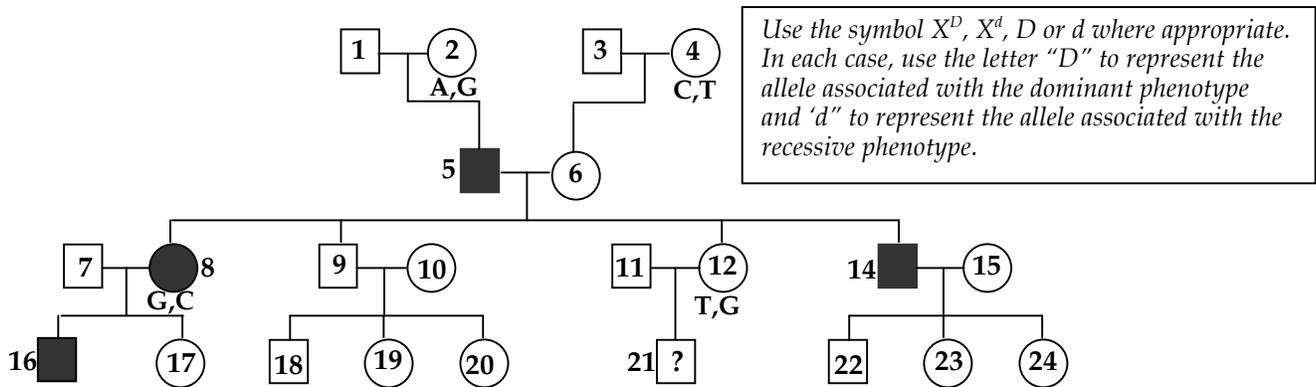


Solution key - 7.013 Problem Set 4- 2013

Question 1

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) Write **all possible genotypes** of the following individuals in the pedigree.

#6: $X^D X^d$

#19: $X^D X^D$

c) What is the probability of **Individual #21** being **affected**?

Individual 12 has the genotype $X^D X^d$ whereas the genotype of individual 11 is $X^D Y$. So their son, individual 21 has a 50% chance of being affected.

d) 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.

i. Identify the SNP(s) that is/are tightly linked with the mutant allele of Gene D.

There are two SNPs in this pedigree that are linked to the mutant allele of Gene D i.e. "G SNP" from individual 2 and "C SNP" from individual 4.

ii. Write the SNP genotypes of the following individuals.

#5: G

#14: C

e) One can use the SNP microarrays to determine the SNP genotype of an individual. Briefly describe how the SNP microarrays work.

DNA microarrays are small, solid supports onto which the oligonucleotides that represent the known SNPs in human genome are immobilized, or attached, at fixed locations. The supports themselves are usually glass microscope slides, but can also be silicon chips or nylon membranes. The DNA is printed, spotted, or actually synthesized directly onto the solid support. On the micro arrays the fluorescent-tagged genomic sample of interest is layered and allowed to hybridize with the oligonucleotides that have complementary sequence. The arrays are read through a laser detector to identify the SNP genotype of the individual test sample.

Question 1 continued

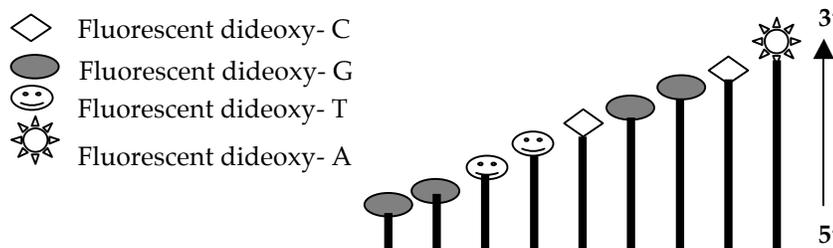
f) Although SNPs can be in the coding, non-coding or intergenic (regions between two genes) of the genes, you observe that this SNP is located within the noncoding region of Gene D. However, it still changes the sequence of the protein encoded by Gene D. Provide an **explanation** that supports this observation.

Most likely, this SNP changes either the splice donor or splice acceptor sites as a result of which the intron is not spliced out. Hence the amino acid sequence of the resulting protein is much longer than the normal variant of this protein.

g) You purify the wild- type and disease- associated forms of Protein D and determine their amino acid sequence. The only difference you find is at the **6th position** shown in **bold** below.

