Molecular interactions on microarrays

Edwin Southern, Kalim Mir & Mikhail Shchepinov

Department of Biochemistry, University of Oxford, South Parks Road, Oxford OX1 3QU, UK. e-mail: ems@bioch.ox.ac.uk

The structural features of nucleic acid probes tethered to a solid support and the molecular basis of their interaction with targets in solution have direct implications for the hybridization process. We discuss how arrays of oligonucleotides provide powerful tools to study the molecular basis of these interactions on a scale which is impossible using conventional analysis.

The DNA microarray is the latest in a line of techniques to exploit a potent feature of the DNA duplex—the sequence complementarity of the two strands. It is remarkable that a molecule of such great structural complexity can reassemble with perfect fidelity from the separated strands. Early studies of duplex melting and reformation, which were carried out on DNA solutions, provided valuable basic information: the dependence of T_m on G+C content and on salt concentration; and the dependence of rate of reassociation on sequence complexity.

The introduction of solid supports greatly increased the range of applications of the method and set the trail to array-based methods. The starting point was the observation that single-stranded DNA binds strongly to nitrocellulose membranes in a way that prevents the strands from reassociating with each other, but permits hybridization to complementary RNA[1]. This simple method has led to the provision of fundamentally important data. For example, it was used to measure the number of copies of repeated genes, such as those for the ribosomal RNAs and tRNAs, in eukaryotes, and to measure changes in the number of copies during processes such as amplification[2]. It helped in the purification of the genes for the ribosomal RNAs by density gradient centrifugation[3] before the advent of DNA cloning. And when cloning arrived, it provided a way of finding those clones which included specific sequences[4]. It was the direct antecedent of the 'blotting' methods, the first of which combined filter hybridization with gel separation of restriction digests[5]. More relevant to microarrays are the methods of 'dot-blotting'[6]. Subsequent automation and miniaturization of the dot-blot showed how hybridization could be used on a large scale to exploit the data emerging from genome programmes[7].

The main distinction between dot blots and DNA microarrays is in the use of an impermeable, rigid substrate, such as glass, which has a number of practical advantages over porous membranes and gel pads[8]. As liquid cannot penetrate the surface of the support, target nucleic acids can find immediate access to the probes without diffusing into pores. This enhances the rate of hybridization, although mixing is important to achieve maximum rates of hybridization even with impermeable supports.

