TECHNIQUES IN GENETIC ENGINEERING

Işıl KURNAZ

CHAPTER 3

DNA libraries

“It is a good morning exercise for a research scientist to discard a pet hypothesis every day before breakfast. It keeps him young.”

Konrad Lorenz (1903 – 1989)

A DNA library is essentially a representation of the entire DNA set (be it genomic or complementary) in a living organism; commonly in bacteria.

In this chapter we will learn about the two major types of libraries, how and why they are made, and how they are screened.

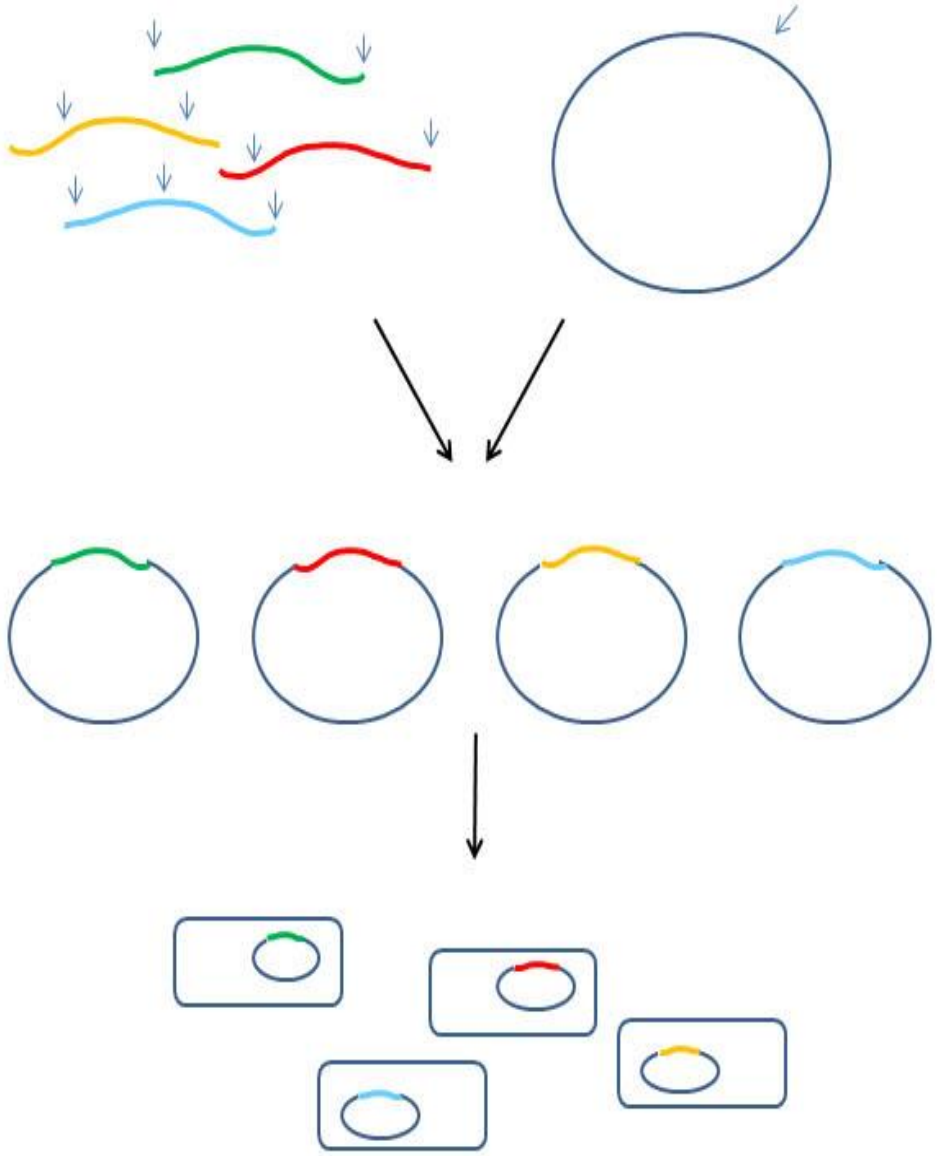
A **genomic DNA library** can be defined as the representation of the entire genomic DNA of an organism, including coding and non-coding regions alike.

