Lecture 9 - Screening cDNA libraries (AMG text pp. 128-134)
September 20, 2000

Most cDNA libraries are representative libraries having a sequence complexity that reflects the starting mRNA sample. However, for functional screening strategies, a normalized cDNA library is preferable because rare sequences are present at the same concentration as abundant sequences. Normalized cDNA libraries are constructed by removing abundant cDNA sequences (e.g., actin) using solution hybridization methods.

cDNA libraries are commonly screened by either of two different approaches:

1. Isolation of clones based on cDNA sequence alone which reflects gene structure.
2. Isolation of clones based on a measurable function of the encoded protein.

Typically, the cloning vector determines which type of screen can be performed. If DNA sequence is the only criteria for identification then most any lambda vector is sufficient, however, if function of the encoded protein is required (or antigenicity), then the vector needs to contain transcriptional and translational control elements.

Three methods are commonly used to screen cDNA libraries by sequence strategies;

1. Using the protein sequence to design degenerate oligonucleotide probes.
2. Using purified proteins as antigens to generate antibodies to detect antigenic proteins.

Degenerate Oligonucleotide Probes
The amino acid sequence of polypeptide fragments of a highly purified protein can usually be determined by automated protein microsequencing based on the Edman degradation procedure or mass spectrometry. Peptide sequences can be used to predict a nucleic acid sequence based on triplet codons in the Genetic Code. Custom oligonucleotide synthesis can then be used to generate a nucleic acid probe for library screening.

Two polypeptide sequences from the same protein can be used to generate separate mixed oligonucleotide pools for use as probes for cDNA library screening. By initially screening the cDNA library with one oligonucleotide pool from a specific region of the protein, and then re-screening all candidate clones with a second oligonucleotide pool, it is possible to identify a subset of double-positive clones that warrant further characterization.
How many different oligonucleotides are required to generate the NT probe based on the Genetic Code?

What is the rationale for using the two probes in successive screenings? Will any cDNAs be missed this way?

Does the oligonucleotide pool contain any probe molecules that are 100% complementary to the cDNA? Explain.

What is included in the hybridization solution to minimize differences in oligonucleotide Tm? Explain.

**Antibody probes to detect antigenic fusion proteins**

Protein-specific antibodies can be used as "probes" to immunologically detect bacterial fusion proteins encoding the protein antigen. Antibody screening of cDNA libraries is facilitated by special lambda phage expression vectors that **produce fusion proteins** which can be detected on membrane filters. The most commonly used cDNA cloning vectors for antibody screening are the **lambda gt11** and **lambda Zap** phage vectors which contain an E. coli transcriptional promoter linked to the beta-galactosidase (**LacZ**) coding sequences.

Why is the filter soaked in IPTG before overlaying the plate, why not just include IPTG in the growth media?

What does a positive signal in an antibody screening strategy tell you about the fusion protein?

What two methods can be used to determine if the "antigen" cDNA corresponds to your gene of interest?

**Screening cDNA libraries based on differential expression**

In this screening strategy, mRNA is isolated from cells that express the phenotype (or protein) of interest (+), and from cells that do not (-). The (+)mRNA is converted to radioactive (+)cDNA using RTase and 32P-dNTPs, and then hybridized to a mass excess of (-)mRNA using solution hybridization. By removing the double-stranded cDNA:mRNA heteroduplexes, it is possible to obtain an enriched probe containing (+)cDNA sequences.

Although subtraction hybridization screening has been useful in the isolation of a number of important cDNAs, this method has largely been superseded by PCR-dependent strategies and high-throughput screening using.

This example illustrates how salt-stress in plants can alter the expression of genes required for osmotic regulation. High specific activity cDNA probes, are hybridized to duplicate filter lifts to identify plaques that contain cDNA inserts corresponding to differentially expressed genes.
Cloned cDNA sequences have the potential to encode active proteins. The technique of cDNA "expression cloning" uses a functional assay to identify positive recombinants without prior knowledge of the gene product.

**Antibody screening** of cDNA expression libraries is a type of functional screen, however it is in principle different than expression cloning since sufficient information about the protein antigen is already available.

Three types of cDNA screening methods based on **functional assays** have been exploited:

1. Protein **activity assays** (table 5.1); see HOT chili peppers.
2. Yeast **two-hybrid system** is based on protein-protein interactions.
3. cDNA **phage display** is a method utilizing solid support interactions.

