Sequence to array: Probing the genome's secrets

Alan P. Blanchard and Leroy Hood

The Human Genome Project has led to the development of two high-throughput technologies that provide powerful approaches to the analysis of complex biological systems: The highly parallel measurements of DNA fragment sizes through the electrophoresis of fluorescent DNAs[1], and the creation of DNA arrays on solid supports for large-scale hybridization experiments[2-5]. The automated fluorescent sequencer has been developed for large-scale DNA sequence analysis and genome-wide genotyping of polymorphic microsatellites—applications lying at the heart of the genome project and the genetics of humans and model organisms. DNA arrays permit the interrogation of information expressed in differing cell types or cell states. The reports by Ferguson et al.[6] and Lockhart et al.[7] (pp. 1681–1684 and pp. 1675–1680) are the latest in the development of DNA arrays as a tool for probing, in a highly parallel fashion, the dynamics of a cell, as revealed through changing mRNA concentrations.

DNA arrays for hybridization experiments can be made by binding DNA from genomic or cDNA libraries on a solid support, such as a nitrocellulose filter. But, these approaches have several drawbacks: First, each clone must be identified by sequence analysis; second, some genomic and cDNA clones have repetitive sequences, leading to confusing cross hybridizations; and third, cDNA clones may differ in expression levels by four orders of magnitude, rendering difficult the identification of genes expressed at low levels.

An alternative is to synthesize oligomer probes directly on a solid support (glass or silicon), or in solution and then subsequently bind them to a surface. This strategy makes use of the exponentially increasing amount of DNA sequence data available to design the oligonucleotide arrays. Edwin Southern made the first oligonucleotide arrays using traditional DNA synthesis on glass plates[3]. Stephen Fodor and his coworkers[4] employed solid-phase DNA synthesis and photolithography to synthesize oligonucleotide arrays on glass with densities that may approach 10^7 per cm^2. We are pursuing a similar strategy using an inkjet technology[5].

The capacity to employ DNA arrays of

genomic fragments, cDNA clones and/or oligonucleotides is revolutionizing our ability to analyze gene expression patterns[6,7], carry out genome-wide genetic mapping[8], clone members of gene families within and across species, scan for mutations in interesting genes[9], define networks of genes controlled by particular transcription factors, and in the future, we will be able to examine DNA–protein interactions on oligonucleotide arrays of double-stranded DNA.

Lockhart et al. label complex mixtures of mRNA with a fluorescent dye and then hybridize against an array of oligonucleotides (20-mers, made by the photolithographic method) complementary to a small number of genes. Different oligonucleotides belonging to the same gene can show large differences in hybridization efficiency, presumably due to secondary structure of the probe (and possibly DNA), and special software is necessary to reliably interpret the data. The specificity of hybridization is checked by incorporating oligonucleotides with single mismatches into the array. By spiking the total mRNA with known control RNAs, they demonstrated the ability to detect 20% changes in mRNA concentration. The overall sensitivity of the system permits the detection of a single mRNA species in the presence of large amounts of total mRNA down to a few copies per cell. Ferguson et al. employed a fiber-optic bundle, with each fiber carrying an immobilized presynthesized oligonucleotide at its end, to monitor fluorescently labeled complementary oligonucleotides. Both strategies provide a rapid and quantitative approach to studying gene expression. The DNA array technology is particularly exciting because it offers the potential of using hundreds of thousands of differing oligonucleotides to analyze changing mRNA populations.

DNA synthesis allows various chemical modifications to the DNA. So-called universal bases, such as inosine and 5-nitroindole, can be incorporated at positions where the desired base is unknown or polymorphic. The charge on the phosphate backbone can be changed by thiolation or methylation. Although unreported, it should be possible to create arrays of peptide nucleic acids, with their extraordinarily high hybridization affinities.

Both of these high-throughput technologies, DNA sequencing and DNA arrays, afford a powerful approach to the analysis of complex biological systems and networks. Biology of the 1970s, 1980s, and early 1990s focused almost exclusively on the study of

individual genes and proteins. The tools of genomics provide the possibility for global studies of biological complexity. The challenge is to break down large and intractable systems (e.g., immunity or inflammation) into analyzable subsystems. Each technique of genomics provides powerful approaches to defining tractable subsystems—DNA arrays through the analysis of expression patterns, DNA sequencing through the analysis of multigene families[10], and genotyping through the identification of polymorphic genes controlling particular traits. Indeed, oligonucleotide arrays will be the next generation genotyping tool[8], with the possibility, for example, of simultaneous interrogation of 6000 randomly scattered single-base polymorphisms—leading to the rapid and more precise chromosomal location of the genes controlling polygenic traits.

Because the entire yeast genome has been sequenced and all of its approximately 6000 genes identified, it offers the ultimate example of systems analysis. Oligonucleotide arrays representing all 6000 genes can be used to interrogate how gene expression patterns change when specific biological activities are triggered (e.g., DNA replication, DNA repair, protein secretion, and change in yeast mating type) or particular genes are destroyed. Indeed, Shoemaker et al. have demonstrated the elegant use of oligonucleotide markers and oligonucleotide arrays to monitor large numbers of yeast deletion mutants under differing selective conditions[11].

The important point is that systems behaviors can only be understood by studying systems (or subsystems) and the high-throughput tools of genomics afford powerful opportunities for systems analysis. More than 50,000 different human genes have already been identified. As the complete repertoire of 100,000 or so human genes becomes available, the complete analysis of complex genetic systems in humans also becomes increasingly tractable.

Alan Blanchard is senior fellow and Leroy Hood is William Gates III professor and chairman of the department of molecular biotechnology, University of Washington, 1705 North East Pacific, Seattle, WA 98195-7730 (lee@nirvana.mbt.washington.edu).

1. Smith, L.M. et al. 1986. *Nature* **321**:674–679.
2. Gress, T.M. et al. 1992. *Mamm. Genome* **3**:609–619.
3. Maskos, U. and Southern, E.M. 1993. *Nucleic Acids Res.* **21**:2267–2268.
4. Fodor, S.P et al. 1991. *Science* **251**:767–773.
5. Blanchard, A.P., Kaiser, R.J., and Hood, L. 1996. *Biosens. Bioelectron.* **11**:687–690.
6. Ferguson, J.A. et al. 1996. *Nature Biotechnology* **14**:1681–1684.
7. Lockhart, D.J. et al. 1996 *Nature Biotechnology* **14**:1675–1680.
8. Chee, M. et al. 1996. *Science* **274**:610–614.
9. Hacia, J.G. et al. 1996. *Nat. Genet.* **14**:In press.
10. Rowen, L., Koop, B.F., and Hood, L. 1996. *Science* **272**:1755–1762.
11. Shoemaker, D.D. et al. 1996. *Nat. Genet.* **14**:In press.