The **cDNA library**, on the other hand, is constructed by the reverse transcription of the mRNA transcripts, and thus only represents the protein coding regions of the genome.

Therefore, the two libraries are used for different purposes that will be described in detail. Based on the type of library, **(a)** the source of DNA, **(b)** the type of vector used, and **(c)** the experimental methodologies differ

The basic principle of all DNA libraries involves:

- Preparation of representative DNA fragments;
- Preparation of the appropriate vector;
- Cloning and transformation;
- Selection and screening.



Genomic DNA libraries

The genomic DNA fragments will contain not only coding sequences, but also introns, promoter regions, other types of regulatory regions, several repeat elements etc. Therefore such libraries would be useful in screening for non-coding regions of DNA.

Figure 3.1. A schematic representation of **genomic DNA library**. Essentially, almost all of the genome is targeted for fragmentation and cloning into different bacterial clones. The cloning vector is digested in this cartoon, and ligated with partial or full digestions of the genome; the ligation products are transformed into bacteria. Ideally each bacterial clone should include a different genomic fragment

cDNA libraries

cDNA is short for **complementary DNA**, ie DNA that has been synthesized as complementary to mature mRNA, representing only the coding regions or exons. Therefore cDNA library refers to a representation of most, if not all, mature mRNA transcripts isolated from a cell type, or in other words, representative of all the expressed genes in a given cell at a given time and a given condition.