Examples of protein activity assays used to screen cDNA libraries.

<table>
<thead>
<tr>
<th>Gene product</th>
<th>cDNA screening strategy</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GMCSF</td>
<td>supernatant from transiently-transfected COS cells was collected and assayed for GMCSF activity using a bioassay</td>
<td>Lee, F. et al. (1985)</td>
</tr>
<tr>
<td>CD2 antigen</td>
<td>transiently-transfected COS cells were selected for cell adherence using anti-CD2 antibody-coated dishes</td>
<td>Seed, B and Aruffo, A. (1987)</td>
</tr>
<tr>
<td>Oct-2</td>
<td>double-stranded oligonucleotide probes were used to screen λgt11 filters for sequence-specific DNA binding activity</td>
<td>Staudt, et al. (1988)</td>
</tr>
<tr>
<td>Xwnt-8</td>
<td>Xenopus embryos were assayed for dorsalizing activity after injection with in vitro transcribed RNA from pooled directionally-cloned cDNAs</td>
<td>Smith, W. and Harland, R. (1991)</td>
</tr>
<tr>
<td>XPCC</td>
<td>UV-resistant cells were selected from a DNA repair deficient cell line that was transfected with an episomal vector expression library</td>
<td>Legarski, R. and Peterson, C. (1992)</td>
</tr>
<tr>
<td>angiopoietin-1</td>
<td>transfected COS cells were formaldehyde-fixed and assayed in situ for the localized secretion of ligands that bound to soluble TIE2 receptor</td>
<td>Davis, S. et al. (1996)</td>
</tr>
<tr>
<td>MNK1</td>
<td>λ phage filter lifts were screened for expression of GST fusion proteins that can be labeled by soluble ERK1 MAP kinase and [γ-32P]ATP</td>
<td>Fukunaga, R. and Hunter, T. (1997)</td>
</tr>
<tr>
<td>capsacin receptor</td>
<td>transfected HEK293 cells were monitored for capsacin-dependent calcium release; sib selections were used to identify a single cDNA clone</td>
<td>Caterina, M. et al. (1997)</td>
</tr>
</tbody>
</table>
List 3 criteria that must be met for these protein activity assays to work for cDNA screening.

What was it about the bioassay that Julius and colleagues used to identify cells expressing the capsaicin receptor that permitted the detection of candidate clones?

How does sib selection work in the context of plasmid pools and the Julius cloning strategy?

Yeast Two-Hybrid screening

Biological processes require protein-protein interactions that can be defined by non-covalent molecular interactions. The idea of using a cloned protein to screen cDNA expression libraries for interacting proteins, follows from the work of Stan Fields and colleagues.

The yeast two-hybrid strategy was based on an earlier finding by Mark Ptashne and co-workers demonstrating that the yeast Gal4 transcription factor is a modular protein. The Ptashne lab showed that the transcriptional regulatory activity of Gal4 requires a sequence specific DNA binding domain (DBD), and a transcriptional activation domain (AD).

The yeast two-hybrid system is used to identify cDNA sequences encoding protein domains capable of interacting with a "bait" protein. The DBD-bait and AD-cDNA plasmids replicate episomally in yeast.

What are the three functional components required for the Two-Hybrid screening strategy to work?

What protein function cannot be present in the bait protein, i.e., what function would give high background?

Why would you want to have both the Gal4 dependent His3 and LacZ reporter genes in the host yeast strain? Which is a genetic selection and which is a genetic screen?

cDNA Phage Display library screening

There are a number of important biological interactions that occur between a protein product and a small diffusible molecule, such as a ligand or antigen. The Phage Display strategy was designed to find functional proteins that bind small ligands.

Filamentous phage cDNA expression libraries can be screened for fusion proteins that bind to a solid support using the panning method. In-frame coding sequences have the potential to produce a fusion protein that is expressed on the phage surface.

Sequential panning and infection cycles are carried out to enrich for phage that bind to the "bait" attached to the solid support. The phagemids are rescued in E. coli and individual picks can be assayed by superinfection with M13 helper phage to produce phage for a 96-well ELISA (enzyme-linked immunosorbent assay).