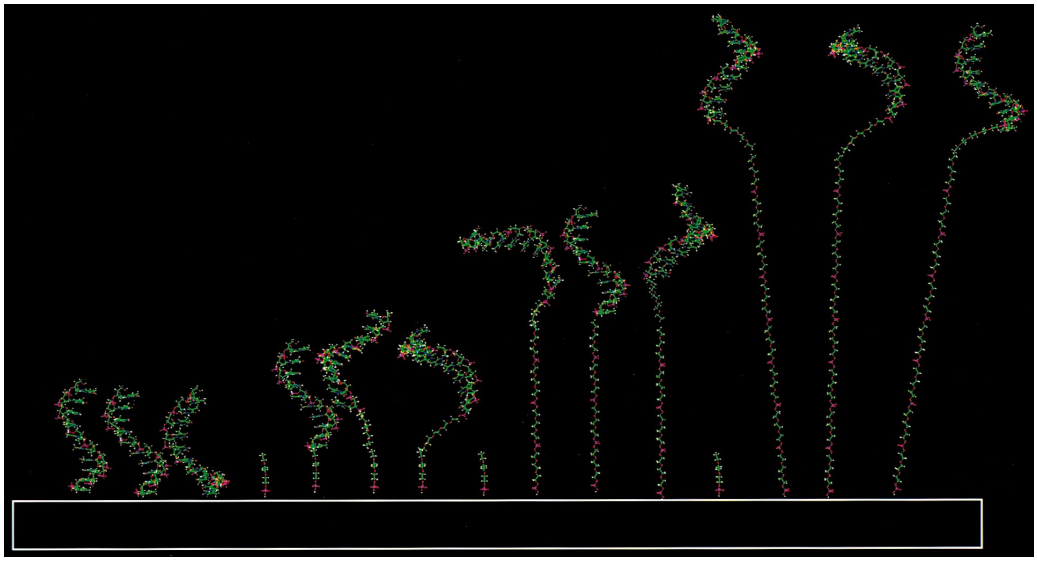

Fig. 1 The density of oligonucleotides on the surface is approximately 10 pmol per mm^2 on aminated polypropylene, approximately 0.1 pmol per mm^2 on glass after ammonia deprotection—-equivalent to one molecule per 39 square angstroms. The oligonucleotides are just about within reach of each other on glass, but rather closely packed on polypropylene supports. Spacers help to overcome steric interference, which can take a number of forms: the ends of the probes closest to the surface are less accessible than the ends furthest away; tethered molecules may crowd each other. Oligonucleotides on long spacers are better able to extend away from their neighbours and from the surface to allow interaction with the target. In this and other figures, the molecules are shown in a stretched conformation. It is likely that the molecules are in a dynamic state which may include this as one extreme, but in which the average state is somewhat more condensed. The linkers illustrated are oligoethylene glycols 26, 60 and 105 atoms in length.

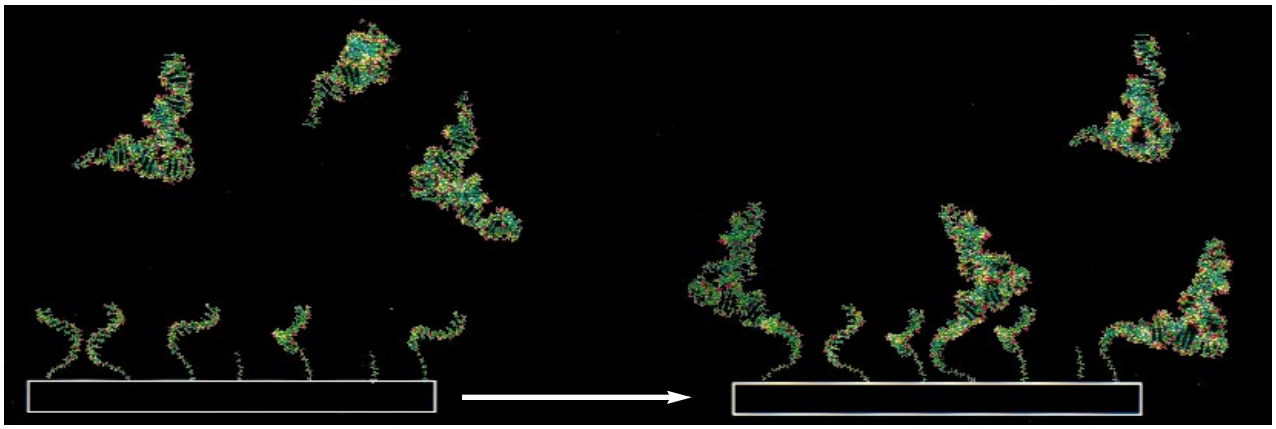

Fig. 2 Long target sequences are likely to fold in on themselves as a result of intramolecular Watson-Crick base pairing. This structure hides parts of the target from the oligonucleotide probes. Large targets are also likely to be inhibited by their bulk from approaching the surface. Illustrated here is tRNAphe in solution, hybridizing to tethered decanucleotides.

The washing step which follows hybridization is also unimpeded by diffusion, speeding up the procedure and improving reproducibility. The flatness, rigidity and transparency of glass supports improve image acquisition and image processing, as the locations of the probes are much better defined than they are on a flexible membrane; high image definition is critical for the small feature sizes which can be achieved on microarrays. Physical rigidity permits incorporation into flow cells for the automated processing which is essential for high throughput analysis. These practical benefits apply to all types of arrays, to cloned DNAs or PCR products as well as to synthetic oligonucleotides. Arrays made on an impermeable support are more readily considered from a theoretical standpoint: the kinetics of the interactions are not complicated by the diffusion of solvent and solutes into and out of pores or by the multiple interactions which can occur once the target has entered a pore[9]. In addition, the molecular attachment of short probes is well defined. For these reasons, we will not discuss in detail arrays made by non-covalent binding of probes to a surface through the bases or the backbone (that is, spotting arrays) nor arrays made on porous supports, although they are useful in many applications[10]. Our focus will be on arrays of oligonucleotides bound to impermeable supports covalently, through one end.

