

Recombinant DNA II

Making, screening and analyzing
cDNA clones
Genomic DNA clones

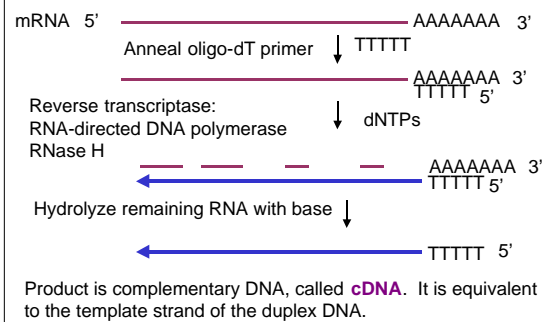
cDNA clones are copies of mRNAs

- Much of the genomic DNA is **not** expressed as mRNA
- Many issues about gene function are best addressed by examining the product that they encode.
- The cDNA copies of mRNA contain primarily sequences that encode protein.
- Therefore, cDNA clones are useful for many studies of gene function.

Construction of cDNA clones

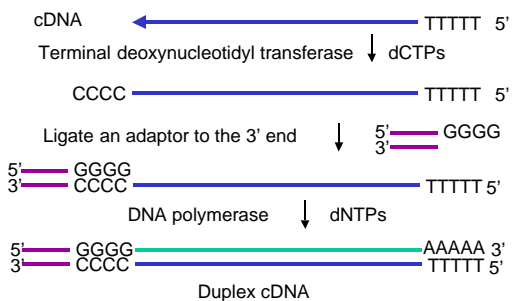
- Use the enzyme *reverse transcriptase* to copy mRNA into complementary DNA, called cDNA. This is equivalent to the template strand of the duplex DNA.
- Use a DNA polymerase to copy that cDNA into the nontemplate (message synonymous) strand.
- Insert the duplex cDNA product into a cloning vector and propagate in a host, e.g. *E. coli*.

cDNA: first strand synthesis

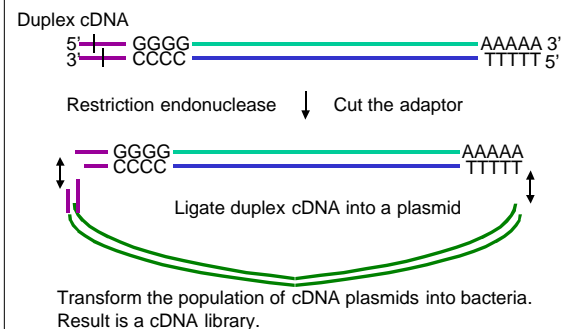


cDNA: second strand synthesis

Problem: How to get a primer for 2nd strand synthesis?



Ligate duplex cDNA into a plasmid



Limitations of cDNA synthesis

- First strand synthesis often does not go to completion.
 - Individual cDNA clones will frequently have the reverse complement of only part of the mRNA.
 - Multiple cDNA clones from a single mRNA will be present in the library
- Priming second strand synthesis is inefficient
 - Some methods necessarily result in the loss of sequences at the 5' end of the nontemplate strand

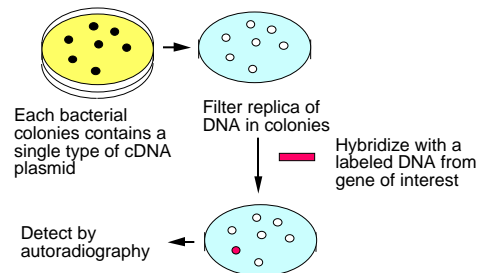
How do you find a cDNA clone from the desired gene?

- A cDNA library has >100,000 individual clones.
- It contains copies of as many as 50,000 different mRNAs .
- The frequency of occurrence of a cDNA from a given gene reflects the abundance of the mRNA for that gene.
- Try to find correct 1 clone in about 100,000.

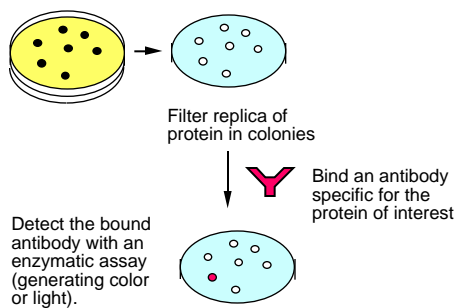
Strategies for screening cDNA clones

- Brute force screen for abundant cDNAs.
- Hybridization with a gene-specific probe.
- Express the cDNA in the host cell (i.e. make a functional protein product)
 - Specific antisera
 - Labeled ligand to a receptor
 - Assay for a function (complementation)
- Differential analysis

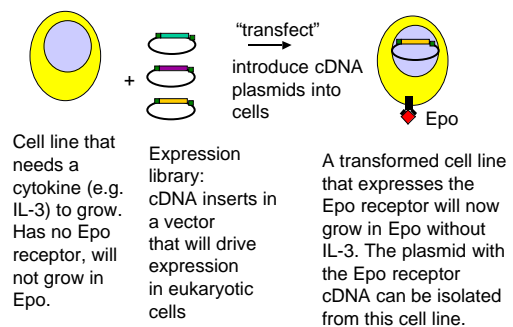
Screening by hybridization



Screening for an expressed product



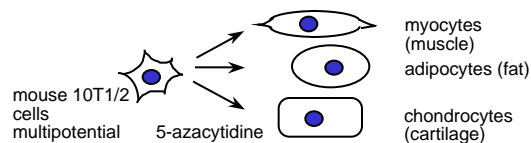
Expression screening in eukaryotic cells



