7.013 Recitation 10 - 2013

Summary of Lectures 15 & 16:

Library: A **genomic library** is a population of host bacteria, each of which carries a DNA molecule that was inserted into a cloning vector, such that the collection of cloned DNA molecules represents the entire genome of the source organism. The DNA molecules of an organism of interest are isolated, partially digested by an endonuclease restriction enzyme and ligated into vectors. The vectors can then be taken up by suitable hosts. The hosts are kept in liquid media and can be frozen at -80°C for a long period of time. Usually the hosts are bacteria that do not contain any plasmids, so as to be sensitive to antibiotics. A **cDNA library** is a collection of cloned complementary DNA (cDNA) fragments inserted into a collection of host cells. A cDNA is produced from fully transcribed mRNA found in the nucleus and therefore contains only the expressed genes of an organism. In eukaryotic cells the mature mRNA is already spliced, hence the cDNA produced lacks introns and can be readily expressed in a bacterial cell. While information in cDNA libraries is a powerful and useful tool since gene products are easily identified, the libraries lack information about introns, and other regulatory elements found in a genomic DNA library.

PCR: This is the technique of Polymerase Chain Reaction, which is used to make lots of copies of a chosen piece of DNA. PCR is simply the process of doing about 30 rounds of DNA replication in a row. This technique requires you to combine in a test tube: the original DNA, nucleotides, thermostable DNA polymerase, and primers. Then, one puts the reaction in a machine that performs 30 cycles in a row of three temperatures. The first temperature denatures the double-stranded DNA molecule, the second allows the primers to base-pair with the template, and the third allows the DNA polymerase to generate the newly synthesized strand 5'→ 3' starting from the 3'OH of the primer.

DNA sequencing: DNA sequencing is the technique by which the nucleotide sequence of DNA is determined. DNA sequencing is simply the process of doing DNA replication in the presence of all four normal nucleotides, and a small percentage of some "bad" nucleotides that do not allow for replication to proceed any farther. This technique simply requires you to combine in a test tube: the original template DNA, the four good nucleotides, DNA polymerase, a primer, and then a small concentration of "bad A" (colored red), "bad T" (colored blue), "bad C" (colored yellow), and "bad G" (colored green). Most of the time, the good nucleotides will be added and the template will be extended normally. Occasionally, a bad C (for example) will be put in across from a G in the template. This will stop replication and label the fragment yellow. You allow the reaction to proceed and then load all of the replication products onto a gel that separates them by size. The sequence can then be read from the result of that gel.

The way that genomes are sequenced is that many copies of the genome are randomly broken into pieces, and then those pieces are sequenced by sequencing machines. A computer then analyzes those sequences and reassembles them into one long continuous piece (for a bacterium with a single circular chromosome) or several linear pieces (for a eukaryote with multiple chromosomes). This technique is called shotgun sequencing. Once the genome is sequenced, one can predict where the genes are in the genome by looking for the signatures of genes, like start codons and stop codons, long open reading frames, promoters, and splice sites (if applicable). Each gene in the genome can then be compared to all other sequences that have ever been determined using the program BLAST. If a new gene shows a large amount of homology to a previously studied gene, the most likely possibility is that the new gene encodes a protein with a similar function to the protein encoded by the previously studied gene.

Microarrays: Not all cancers respond equally well to all treatments, so knowing the specific type or subtype of cancer is important to successful treatment. Some cancers may be identified using histopathology, but other are best characterized by determining the gene expression profile. A DNA microarray is a multiplex technology that allows comparison of cells based upon the expression of many different genes. A DNA microarray consists of an arrayed series of thousands of microscopic spots of DNA, each spot representing a gene. When a DNA microarray is probed with mRNA isolated from cells, individual mRNA molecules will hybridize to the appropriate DNA spot. If a gene is highly expressed, more mRNA will be made from that gene, so more mRNA will hybridize to the corresponding DNA spot, and the signal from that spot will be greater.