At the present time, the main large-scale application of microarrays is comparative expression analysis[10]. Another application, the analysis of DNA variation on a genome-wide scale, is becoming an increasingly viable prospect[11]. Both of these applications have many common requirements, but differ in some important respects. For the analysis of variation, it is important that the reaction forming duplex between target and probe is able to discriminate a single mismatched base pair (see page 56 of this issue (ref. 11)); the high degree of discrimination required is possible only with short probes. Sequence discrimination is less important for the measurement of expression levels, but in this case, quantitative measurement over a wide dynamic range is important. Other applications include the characterization of molecular interactions and the discovery of effective antisense reagents[12].

Array fabrication

In situ synthesis has a number of advantages over deposition of pre-synthesized oligonucleotides. Yields are high and consistent over the surface of the support from one cell of the array to another. It permits combinatorial strategies, for the fabrication of large arrays of oligonucleotides in few coupling steps. Three approaches have been used to direct oligonucleotide synthesis to defined areas of a support for *in situ* fabrication of arrays. The photochemical deprotection method is described by Robert Lipshutz and colleagues on page 20 of this issue (ref. 13). Ink-jet delivery of nucleotide precursors to the surface[14] has been developed by a number of companies, but is not yet in commercial production. These two methods enable manufacture of 'random access' arrays; that is, the oligonucleotide in any position can have any chosen sequence. Synthesis can also be localized by

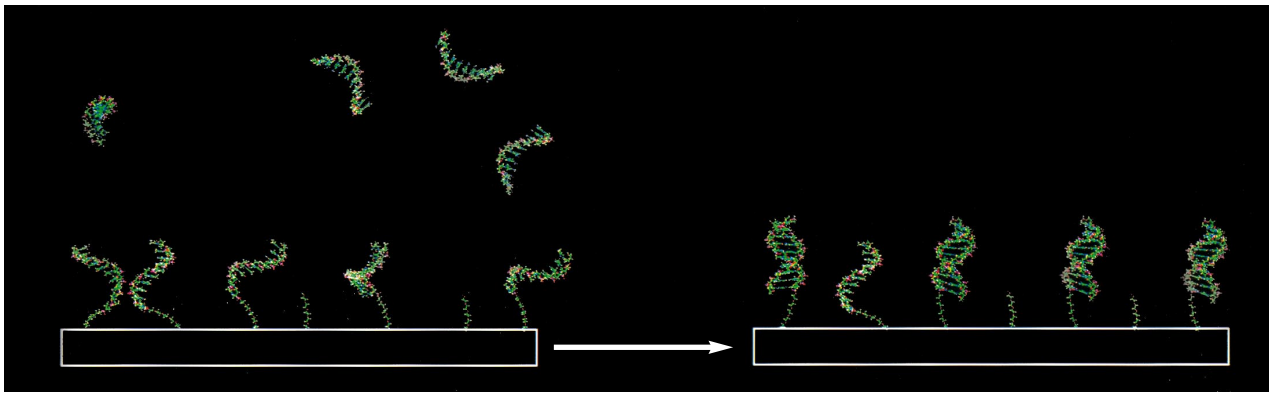

Fig. 3 Short targets are better able than large targets to interact with tethered oligonucleotides: they are less likely to have bases hidden from duplex formation by intramolecular base pairing; and, as they are less bulky, they will more readily penetrate the closely packed lawn of oligonucleotides. Ideally, target and probe should have the same length, as illustrated.