Wild- type variant: N-gly⁵-**trp**⁶-ala⁷-C
Mutant variant: N-gly⁵-**ser**⁶-ala⁷-C

Shown below is a portion of the fluorescence dideoxy- sequencing gel, which gives the sequence of the **non- template/ mRNA like strand** of the DNA that corresponds to amino acids 5-7 of the **mutant form of Protein D**.



Write the sequence of the double stranded DNA that corresponds to amino acids 5-7 of the **wild-type form of Protein D** and label its 5' and 3' ends. Note: A codon chart is provided on the last page of the problem set.

5' GGT TGG GCA 3'
 3' CCA ACC CGT 5'

h) The following is the DNA sequence of the **wild type allele of Gene D** that you want to amplify using the polymerase chain reaction (PCR).

5' CTCGAGGTGAATATGAAAG-----Gene D-----CATT TGGCGCGTAATCGATA 3'
 3' GAGCTCCACTTATACTTTC-----G TAAACCGCGCATTAGCTAT 5'

- i. 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. You should also have the pair of primers and dNTPs for reasons described above.

- ii. Give the sequence (10 bases long) of a **set of primers**, which you would use for the PCR reaction. 5'CTCGAGGTGAATAT3' (as primer that binds 5'->3' to the bottom strand) and 5'TATCGATTACGCAAAT3'(as primer that binds 5'->3' to the top strand).

- iii. 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. 94°C for Step 1, 60°C for step 2 and 72°C for Step 3. Briefly **explain** why the three steps are performed under different temperatures.

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).

Question 1 continued

- iv. When PCR amplification is used to provide a specific DNA fragment for cloning, the resulting clones are sequenced to make sure that inserts have the correct base sequence. What activity does Taq DNA polymerase enzyme lack that may explain the errors that occur during PCR amplification?

Taq DNA polymerase lacks the 3' → 5' exonuclease activity and can therefore not proofread the newly synthesized DNA strand for any incorrectly incorporated nucleotide base.

Question 2

You need a large amount of Protein D encoded by Gene D. Therefore you decide to engineer a mouse cell line that will secrete a large amount of Protein D, so that you can purify Protein D from the medium. **Note: A cell line is a single type of cell, which continuously grows in culture.**

- a) List **any four components**, of the host eukaryotic cell translation machinery, which are **absolutely** required for synthesis of proteins, and briefly (few words) indicate what each does.

Many answers can be accepted here including: Ribosomes, amino acids, tRNAs, mRNA template.

Ribosomes: They serve as the machinery for protein synthesis. They bind onto the ribosome-binding site (RBS) of the **mRNA template** and start moving along the length of mRNA template looking for the start codon (5' AUG3'), which defines the reading frame of the mRNA. From here on the **tRNA**, each charged with a specific **amino acid** binds to the codons of the mRNA template and also to the ribosomes and starts adding the amino acid to the growing peptide chain. The process continued till the translation machinery reaches the stop codon, at which point it falls apart and the newly synthesized protein is released.

- b) List **two components**, of the host eukaryotic cell translation machinery, which are **absolutely** required for **export of Protein D from the cell?**

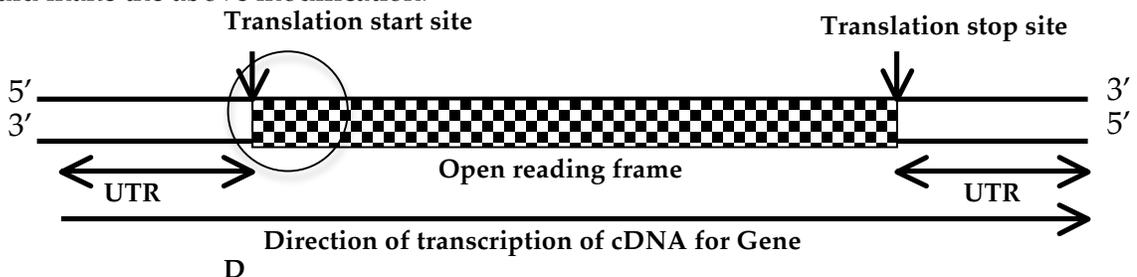
Many answers can be accepted here including: Signal recognition peptide (SRP), SRP receptor on Endoplasmic reticulum (ER), translocon, ER, golgi, ribosomes, vesicles, etc.

