Plenary Lecture: DNA Microarrays—the How and the Why

Ed Southern
Department of Biochemistry
University of Oxford
South Parks Road
Oxford OX1 3QU
U.K.
ems@bioch.ox.ac.uk

Abstract

Much of the appeal of DNA chip technology to computer scientists—several lectures in previous RECOMB meetings have addressed the topic—no doubt derives from the similarity of the devices to the familiar chips in their computers. But the cross-references go beyond this physical similarity. Among the earliest proposals for applications of DNA chips was "sequencing by hybridisation", or "SBH". The conceptual basis for this method resembles an algorithm for DNA sequence comparison in the computer, and it was not long before the field of SBH was taken over by theoreticians who came up with ingenious methods for interpreting the data and for improving the efficiency of the process. However, experimentalists have not stood still and several useful applications for arrays have emerged in recent years.

Much of the work of our group has been directed to improving the technology and to studying the molecular basis of the processes underlying the use of DNA arrays. Hybridisation to microarrays is the latest in the line to exploit the potent feature of the DNA structure that enables the two strands to come together—their sequence complementarity. It is remarkable that a molecule of such great structural complexity as the DNA double helix can reassemble with perfect fidelity from the separated strands.

In the analytical procedures which exploit the hybridisation process, a strand of known sequence—the probe—permits the detection or measurement of its complement in the unknown—the target. Most procedures, such as blotting and the PCR, use one or two probes at a time. The microarrays, which comprise hundreds to thousands of probes, are particularly useful tools for the analysis of the large number of targets characterised by genome projects.

Two different types of arrays have been described. Arrays of probes specific to known sequences are useful for the analysis of gene expression and sequence variation on scales embracing the whole gene set.

Of greater appeal to theoreticians, particularly computer scientists, are the generic arrays, comprising all oligonucleotide sequences of a given length, e.g. all 65,536 octanucleotides, which can be used to analyse any sequence, even those which are unknown. For some applications, such as the analysis of variation the hybridisation process must distinguish between sequences which differ by only one base. This can only be achieved with relatively short probes, and it is in this area that the generic arrays are likely to find their main applications.

Surprisingly little is known about the mechanism by which Watson-Crick paired structures are formed from the single strands. We have studied this in some detail, especially the influence that intramolecular structure in long target sequences may have on the process. This will be one of the main topics of the presentation. We will also show how arrays are helping to characterise the functions of genes of unknown function by the selection of effective antisense reagents to use in gene knockout experiments.