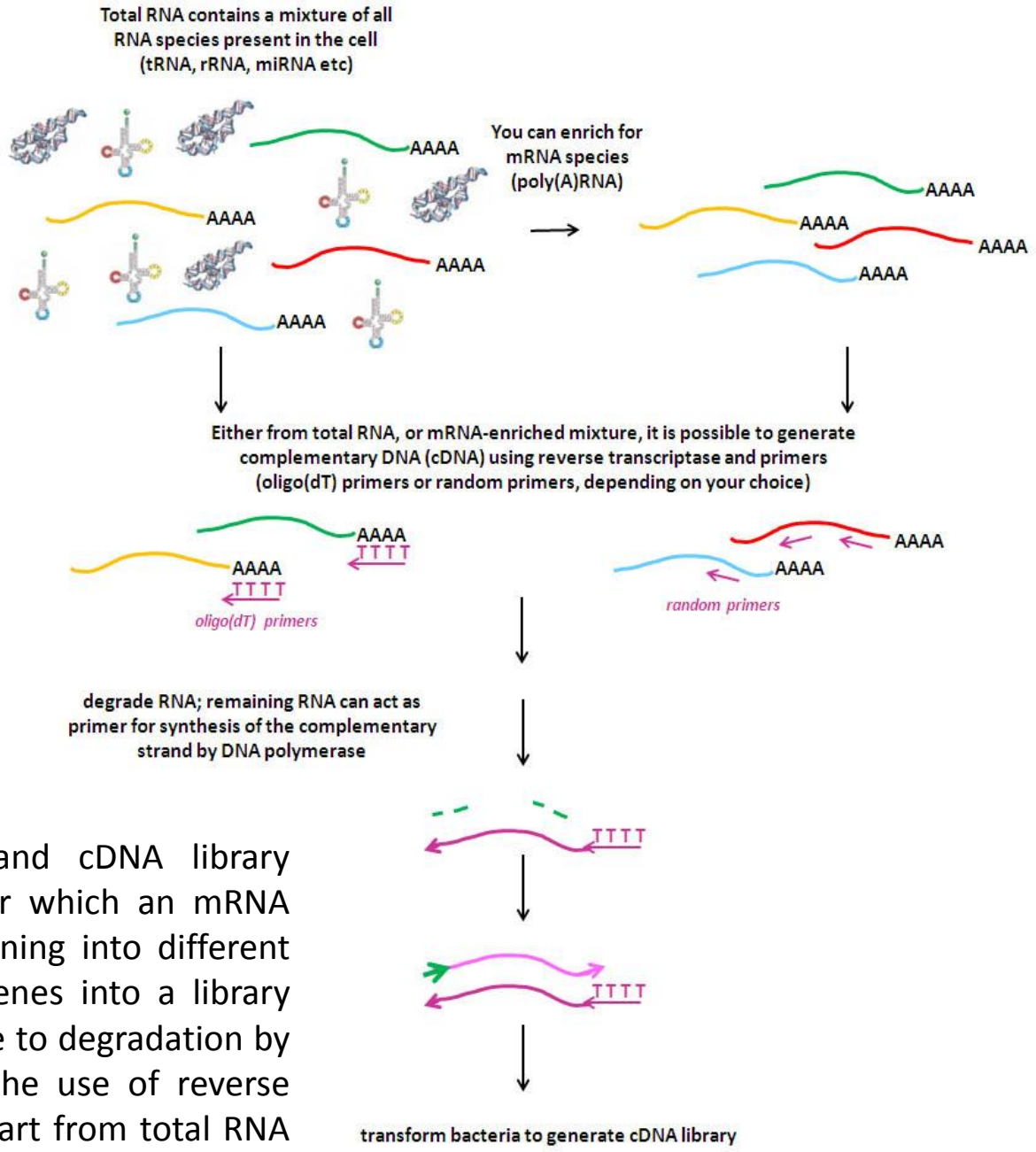


Figure 3.2. A schematic representation of cDNA synthesis and cDNA library construction principle. Essentially, only the expressed genes (for which an mRNA transcript would be present in the cytoplasm) is targeted for cloning into different bacterial clones. However, in order to clone these expressed genes into a library vector, one must first convert the mRNA (which is highly susceptible to degradation by RNases) to cDNA (which is resistant to RNase degradation) by the use of reverse transcriptase and primers. The cDNA synthesis procedure could start from total RNA mixture (top left), or one that is enriched for mRNA (top right).

Library Screening

Once the library is thus obtained, then one needs to screen this library for a particular gene of interest, usually using a homologous DNA as a **probe**, which has to be labeled either radioactively or non-radioactively. This probe could either be the homologue of the DNA fragment from a different species (such as the coding region of a human gene used as a probe to screen a mouse cDNA library from the tissue of interest), or part of a coding region to "hunt" for a promoter region, or part of a gene from one tissue to screen for homologues in a different tissue, etc. The basic principle behind probe-based screening is that single stranded DNA or RNA species will hybridize to complementary sequences on the library

Monitoring transcription

Reverse Transcription PCR (RT-PCR)

(not to be confused with Real Time PCR; prefer to use Q-PCR, or quantitative PCR, to avoid confusion)

Since RNA is relatively less stable than DNA, in other words more amenable to degradation by RNases that are quite abundant everywhere, it is more difficult to work with the mRNA species for comparison of expression levels. Researchers therefore often refer to the cDNA that is synthesized from mRNA by reverse transcription for comparison of expression levels between different samples. As with every other experiment, an internal control is required to maintain some level of quantification: mostly a housekeeping gene transcript such as that of glyceraldehyde-3-phosphate dehydrogenase (GAPDH or G3PDH) enzyme or b-tubulin gene is used for normalization among different samples. The cDNA mix (usually 1^{st} strand is sufficient for this purpose) generated from mRNA from different samples are then subjected to PCR amplification by primers specific for the gene of interest