- c) You isolate a cDNA for Gene D. However, when you transfect the normal cell line with a vector containing Gene D (*i.e. insert a vector containing Gene D into the cells*), you find that Protein D is produced in the cytoplasm but not secreted. You conclude that something is therefore wrong with the Gene D.

- i. What modifications would you make to the cDNA for Gene D that would plausibly allow Protein D to be secreted?

Include DNA that encodes a signal sequence near the 3' end of the template DNA strand that, when transcribed, will be on the 5' end of the mRNA and ultimately on the N-terminus of the translated protein. This signal sequence will allow the protein to bind to the SRP. At this point the protein synthesis will temporarily halt. The signal sequence –SRP complex will then be recognized by the SRP receptor/translocon. The signal sequence –SRP complex will then pass through the SRP receptor to reach the ER lumen. The SRP will then detach and the signal sequence will be cleaved and protein synthesis will resume. Once synthesized in the ER lumen, the protein will be packaged in vesicles that will be transported to golgi for further modifications (adding carbohydrates, lipids or phosphate groups) and then to the cell membrane. The vesicles will then fuse with the plasma membrane and protein inside them will be secreted in the surrounding environment.

- ii. The following is a schematic of the cDNA for Gene D. **Circle** the region in Gene D, where you would make the above modification.

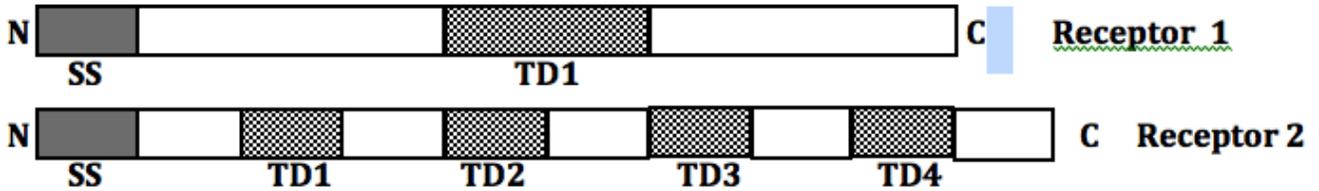


Question 2 continued

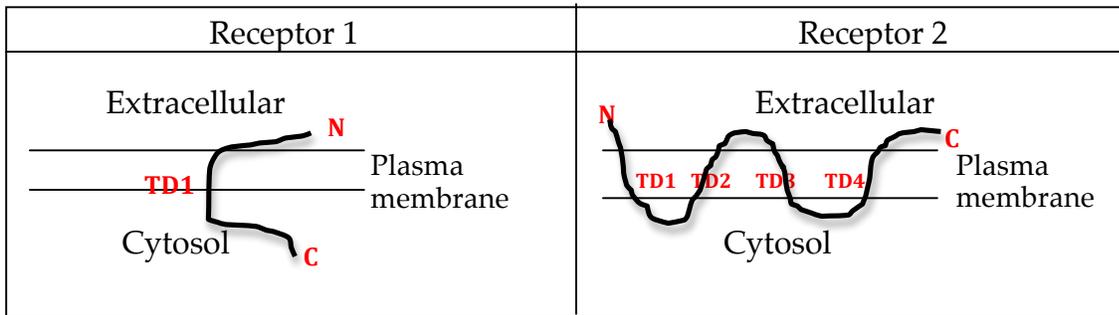
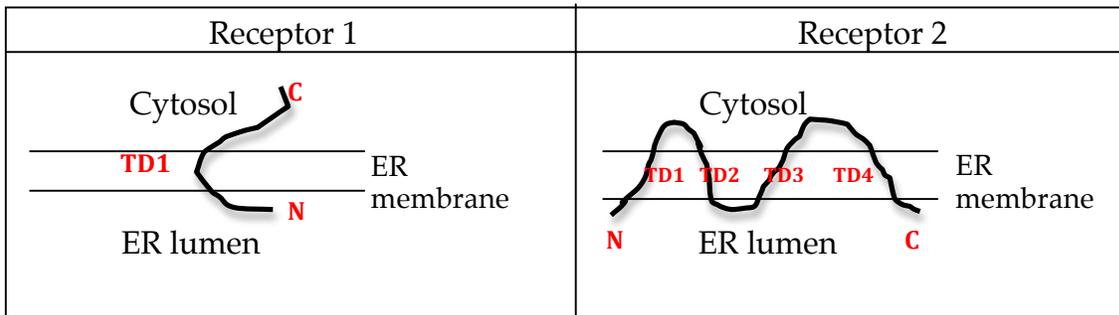
- iii. Assuming your manipulation in part (i) is successful, would the modified/secreted version of Protein D be of the same size, larger than, smaller than its unmodified/cytoplasmic version? Explain your answer.