Gene therapy: Gene therapy is the insertion of genes (using an appropriate vector i.e. retrovirus, adenovirus etc) into an individual's cells and tissues to treat a disease, such as a hereditary disease in which a deleterious mutant allele is replaced with a functional one. Although the technology is still in its infancy, gene therapy has been used with some success. One can do **somatic cell gene therapy**, which can provide some relieve to the patient if successful or germ-line/ reproductive gene therapy, which can cure the disease for subsequent generations. The technique is still in its infancy. We need to enhance the efficiency of delivery. Also there are risks i.e. for example the SCID patients who received gene therapy later on developed leukemia.

One can also alter the genetic make-up of an animal either by putting in the genes (**knock in organisms**) or by knocking of the genes from them (**knock out organisms**).
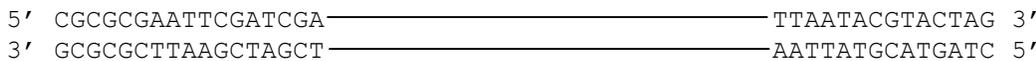
In order to create a **chimera** you take the embryonic stem cells (ES cells) from a 3 -5 days old embryo. These cells are pluripotent and can differentiate into any cell type of the body. You can then put these cells into a recipient embryo. The recipient embryo can then be implanted in a pseudo-pregnant female. The newborn produced is a chimera since it has cells of two different genetic makeup.

Questions:

1. How is a cDNA library different from a genomic library?

2. 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 should bind in this region   Primer 2 should bind in this region

```
5' CGCGCGAATTCGATCGA————————————————————TTAATACGTACTAG 3'
3' GCGCGCTTAAGCTAGCT————————————————————AATTATGCATGATC 5'
```

          Primer 1   Primer 2

    Set A: 5' TCGATCGAATTC 3'   AND   5' TAATACGTACTA 3'

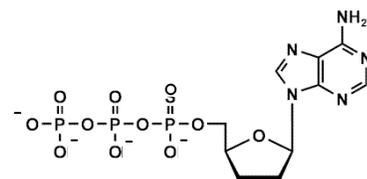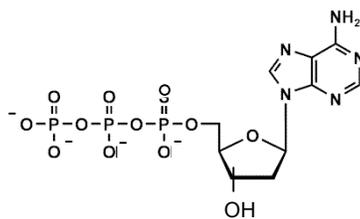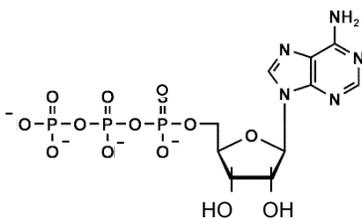   Set B: 5' GCTTAAGCTAGC 3'   AND   5' GATCATGCATAA 3'

    Set C: 5' GAATTCGATCGA 3'   AND   5' CTAGTACGTATT 3'

3. Take a look at the molecules drawn below
   #1   #2   #3

a) Which of the above molecules is non-physiological and is only used in DNA sequencing?

b) Which one of the above molecules is used in RNA? ?

c) Which one of the above molecules is used in DNA?

d) Which one of the above molecules is used as the major source of energy in cells?

4. You decide to sequence the following piece of DNA using the following primer:

```
5'-…GTATAGCCTCCGT…-3'
3'-…CATATCGGAGGCA…-5'
```
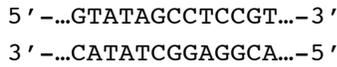
The primer you use is 5'-ACG-3'   (Note: real primers are usually 18-22 nts long). You set up a DNA sequencing reaction with this DNA, this primer, DNA polymerase, dATP, dTTP, dGTP, dCTP, and ddATP.

i. How many different products would form from this reaction?