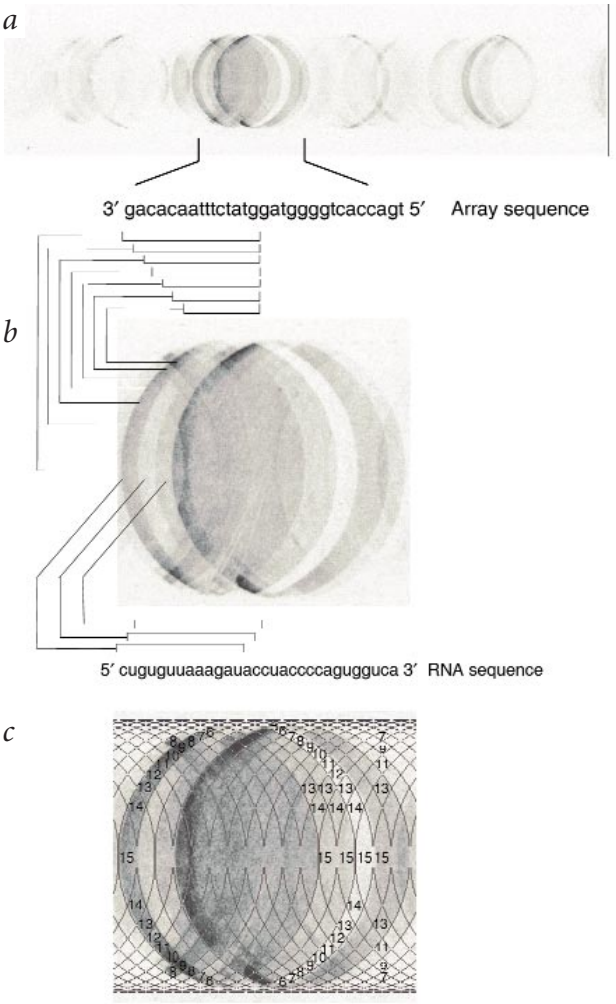

Fig. 4 A ^{32}P-labelled RNA transcript of rat *Ox40*, encoding a cell surface glycoprotein, hybridized to a scanning array in 3.5 M TMACl at 4 °C. ***a***, the array represents a tiling path of oligonucleotides complementary to the target sequence. There are sharp transitions in hybridization intensity; regions of high yield punctuate regions of low yield. ***b,*** A region of high yield is enlarged. Contrasts in hybridization yield are seen within this region. Successive positions on the array, which step through the target sequence are indicated. Arc motifs, which indicate hybridization of related sequences, are embedded in the hybridization pattern. The sequence of oligonucleotides is indicated for the first of two arcs of high intensity which contain oligonucleotides with common 5′ ends that are successively shortened at the 3′ end. Oligonucleotides in the region covering these two arcs contain a low GC content and although the longer sequences are complementary to runs of ribopurines, the highest yield on the array is in the second arc from GAUACCUA, which contains two uridines (these are normally destabilizing). An arc of low hybridization yield in which oligonucleotides have a common 3′ end is also evident. The base sequence at the common end has an inordinate influence on hybridization yield; oligonucleotides in positions shifted by one increment differ starkly in yield from oligonucleotides within the arc. ***c***, a template placed over the image shows the positions of cells containing individual sequences; yield along an arc, going from longer sequences to shorter constituent sequences, can increase and then drop again—effects that are probably due to secondary structure in the target sequence.

confining chemicals physically, for example, using masks or physical barriers[15]; by this means, complex arrays comprising many different, related sequences can be made in a few coupling steps by combinatorial methods. Flooding the precursors through orthogonally intersecting channels has been used to make arrays of all sequences of a chosen length[16,17]: circular or diamond shaped reaction chambers are used to make 'scanning' or 'tiling-path' arrays by applying the precursors in a series of overlapping areas on the surface of the support[18]. These arrays have been useful for the studies of hybridization behaviour.

It is difficult to assess the quality of the oligonucleotides made on a surface. The amount of material—approximately 10 pmol per square millimeter at densest packing—is small. However, analysis of oligonucleotides made on cleavable linkers suggests a high quality (S. Case-Green, pers. comm.). Non-destructive measurements can be made by ellipsometry or interferometry[19], methods recruited from the field of materials science. These techniques could be used for routine quality control, but are not available to most biology laboratories. Pre-synthesized oligonucleotides, on the other hand, can be assessed before they are attached to the surface[20], but it is not presently economical to make large arrays in this way. When large numbers of arrays with the same probes are needed, deposition may be more economical than *in situ* synthesis. Deposition is also the method of choice for long sequences which are available as PCR products. The technology for making spotting arrays is more accessible than that for *in situ* fabrication (see pages 10 (ref. 21) and 15 (ref. 22) of this issue).

