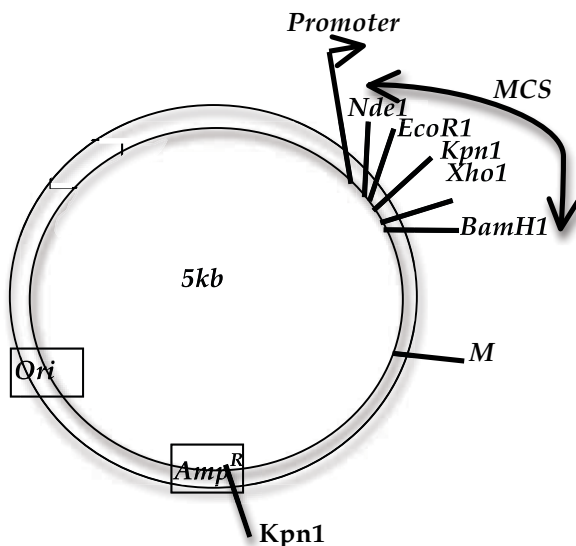
Most likely "No". There are multiple correct answers some of which are below. Bacterial cells may have different sets of arginine synthesizing enzymes compared to the eukaryotic yeasts. The post- translational modifications that are critical for enzyme activity may be different in prokaryotes and eukaryotes.

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	<i>EcoR1</i>	<i>EcoR1</i>
2 "✓"	<i>EcoR1</i> & <i>Sal1</i> ("✓")	<i>EcoR1</i> & <i>Xho1</i> ("✓")
3	<i>BamH1</i> & <i>Sal1</i>	<i>BamH1</i> & <i>Xho1</i>

b) Circle the strategy (s) in the table above, that would allow directional cloning of Gene Z cDNA. Strategies 2 & 3.

c) Put a "✓" next to the strategy in the table above, that would allow the expression of Gene Z cDNA. Strategy 1 will also work, but only 1/2 the time unlike Strategy 2.

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 <i>E. coli</i> cells	Transformed 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.

The E. coli cells are ampicillin sensitive.

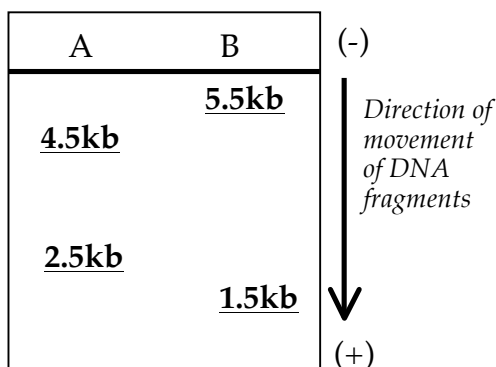
- ii. Although the *E. coli* cells in Row 3 of the table above have been transformed in the absence of ligase, you still observe a significant number of colonies. Why is this so?

The Endogenous bacterial ligase was able to ligate the plasmid at low frequency thus producing colonies both on plates.

- 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?

No, besides the bacterial clones that are transformed with the recombinant plasmid, you will always have some E. coli cells that would be transformed with the self- ligated plasmid (that has no Gene Z cDNA insert) and form colonies. You will also have some E. coli cells (although very few) that will spontaneously revert from amp^S to amp^R phenotype.

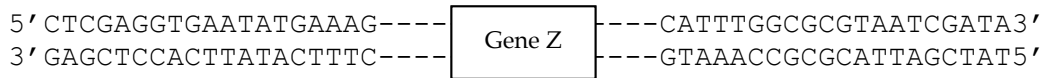
e) You next select two transformed colonies (Colony A & Colony B) that have the Gene Z cDNA insert. You realize that the gene could have been inserted into the plasmid in two different orientations. So you isolate the recombinant plasmids from two colonies, digest them with restriction enzyme M, resolve 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.



Colony A contains the recombinant plasmid that has the Gene Z cDNA inserted in the proper orientation with respect to the bacterial promoter for transcription unlike Colony B. Therefore, only the plasmids in colony A will express Gene Z cDNA.

Question 3

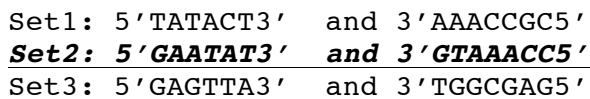
The following is the **wild- type allele of Gene Z** that you want to amplify using the polymerase chain reaction (PCR).



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.

You would need the **thermostable Taq DNA polymerase**, which can catalyze the DNA polymerization reaction and add the dNTPs to the 3' end of the primers, which anneal with the complementary bases in the template DNA strands. So you should have the pair of **primers, template strand and dNTPs** and buffer to carry out the PCR reaction.

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



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. 95°C for Step 1, 55°C for step 2 and 70°C for Step 3. Briefly **explain** what occurs at each of these three steps.