The nascent/primary protein made from the modified gene would be longer by the amount of the signal sequence. However, signal sequences are removed in the ER so the secreted version of the protein would be the same.

d) Interestingly, you find that this protein, when secreted, can bind to either of following cell membrane receptors (**Receptor 1 and Receptor 2**). The transmembrane domains (TD) and the signal sequence (SS) are shown in the schematic.



Draw the two receptors as they would be inserted into the Endoplasmic reticulum (ER) and plasma membrane, label the N and the C termini and include all the TD domains that are shown in the schematic above.



Question 3

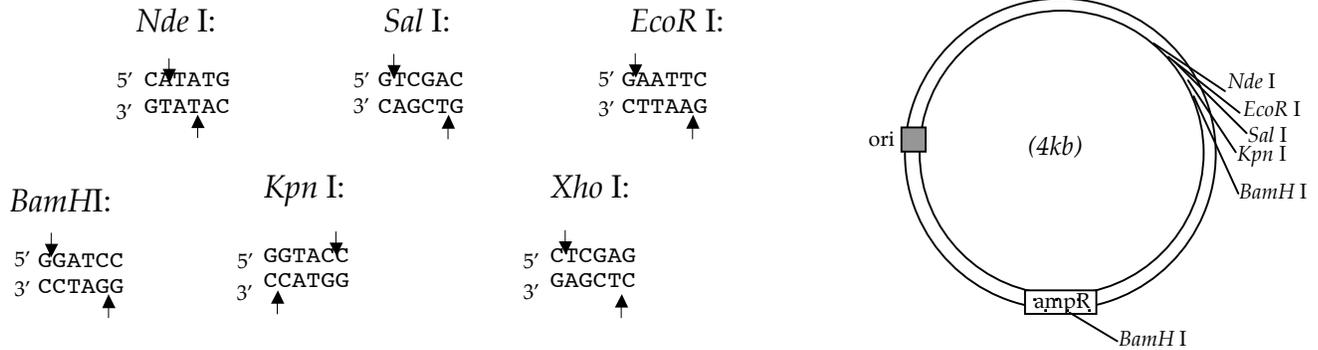
You have identified an enzyme (E1) in a yeast strain (Strain 1) that catalyzes a step in a biochemical pathway, which results in the synthesis of the amino acid Arginine. This enzyme is encoded by Gene A. You isolate Gene A from the yeast cells, clone it into a plasmid vector and **amplify** it in the bacterial cells.

- a) List the minimum features that a plasmid vector **must** have to allow the cloning and amplification of Gene A in bacterial cells.

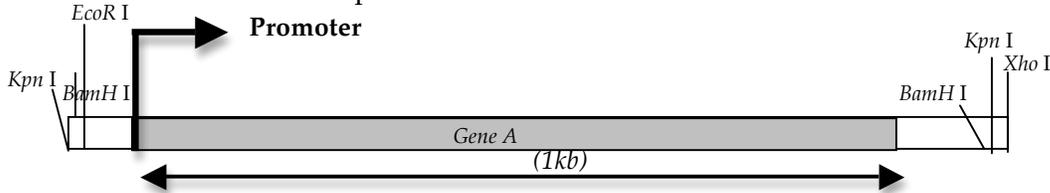
Origin site (Ori site) for replication, a site that can serve as the recognition sequence for restriction enzyme so that the plasmid can be cut open and used as a vector to clone the desired sequence and a reporter gene (i.e. antibiotic resistant gene) that can be used to differentiate between the untransformed host cells and the host cells that obtained a plasmid.