Effect of support on duplex yield

Oligonucleotides cannot be coupled directly to the surface silanol groups of silicate glass or to most plastics. It is necessary to functionalise the surface with a group from which to initiate the growth of the oligonucleotide chain. Oligoethylene glycols have been used for glass[22], and polypropylene is readily aminated in a plasma discharge[23].

Tethering one end of an oligonucleotide to a surface is expected to affect formation of duplex with a target in solution. The bases nearest the surface are less accessible than those furthest away. The packing of oligonucleotides made by synthesis *in situ* on a glass or polypropylene surface is high: so high that there is steric crowding. Fortuitously, the ammonia used to deprotect the bases appears to dissolve enough from the surface of glass to obviate steric hindrance but leaves enough to yield good hybridization signal[19,24]. Ammonia does not remove oligonucleotides from polypropylene, however, and in this case there is evidence of an effect of steric crowding on the interaction between substrate and probe. Hybridization yields are increased up to two orders of magnitude by introducing spacers between the surface and the oligonucleotides (Fig. 1). Spacer length has a marked effect[21,25]; interestingly, there is an optimum length beyond which hybridization yield declines, presumably because the oligonucleotides 'dissolve' in the linker and become less accessible to the target. Clearly, there is scope for further studies of materials for array fabrication, in particular, for improvements to the attachment chemistry.

Effects of base composition and sequence

Base composition has a large effect on duplex yield in the solvents normally used for hybridization. The effect is undoubtedly due to lower stability of A:T versus G:C pairs. Short oligonucleotides may have extreme biases in composition and oligonucleotides of the same length have correspondingly large differences in T_m. As a rough rule, adding an A:T base pair increases T_m by 2 °C, compared with 4 °C for a G:C pair.

Arrays offer rich opportunities to analyse large numbers of sequence interactions. In a study using an array comprising all 256 octapurine sequences and a target comprising decanucleotides with the composition $A(C,T)_8A$ (ref. 17) it was found that high concentrations of TMACl had a large effect on hybridization yields. In 1M NaCl, hybridization to A_8 was barely detectable under conditions where G_8 gave good duplex yield. But in 3–4 M TMACl, yields of all duplexes increased, with the

greatest effect on A_8; in this solvent, the largest difference in yield was around fivefold. The study also showed that terminal G:C pairs produce significantly increased yields: $GN_6G > GN_6A \sim AN_6G > AN_6A$ for octanucleotides of the same overall base composition. This study and others suggest that sequences of the same composition but different sequence give different yields. Sequence effects are expected, as it is known that base stacking interactions, which depend on nearest neighbours, significantly affect duplex stability. More subtly, it has been shown that unpaired bases, which stack on the end of the duplex in the case when the target overlaps the probe oligonucleotide, may have a strong effect on yield[26]. While duplex stability is a determinant of duplex yield, the large magnitude of the differences in duplex yield and other evidence suggest that it is not the main factor.

Duplex formation and folding in the target

Duplex formation is reversible. However, even for relatively short duplexes, 10-mers or longer, the rate of strand separation is very low under hybridization conditions used with arrays (for example, 1 M NaCl or 3 M TMACl at temperatures below 30 ºC (S. Case-Green, pers. comm. and refs 27–29). Thus, differences in yield are not due to differences in overall stability of duplexes but to different rates of forward reaction. To find the underlying cause of the dramatic differences in duplex yield arising from different regions of the target[12,18,30], we must look into the mechanism of duplex formation. It is highly unlikely that all base pairs of the duplex form simultaneously: it is much more likely that the process begins by the formation of a transient nucleation complex from the interaction of very few base pairs[31]. Duplex formation proceeds, one base pair at a time, through a zippering process. At any point the reaction may go in one of two directions—pairing or separation: if bases are complementary and freely available for pairing, duplex formation is more likely to proceed; if bases are non-complementary or a stable structure inhibits base pair formation, the block to the zippering process may drive the nucleation complex to fall apart. Duplex formation, and hence duplex yield, will be determined by the stability of the nucleation complex and of intermediates up to the point in the zippering process where the likelihood of strand separation is negligible. It is evident that a number of factors must contribute to these early interactions, but there have been few systematic studies.