At 95°C the two strands of the DNA duplex unwind and separate from each other. DNA template melts or becomes single-stranded (denaturation). At 55°C the primers anneal to the template i.e. they undergo complementary base pairing with the respective template strands of the DNA (annealing). At 70°C the primers are extended by Taq DNA polymerase in the 5' to 3' direction (elongation).

d) Would all the PCR amplified fragments using the primers in part (b) be of the same size (Yes/ No)?
No, you will have fragments of variable lengths. The initial products will be larger compared to the fragments in the exponential phase of PCR.

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.



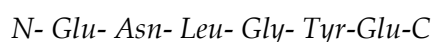
i. Write the DNA sequence for the region that corresponds to the amino acids 2-7 of the protein encoded by Gene Z and label the 5' and 3' ends of both strands.



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.

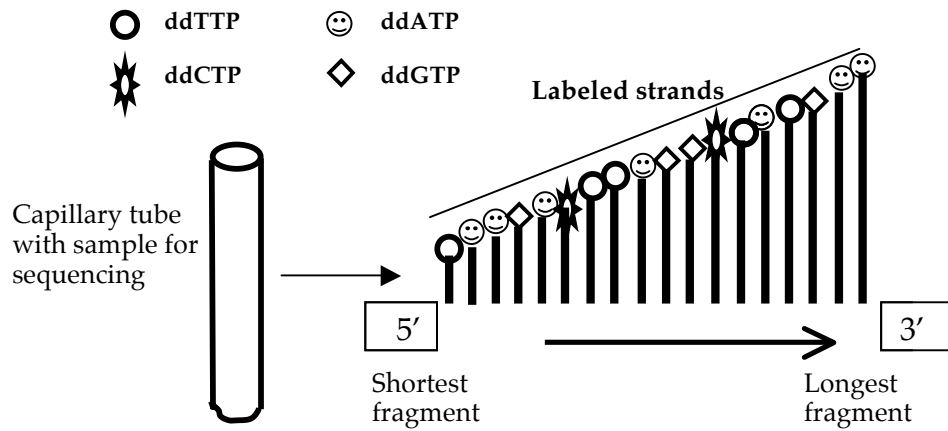


iii. Write the sequence of the amino acids 2-7 of the protein encoded by Gene Z and label its N and C termini. **Note:** A codon chart is given on the last page of this problem set.



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.

5' **T**AAGACTTAGGCTATGAA3' (coding strand)
 3' **A**TTCTGAATCCGATACTT5' (non-coding strand)

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	
Stop codon	Glu ²	Nonsense
Asp ³	Asn ³	Missense

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.**

The SNP can be in the coding sequence (i.e. exons) of the gene that codes for the amino acids of the protein. It can change the amino acid sequence if its presence results in a missense, nonsense or frameshift mutations. The SNP can also be in the non-coding regions (introns) of the gene and its presence may prevent the intron from being removed. So now the mature mRNA will be much longer since it has the introns within and will code for a longer protein sequence which will most likely be non-functional.

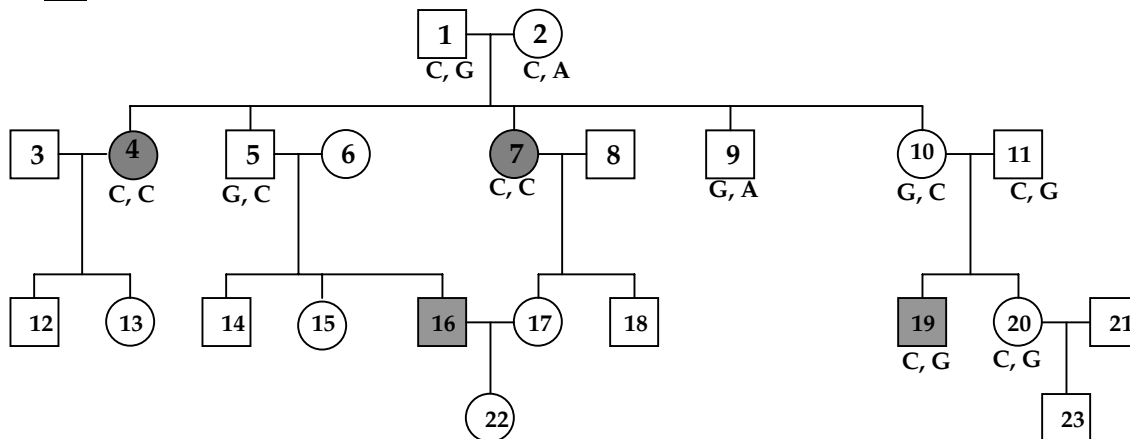
ii. It does not change the amino acid sequence of a protein? Include **all** the possible options and give an **explanation** for each selected option.

It may be in the non-coding region (i.e. promoter region) of the gene where it can influence the expression of the gene but not the sequence of the protein encoded by the gene. It can also be in the intergenic region, which will have no influence on the amino acids sequence of the protein.