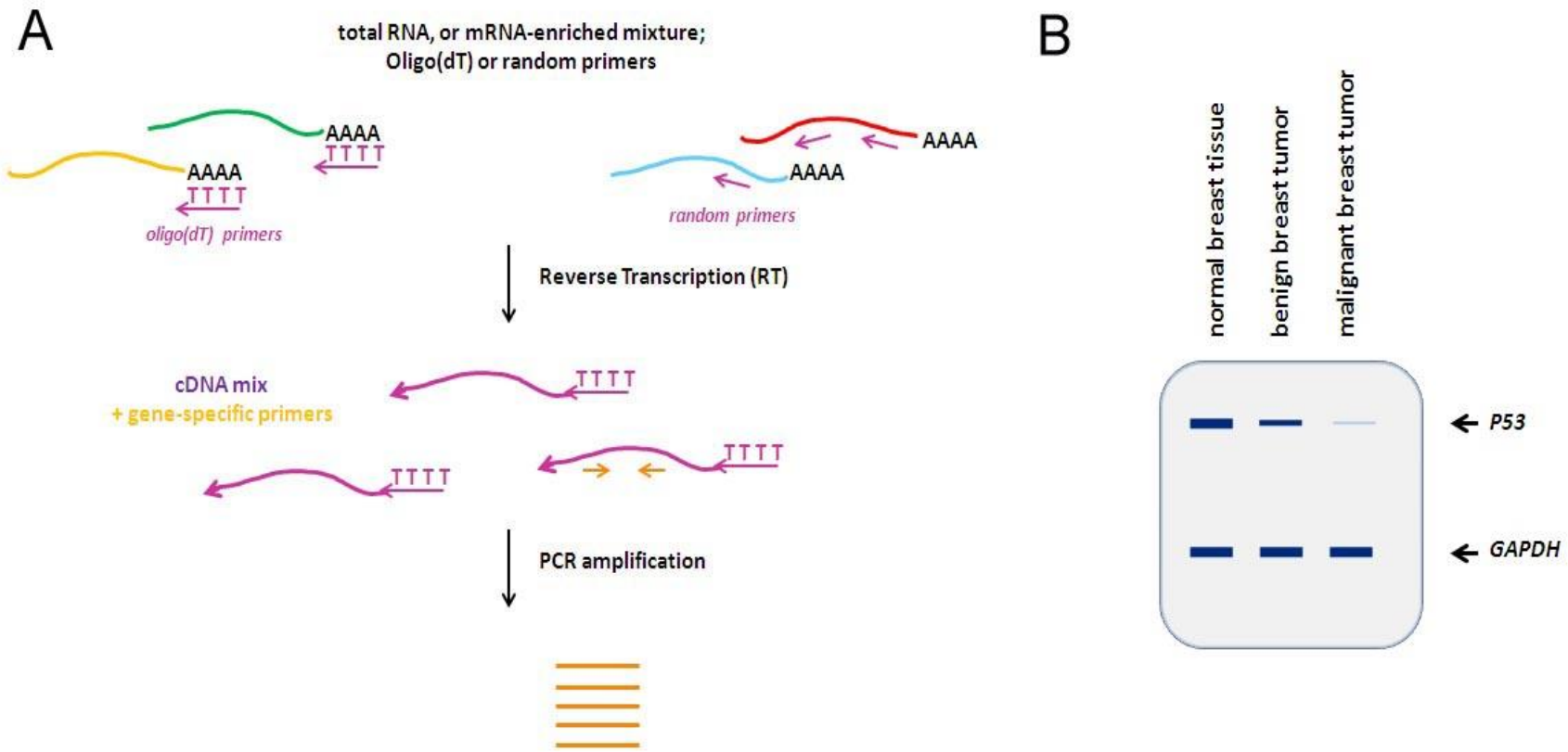


Figure 3.3. Reverse Transcription Polymerase Chain Reaction (RT-PCR). (a) A schematic summary of RT-PCR methodology; (b) a hypothetical example of an RT-PCR result on agarose gel, showing amplifications with primers specific for *GAPDH* internal control and *P53* transcript in three different tissue samples.

If no gene sequence is available for your favorite protein, but for example protein sequences are available from various species, for instance, then degenerate primers can be constructed to “fish out” the gene from cDNA mix without the need for a library construction and screening (but word of caution: with quite high false positives, so thorough follow-up assays are a must!)