Question 3 continued

b) You decide to use the following plasmid vector to clone Gene A. The recognition sequence for each restriction enzyme is given below. An arrow represents the site at which the restriction enzyme cuts.



i. A schematic of Gene A is given below. You want to clone Gene A into the plasmid vector. Give three different strategies that you could use to clone Gene A into the vector, and obtain colonies that contain a recombinant plasmid.



Strategy	Restriction enzyme(s) used to cut....	
	Gene A	Plasmid vector
1	Kpn1	Kpn1
2	EcoR1 & Xho 1	EcoR1 & Sal1
3	EcoR1 & kpn1	EcoR1 & kpn1

ii. Which strategies (Choose from Strategy 1, 2, 3) would allow a **directional cloning**? 2 & 3

c) You then transform the bacterial cells with the ligation mix.

i. Briefly outline a procedure that would allow you to distinguish the **transformed** bacterial cells from the **untransformed** ones.

You plate the ligation mix on a master plate that has no ampicillin and let the bacterial cells form colonies. ALL bacterial cells, both transformed and untransformed will grow and form colonies on the master plate. Then you replica-plate these bacterial colonies on a plate that contains ampicillin. Only the bacterial cells that have been transformed with the plasmids will survive and form colonies since they will have the amp^R gene and can make the enzyme that can degrade ampicillin. In comparison, the untransformed cells in the absence of this enzyme will not survive and hence will NOT form any colonies following replica plating.

ii. Briefly outline how the DNA gel electrophoresis would allow you to distinguish between the bacterial cells transformed with the **self-ligated plasmid** from those transformed with the **recombinant plasmid**.

You will isolate the plasmids from the bacterial colonies. The self-ligated plasmid (4 kb) and the recombinant plasmid with cloned Gene A (5kb) will be of different sizes and hence they will resolve differently when subjected to DNA gel electrophoresis.

Question 3 continued

d) You grow and amplify the recombinant plasmid in the bacterial cells. You then purify the recombinant plasmid from the bacterial cells and transform the yeast cells with it.

- i. Why did you use the bacterial cells only for the amplification of the plasmid but **NOT** for the expression of Gene A?

Bacteria, being prokaryotes will not be able to splice out the introns and make a functional protein. You can also argue that if this eukaryotic protein requires further post-translational modifications the bacterial cells may not be able to do so i.e. the process is different in prokaryotes and eukaryotes.

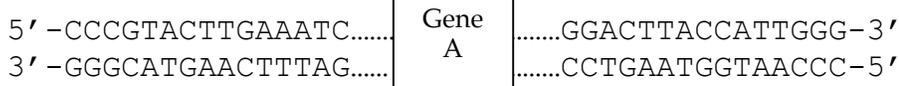
- ii. Give an **alteration** in your experiment that would have allowed you to both amplify and express Gene A in the bacterial cells.

Instead of cloning Gene A, you may clone the cDNA corresponding to Gene A. The cDNA will be made by using mature mRNA corresponding to Gene A and therefore will have ONLY exons but NO introns and NO regulatory region. So as long as the cDNA is inserting in the plasmid in the correct orientation with respect to the promoter, the bacterial cells translation machinery can be used to translate it into the desired protein.

- iii. What media would you use to select the yeast cells that are expressing the enzyme encoded by Gene A?

You should plate the transformed yeast cells in the minimal media that lacks arginine (arg- plate). Only the cells that have been transformed with Gene A will be able to make the enzyme needed for Arg biosynthesis. These cells will form colonies on arg- plate.

e) Your friend also does the same experiment to clone Gene A with the following flanking sequence. But he decides to select the bacterial cells transformed with the recombinant plasmid using colony hybridization.



- i. Give the sequence of the probe that he would use for colony hybridization and label the 5' and the 3' ends.

Your friend can either use 5'-CCCGTACTTGAAATC-3', 5' GGACTTACCATTGGG-3' , 3' GGGCATGAACTTTAG5 or 3'-CCTGAATGGTAACCC-5' sequence, radio label it and use it as a probe.

- ii. Briefly describe the major steps of colony hybridization that would allow your friend to select the transformed bacterial colony.