ii. How many nucleotides long would each different product of this reaction be?
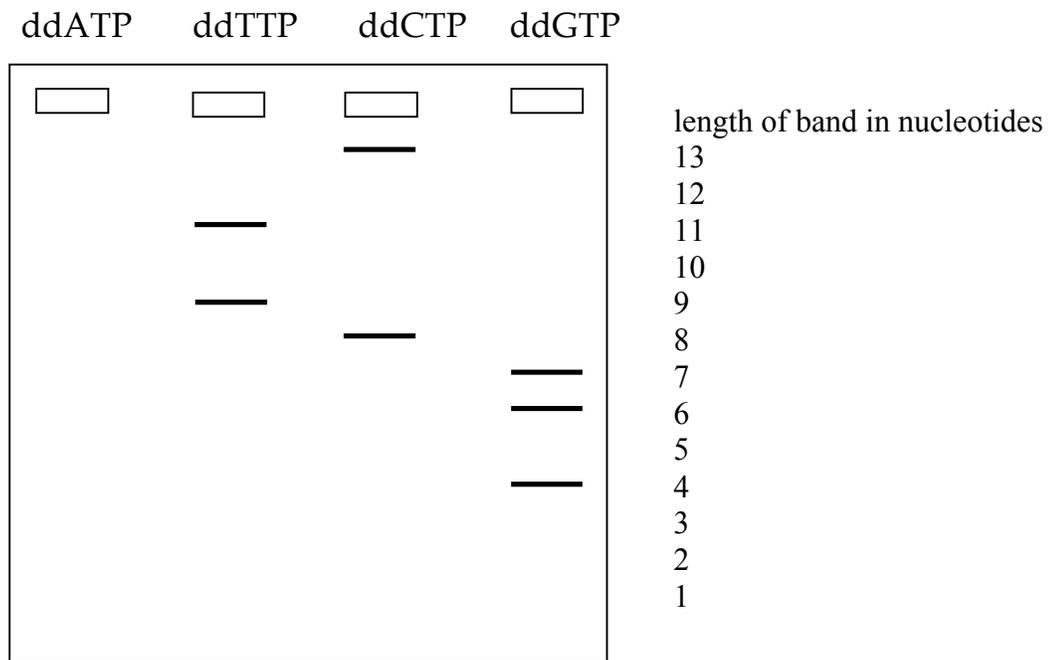
iii. Write the sequence of DNA that you read from this gel below. Label the 5' and 3' ends of the molecule.

iv. Would you be able to tell the sequence of the molecule if you had loaded into a single lane a reaction in which all four ddNTPs had been added from a lane?

5. You identify a hypothetical Gene R in humans that encodes for protein "R" which is involved in maintaining low blood cholesterol level. This gene shows an autosomal dominant mode of inheritance and the affected individuals are at a higher risk of developing a cardiac disorder (CD). You come across a SNP (Single nucleotide polymorphism) that is tightly linked to Gene R. You decide to use this SNP as a marker for CD. The two alleles of SNP (C and G) are shown for each individual in the following pedigree. *Individuals affected by CD are shaded in black.*

1   2

C/G   C/G

G/C   C/C   G/C   G/C   G/C   G/G   G/C   C/C

3   4

C/C   C/C   G/C

?

C/C
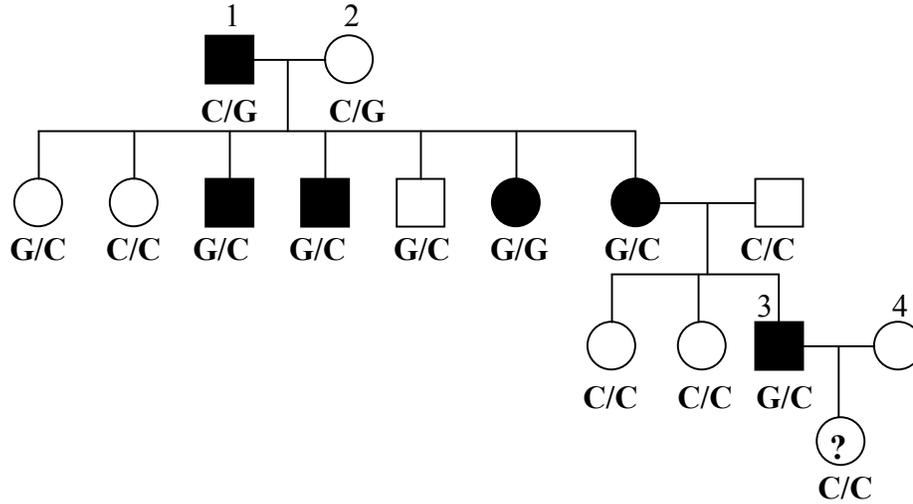
i.   Assuming no recombination between this SNP and Gene R, which allele of the SNP is linked to the disease associated **R** allele in individual #1?

ii. Assuming no recombination between this SNP and Gene R, what is the probability that the offspring of individual #3 and individual #4 has the disease?

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.