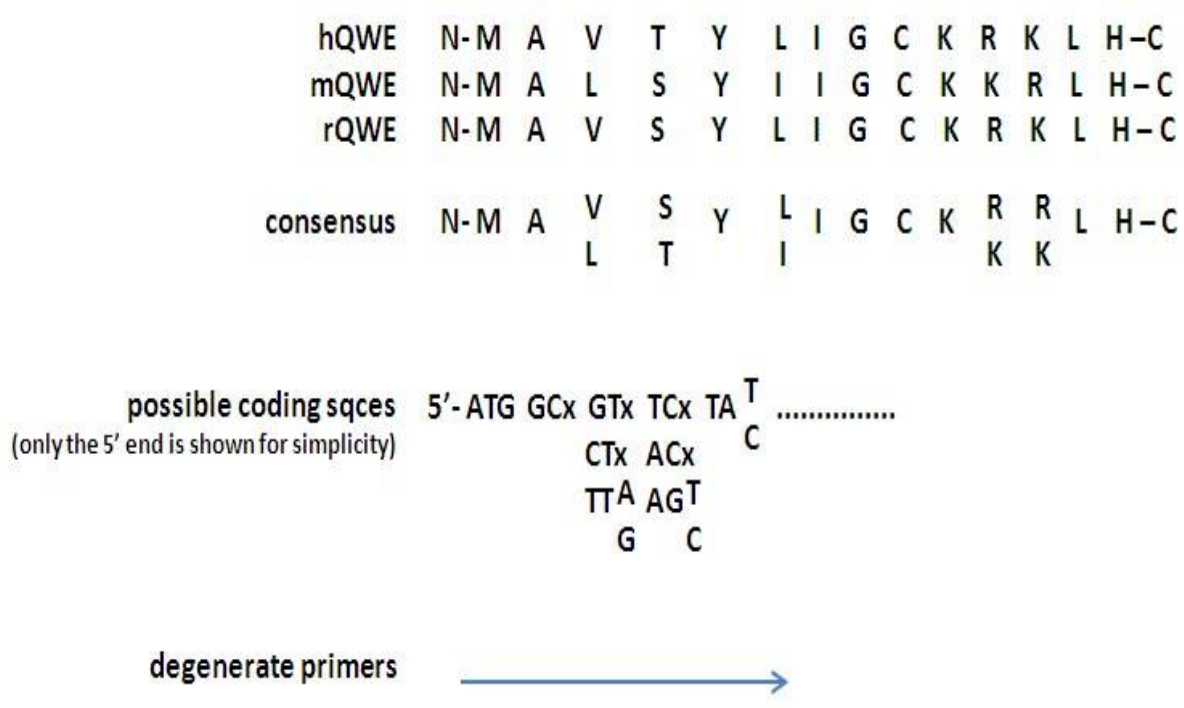


Figure 3.4. A hypothetical example of how degenerate primers could be deduced from known protein sequences through alignment. Online bioinformatic tools are also available for automated design of degenerate primers

Northern blotting

Northern blotting is technique used for the analysis of RNA, and most commonly mRNA, among samples, and was developed by James Alwine, George Stark and David Kemp. It was named as such after the Southern blot, developed by Edwin Southern, used for the detection of DNA. In both Southern and Northern blotting, the nucleic acids in the sample (DNA in the former and RNA in the latter) are separated in electrophoresis gel and thereafter “blotted” on a membrane, and then probed for screening

