**STUDYING GENE EXPRESSION**

 All genes have to be expressed in order to function. The first step in expression is transcription of the gene into a complementary RNA strand. For some genes-for example, those coding for transfer RNA (tRNA) and ribosomal RNA (rRNA) molecules—the transcript itself is the functionally important molecule. For other genes, the transcript is translated into a protein molecule.

 To understand how a gene is expressed, the RNA transcript must be studied. In particular, the molecular biologist will want to know whether the transcript is a faithful copy of the gene, or whether segments of the gene are missing from the transcript. These missing pieces are called introns and considerable interest centres on their structure and possible function. In addition to introns, the exact locations of the start and end points of transcription are important. Most transcripts are copies not only of the gene itself, but also of the nucleotide regions either side of it. The signals that determine the start and finish of the transcription process are only partly understood, and their positions must be located if the expression of a gene is to be studied.

**Studying the RNA transcript of a gene**

 Over the years a variety of techniques have been devised for studying RNA transcripts. Some of these techniques merely detect the presence of a transcript and give some indication of its length, others enable the start and end of the transcript to be mapped and the positions of introns to be located.

***Detecting the presence of a transcript and determining its nucleotide sequence***

 Before studying the more sophisticated techniques for RNA analysis, we must consider the methods used to obtain basic information about a transcript. The first of these methods is **northern hybridization**, the RNA equivalent of Southern hybridization which is used to measure the length of a transcript. An RNA extract is electrophoresed in an agarose gel, using a denaturing electrophoresis buffer (e.g., one containing formaldehyde) to ensure that the RNAs do not form inter- or intramolecular base pairs, as base pairing would affect the rate at which the molecules migrate through the gel. After electrophoresis, the gel is blotted onto a nylon or nitrocellulose membrane, and hybridized with a labeled probe. If the probe is a cloned gene, the band that appears in the autoradiograph is the transcript of that gene. The size of the transcript can be determined from its position within the gel, and if RNA from different tissues is run in different lanes of the gel, then the possibility that the gene is differentially expressed can be examined.

***Transcript mapping by hybridization between gene and RNA***

 Nucleic acid hybridization occurs just as readily between complementary DNA and RNA strands as between single-stranded DNA molecules. If a hybrid is formed between a DNA strand, containing a gene, and its mRNA, then the boundaries between the double- and single-stranded regions will mark the start and end points of the mRNA. Introns, which are present in the DNA but not in the mRNA, will “loop out” as additional single-stranded regions.

 Now consider the result of treating the DNA–RNA hybrid with a single-strand specific nuclease such as S1. S1 nuclease degrades single-stranded DNA or RNA polynucleotides, including single-stranded regions at the ends of predominantly double stranded molecules, but has no effect on double-stranded DNA or on DNA–RNA hybrids. S1 nuclease will therefore digest the non-hybridized single-stranded DNA regions at each end of the DNA–RNA hybrid, along with any looped-out introns. The single-stranded DNA fragments protected from S1 nuclease digestion can be recovered if the RNA strand is degraded by treatment with alkali.



Figure: A DNA-mRNA hybrid and the effect of treating this hybrid with a single-strand-specific nuclease such as S1.

***Transcript analysis by primer extension***

 S1 nuclease analysis is a powerful technique that allows both the 5′ and 3′ termini of a transcript, as well as the positions of intron–exon boundaries, to be identified. The second method of transcript analysis that we will consider—**primer extension**—is less adaptable, because it can only identify the 5′ end of an RNA. It is, nonetheless, an important technique that is frequently used to check the results of S1 analyses.

 Primer extension can only be used if at least part of the sequence of the transcript is known. This is because a short oligonucleotide primer must be annealed to the RNA at a known position, ideally within 100–200 nucleotides of the 5′ end of the transcript. Once annealed, the primer is extended by reverse transcriptase. This is a cDNA synthesis reaction, but one that is very likely to proceed to completion as only a short segment of RNA has to be copied. The 3′ end of the newly synthesized strand of DNA will therefore correspond with the 5′ terminus of the transcript. Locating the position of this terminus on the DNA sequence is achieved simply by determining the length of the single-stranded DNA molecule and correlating this information with the annealing position of the primer.



Figure: Locating a transcription start point by primer extension.

***Studying transcriptome by microarray or chip analysis***

 Techniques that enable more accurate comparisons of the amounts of individual mRNAs were first developed as part of the yeast post-genomics project. In essence, these techniques involve a sophisticated type of hybridization analysis. Every yeast gene—all 6000 of them—was obtained as an individual clone and samples spotted onto glass slides in arrays of 80 × 80 spots. This is called a **microarray**. To determine which genes are active in yeast cells grown under particular conditions, mRNA was extracted from the cells, converted to cDNA and the cDNA labeled and hybridized to the microarrays. Fluorescent labels were used and hybridization was detected by examining the microarrays by confocal microscopy. Those spots that gave a signal indicated genes that were active under the conditions being studied, and the intensities of the hybridization signals revealed the relative amounts of the mRNAs in the transcriptome. Changes in gene expression, when the yeast were transferred to different growth conditions (e.g., oxygen starvation), could be monitored by repeating the experiment with a second cDNA preparation.

 Microarrays are now being used to monitor changes in the transcriptomes of many organisms. In some cases the strategy is the same as used with yeast, the microarray representing all the genes in the genome, but this is possible only for those organisms that have relatively few genes. A microarray for all the human genes could be carried by just 10 glass slides of 18 × 18 mm, but preparing clones of every one of the 20,000–30,000 human genes would be a massive task. Fortunately this is not necessary. For example, to study changes in the transcriptome occurring as a result of cancer, a microarray could be prepared with a cDNA library from normal tissue. Hybridization with labeled cDNA from the cancerous tissue would then reveal which genes are up- or down-regulated in response to the cancerous state.

 An alternative to microarrays is provided by **DNA chips**, which are thin wafers of silicon that carry many different oligonucleotides. These oligonucleotides are synthesized directly on the surface of the chip and can be prepared at a density of 1 million per cm2, substantially higher than is possible with a conventional microarray. Hybridization between an oligonucleotide and the probe is detected electronically. Because the oligonucleotides are synthesized *de novo*, using special automated procedures, a chip carrying 20,000–30,000 oligonucleotides, each one specific for a different human gene, is relatively easy to prepare.