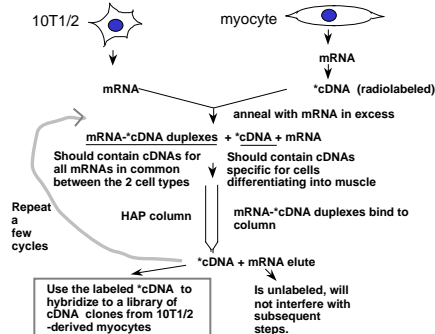
Differential analysis

- Instead of looking for one particular cDNA, look for cDNAs from **all** genes whose expression differs in the process under study
 - Differentiation from mesoderm to muscle
 - Response to different nutrients
 - Progression through S phase of the cell cycle
- Methods:
 - Subtractive hybridization
 - Differential display
 - Hybridization to massively parallel arrays of cDNAs.

Differential analysis applied to muscle differentiation



Subtractive hybridization



Differential display of RT-PCR products

- Make cDNA from all mRNA in the two different cellular states (RT = reverse transcriptase).
- Use several sets of PCR primers to amplify a representative sample of all the cDNAs.
- Resolve those RT-PCR products on a gel.
- Find the products that are *present in only one* of the two cellular states being compared.
- Try to isolate the corresponding gene.

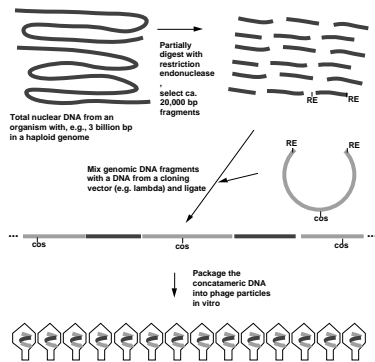
Sequence everything, find function later

- Determine the sequence of hundreds of thousands of cDNA clones from libraries constructed from many different tissues and stages of development of organism of interest.
- Initially, the sequences are partials, and are referred to as expressed sequence tags (ESTs).
- Use these cDNAs in high-throughput screening and testing, e.g. expression microarrays (next presentation).

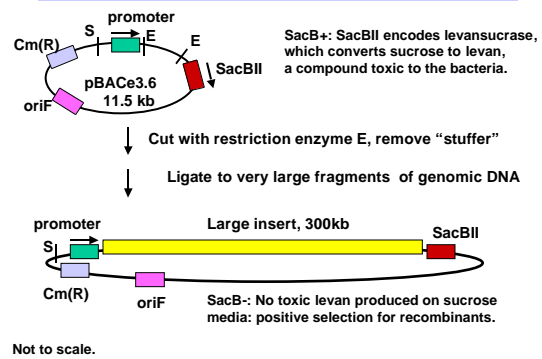
Genomic DNA clones

- Clones of **genomic DNA** contain fragments of chromosomal DNA. They are used to:
 - obtain detailed structures of genes
 - identify regulatory regions
 - map and analyze alterations to the genome, e.g. isolate genes that when mutated cause a hereditary disease
 - direct alterations in the genome
 - sequence the genome.

Construction of libraries of genomic DNA



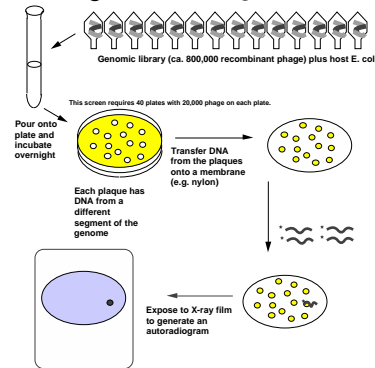
BAC vectors for large DNA inserts



How many clones make a representative library?

- P = probability that a gene is in a library
- f = fraction of the genome in a single recombinant
- f = insert size/genome size
- For N recombinants, $1-P = (1-f)^{expN}$
- $\ln(1-P) = N \ln(1-f)$
- $N = \ln(1-P) / \ln(1-f)$
- For a lambda library with an average insert size of 17 kb and a genome size of 3 billion bp, then one needs a library of 800,000 clones to have a probability of 0.99 of having all genes in the library.
- For a BAC library, with an average insert size of 300 kb and a genome size of 3 billion bp, then the library size required for $P=0.99$ is reduced to about 46,000 clones.

Screening libraries of genomic clones



Sequence everything: genomics

- Instead of screening for one gene at a time, an entire genome can be sequenced, and one can use experimental and bioinformatic approaches to find many (all?) genes of interest.
- Made possible by
 - Substantial increases in speed of sequencing
 - Larger insert libraries for larger genomes
 - Combination of hierarchical sequencing (based on maps) and whole genome shotgun sequencing