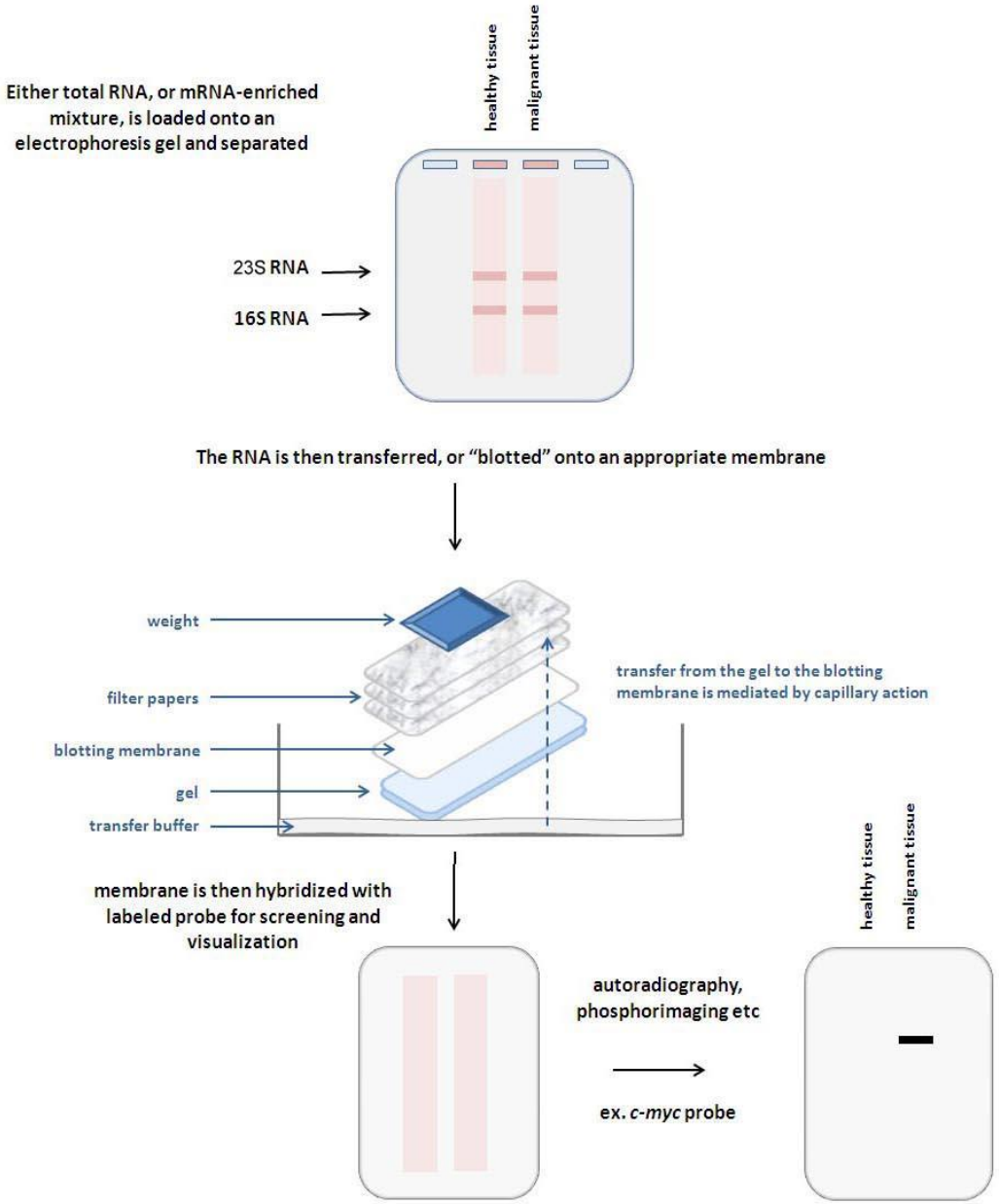


Figure 3.5. A schematized summary of Northern blotting. The RNA isolated from different samples (in this example, from healthy tissue vs malignant tumor) are loaded onto an electrophoresis gel and separated. These bands are then transferred to a blotting membrane by capillary action. The membrane is then hybridized with a labeled probe (such as a *c-myc* probe in this hypothetical example), and the signal visualized by a number of different methods (such as autoradiography, phosphorimager or another appropriate method).