You will take an imprint of the bacterial colonies on a nitrocellulose paper / nylon membrane and disrupt the cells leaving behind the genomic DNA on the nitrocellulose paper / nylon membrane. You will add the P³² labeled oligonucleotide probe onto the paper and let it hybridize with the complementary genomic DNA. You put an X ray film on the nitrocellulose paper and take an imprint of the bacterial colony that gets radiolabeled and hence lights up on the X ray film. This colony contains the plasmid that contains the gene of your interest.

Question 4

Individuals who are homozygous for the mutation in Gene A (genotype = aa) suffer from a liver disease that shows an autosomal recessive mode of inheritance. You decide to do a UROP in a lab that is dedicated to using gene therapy to cure this hereditary disorder, in a mouse model.

a) Using **virus** as a vector that stably integrates into the genome of the host cell, you introduce a wild-type allele of Gene A in the liver cells of newborn mice whose genotype is "aa". If all cell types of the mouse have roughly the same genome why would you use only the liver cells to introduce the wild-type allele of Gene A?

Although ALL cells will have almost the same genome, different cell- types will express different sets of genes and hence will have different proteins. Gene A is expressed only in Liver cells and hence you use these cell-types in your experiment.

Question 4 continued

b) Based on what you have learnt in 7.013 lectures why is the **ex-vivo** gene therapy more successful than the **in- vivo** gene therapy? Give one reason.

Ex-vivo gene therapy gives you the ability to introduce a gene and screen for the transfected cells prior to introducing them back into the patient. So it is generally more successful than in- vivo gene therapy. But still there are multiple issues:

- 1. Efficiency of infection: The virus that carries the cDNA of the interest does not transfect sufficient number of cells to improve the symptoms of the disease.*
- 2. The cDNA even if introduced may not be efficiently expressed or may not stay expressed for a long time to provide relief from symptoms for a long time.*
- 3. The recipient immune system may elicit an immune response to the virus. Or the virus may itself prove toxic to the recipient.*

c) Your supervisor suggests that you use embryonic cells instead of liver cells for ex-vivo gene therapy. **Explain** what may be the potential advantage of using embryonic cells compared to the liver cells.

(NOT GRADED)

Embryonic cells are pluripotent and can therefore acquire different cell fates. They can also serve as the precursor for the germ-line cells. If this is the case, then the defect can be cured not only in the individual produced from the developing embryo but also in his descendants. Also, since the embryonic cells are undifferentiated cells they may better adopt to expressing new proteins.

d) You isolate embryonic cells from an affected mouse (genotype: aa) and infect them with a viral vector that has a wild- type allele of Gene A. You select the cells that have been infected with the virus containing the wild- type copy of Gene A and re-introduce them into the developing embryo of the affected mouse to obtain newborns. You trace the location and expression of Gene A in the newborn mice by adding a blue color dye that specifically binds to the protein encoded by Gene A. You obtain the following two sets of results. **Note:** *You may assume that the level of expression of the Gene A correlates with the intensity of the blue color in the cells. In wild-type mice, the dye stains only the liver cells.*

- **Set 1:** *When you add the dye you find that most of cells in the mouse, including the liver cells, turn blue.*
- **Set 2:** *Based on the results of Set 1, you modify your viral vector that contains Gene A, reinsert it into the embryonic cells and obtain newborns by following the same steps that were described above. When you add the dye you find that only the liver cells turn blue and the color is of the same intensity as in the liver cells of wild-type mice. In addition, these mice do not show the manifestations of the disease.*

How did you modify the viral vector so that the introduced Gene A was expressed only in the liver cells of newborns obtained in Set 2?

The introduced gene in Set 1 is under the regulation of a generic promoter. In contrast, the introduced gene in Set 2 is under the regulation of a tissue specific promoter.

e) 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 infect these cells with a modified vector from Set 2 that has a wild type copy of Gene A. 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. **(NOT GRADED)**

- i. Give **all** the possible genotypes of the newborn obtained from this strategy? Briefly **explain** why you selected this genotype. **Note:** *Use the uppercase A to represent the allele responsible for the dominant phenotype and lowercase a to represent the allele responsible for the recessive phenotype.*

This strategy 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.

Question 4 continued

- ii. You allow the newborn obtained from the strategy outlined above to mate with a wild- type female mouse (genotype: AA). Would you expect all the mice from this mating experiment to have a normal phenotype (Yes/ No)? **Explain** why you selected this option.

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.

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

7.013 Introductory Biology
Spring 2013

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