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

The SNP is in the non-coding regions (introns) of the gene and its presence may prevent the intron from being removed. So now the mature mRNA will be much longer since it has the introns within and will code for a longer protein sequence which will most likely be non-functional.

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.
SNP C

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: Dd

Individual 5: Dd

Individual 9: DD

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.**

The genotype of individual #20 is DD since she inherits the "G" SNP from individual #10 and "C" SNP from individual #11. Since both individual #20 and 21 have the genotype "DD" their child has a 0% probability of being a carrier.

- 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.

DNA sequencing (as discussed in lectures), Restriction fragment length polymorphism (RFLP), microarrays; all are correct.

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.

Strategy 2 will produce a chimeric mouse. This mouse will have some cell-types that will be derived from the modified embryonic cells where one allele associated with the disease has been replaced by a wild-type copy of the Gene A through homologous recombination (genotype Aa). The same mouse will also have some other cell-types that are derived from the original embryonic cells (genotype aa) that did not receive a wild-type copy of Gene A. So the genotype of this chimeric mouse will be Aa in some cell-types and aa in other cell-types based on the embryonic cell types from which they originated through cell divisions.

- 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.

Aa or aa (or AA if both copies have been replaced by the introduced copy of Gene A). But No cell will have the genotype Aa

- 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.

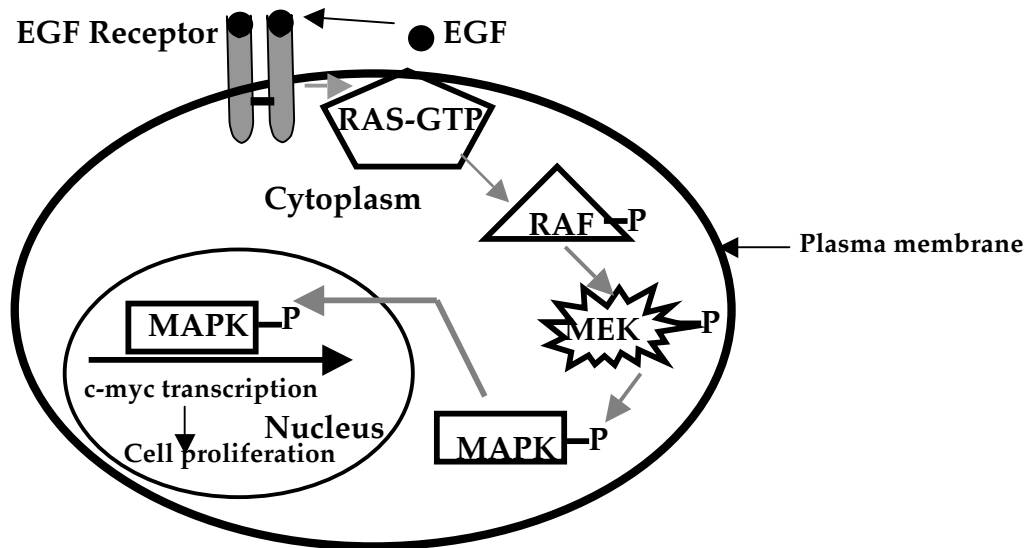
The gametes derived from the mouse obtained by strategy 2 can either have "a" or "A" genotypes. When they fuse with the gametes of a wild-type female (genotype "A") they will produce newborns that can either have AA or Aa genotype and will therefore be phenotypically normal.

- 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. *The virus has the reverse transcriptase that reverse transcribes DNA from RNA, which can integrate into the host genome.*

Question 6

Following is the schematic of a signal transduction pathway that is activated by the binding of an 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.

Ras will remain in its GTP bound active form irrespective of whether EGF is active or not. Hence Ras will activate the downstream components of the signaling cascade thus resulting in increased/ uncontrolled cell proliferation compared to the control cells.

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.

In the absence of the kinase domain MAPK will not be phosphorylated and hence will continue to remain in its inactive state. Thus you will observe decreased or no cell proliferation compared to control cells.

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.

In the absence of the signal sequence the protein will be translated and stay in the cytosol since the protein translation machinery will never be directed to the Rough endoplasmic reticulum (RER).

- 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.

The EGF receptor variant in this cell line has the signal sequence but lacks the hydrophobic / transmembrane domain sequence. Therefore, this variant of protein will be secreted into the medium. Thus the EGF receptor will not be available at the cell membrane to bind to EGF ligand and activate Ras-Raf-MEK-MAPK kinase pathway that is required for cell proliferation.

- 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.

The EGF receptor variant in this cell line lacks both the signal sequence and the hydrophobic / transmembrane domain sequence. Therefore this protein variant will remain localized in the cytosol and will eventually be degraded. Thus the EGF receptor will not be available at the cell membrane to bind to EGF ligand. Therefore the Ras-Raf-MEK-MAPK kinase pathway will not be stimulated, there will be no transcription of c-myc gene and hence there will be no cell proliferation.

MIT OpenCourseWare
<http://ocw.mit.edu>

7.016 Introductory Biology
Fall 2014

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.