Preparing the probe

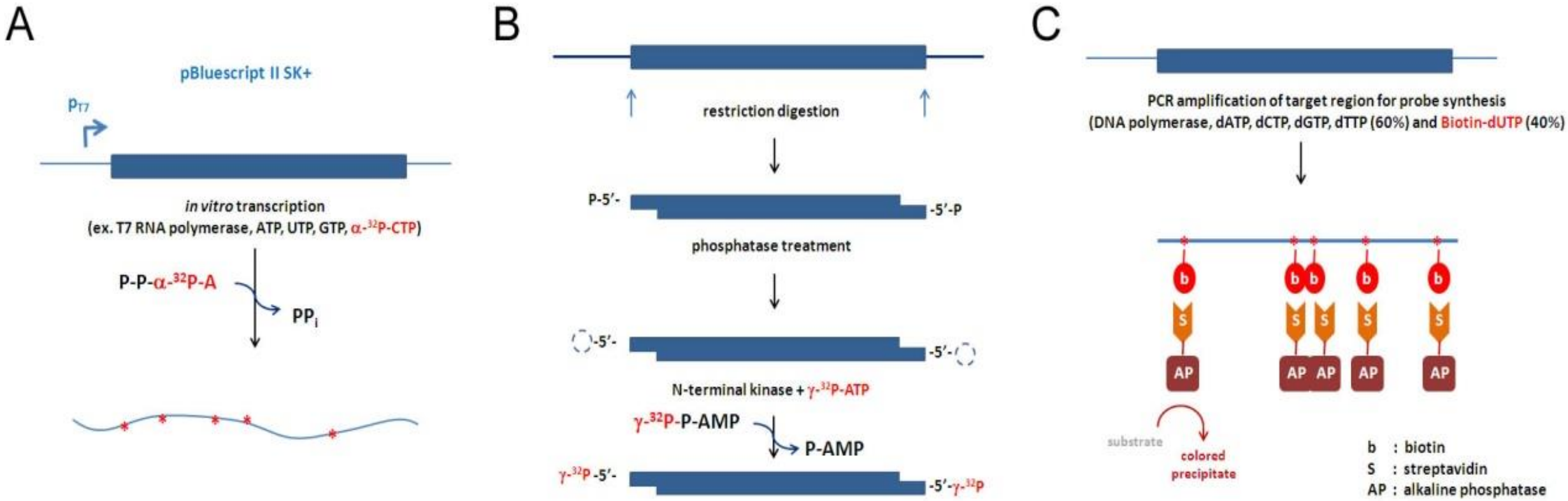


Figure 3.6. A simplified schema of probe preparation strategies. **(a)** Preparation of RNA probe by *in vitro* transcription. For radioactive labeling, a-Phosphate of the nucleotide that gets incorporated to the transcript should be labeled with the radioisotope. **(b)** end-labeling of a cDNA fragment. The g-Phosphate of ATP gets transferred to the DNA fragment. **(c)** PCR-based probe generation using biotinylated deoxynucleotide. Note that a mixture of unlabeled (dTTP) and labeled (biotin-dUTP) can be used when making the probe. The affinity of biotin to streptavidin is exploited for detection: an enzyme such as alkaline phosphatase can be conjugated to streptavidin, which then converts a colorless substrate to a colored precipitate for detection

Nuclease Protection Assay

Nuclease protection assay, and in particular ribonuclease protection assay (RPA) that will be discussed here, is more sensitive than the traditional Northern blot for the detection and quantification of specific RNA species in the total RNA isolate.

The principle of nuclease protection assay is the hybridization of specific target RNA (or RNAs; multiple RNA species can be detected simultaneously, so long as the probes are of different lengths) with a target-specific antisense probe, generating a double-stranded hybrid (DNA/RNA or RNA/RNA). Any unhybridized probe or unhybridized RNA in the sample will be digested with single-stranded RNA-specific nucleases (or S1 nuclease when probe is a DNA molecule). Thereafter, nucleases are inactivated, the double stranded nucleic acid hybrids are precipitated, and analyzed in gel electrophoresis

Microarray analysis

In either RT-PCR or Northern blotting, one must have an idea about which gene could be expressed so that probes specific for that transcript can be prepared. However, if one wishes to address a more general and unbiased question of “which genes may be expressed differently among samples”, then a more general method for screening gene expression is required. Microarray or DNA chip technology is rather useful for this purpose and has become a standard for studying global gene expression profiles.

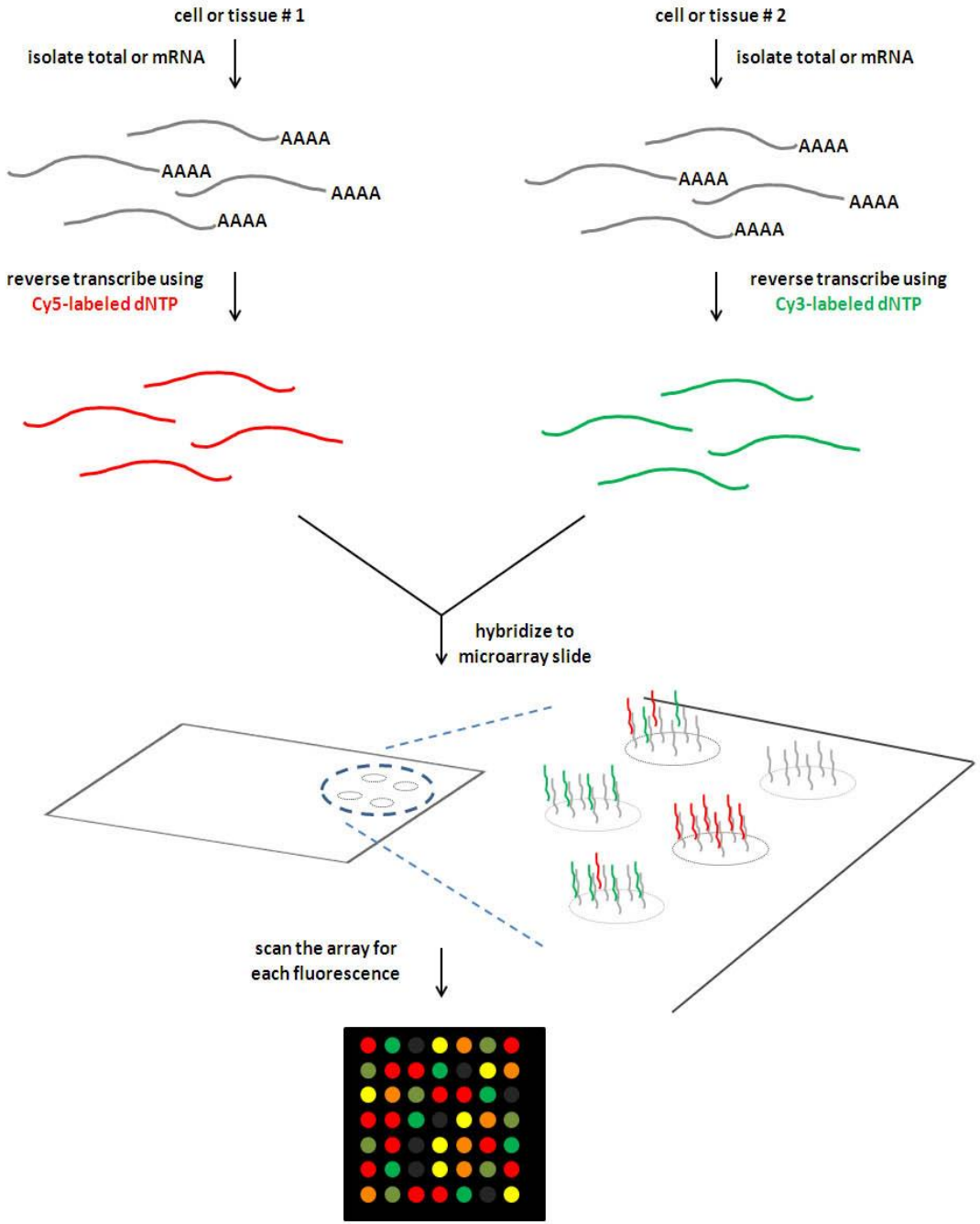


Figure 3.7. An overview of microarray analysis for gene expression profiling. Usually, RNA from two different samples are isolated, and reverse-transcribed using labeled deoxynucleotides (such as Cy3- and Cy5-labeled dNTPs). These labeled cDNAs are then mixed and hybridized with the microarray. Each spot on microarray corresponds to a different gene (shown here as grey wavy lines in the zoomed in region), and depending on which genes are transcribed in each sample (and in which amount), the labeled cDNAs will hybridize differently to each spot. When the array is scanned for each fluorescence and the information is merged, each spot will not only show a different color (different scales of red, green or yellow), but also have a different intensity measurement recorded in an Excel file or similar for further analysis.

“A scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die and a new generation grows up that is familiar with it.”

Max Planck (1858-1947)