When the probes on the array are short and the targets are long (Figs 2,3), as is the case for most applications, initiation must begin at internal sites in the longer strand. The availability for nucleation of sites in the target will be determined by secondary structure, as intramolecular base pairing is stable under the non-stringent conditions used to hybridize to oligonucleotides. The hybridization behaviour of a tRNA with known tertiary structure is consistent with this view. Only four regions are open to duplex formation (ref. 30; K.M., unpublished observations). Each strong heteroduplex takes in one side of a stem; significantly, in all four cases, there are unpaired bases stacked on the end of the stem. The mechanism of duplex formation suggested by this result is one in which nucleation begins at the unpaired bases and propagates into the stem by strand displacement. The significant structural feature seems to be that the bases incorporated into the heteroduplex are already in a helical conformation in the native tRNA. Heteroduplex formation involves minimal perturbation of existing structures.

Effects of target structure (Fig. 4) are unlikely to be relevant to hybridization to spotted clones or PCR products as these are carried out under more stringent conditions, which should melt most secondary structure. However, these theoretical considerations of mechanism have significance for the practical application of hybridization to oligonucleotide arrays.

Minimizing secondary structure of targets

Most analyses are directed to complex targets, for example, human genomic DNA. In general, it is preferable to reduce sequence complexity to produce good hybridization signal within a reasonable hybridization time. Consequently, amplification by PCR is a standard part of target preparation. A favoured procedure for producing single-stranded targets from PCR products is to include a promoter for a RNA polymerase in one of the primers, from which RNA is transcribed[15]. However, RNA has stable secondary structure which can interfere with hybridization, as we have seen. Steps must be taken to reduce these effects, such as fragmenting the RNA, preferably to a size close to that of the oligonucleotides on the array[13]. Secondary structure is less of a problem with DNA targets and PCR products can be made single-stranded by treatment with exonucleases if one of the primers is blocked by incorporating resistant groups such as phosphorothioates[32] or a dendritic cap[33].

Polymerase and ligase extension

DNA polymerase and ligase enhance and complement the discriminating power which can be achieved by hybridization alone. It is well known that mismatches close to the centre of an oligonucleotide have a strongly destabilising effect on the duplex; mismatches at the ends are less destabilising and thus more difficult to discriminate by hybridization. Polymerases and ligases, by contrast, are affected more by terminal than internal mismatches. Polymerase uses the tethered oligonucleotide as a primer in an extension reaction in which the subtrate is a dideoxy-nucleotide triphosphate (ddNTP). The enzyme incorporates and extends only one base—that which is complementary to the next base in the target. This process is known as minisequencing[34], or genetic bit analysis[32]. Related methods that use ligase have been developed[35,36] and are being adapted for use with arrays. Heat stable enzymes help to overcome the problems which arise from carrying out the reactions at low temperature, conditions which we have seen encourage formation of intramolecular folding in the target. Thermostable polymerases and ligases are able to function at high temperatures, where the oligonucleotide duplexes have short lifetimes. Evidently, the enzymes stabilise transient duplexes or act quickly enough to form product during the short life of the duplex. In this way, the enzymes drive the reaction further than is possible by hybridization alone under conditions of high stringency.

DNA arrays have proved their value as tools for large-scale nucleic acid analysis. There is no doubt, however, that many facets of the system can be improved. It is likely that better substrate materials remain to be discovered, and the nature of the interactions between probe and target remains only partially characterized. There is room for development of automated systems for making arrays and for processing them through hybridization and reading; more sensitive detection methods, especially methods that permit analysis in real-time, would extend the range of application. It seems likely that new enzymatic techniques will be adapted to array format. In short, there is a need for systematic, basic studies, for which the arrays themselves offer a powerful experimental platform[16,37].

1. Gillespie, D. & Spiegelman, S. A quantitative assay for DNA-RNA hybrids with DNA immobilised on a membrane. *J. Mol. Biol.* **12**, 829–842 (1965).
2. Ritossa, F., Malva, C., Boncinelli, E., Graziani, F. & Polito, L. The first steps of magnification of DNA complementary to ribosomal RNA in *Drosophila melanogaster*. *Proc. Natl Acad. Sci. USA* **68**, 1580–1584 (1971).
3. Birnstiel, M., Speirs, J., Purdom, I., Jones, K. & Loening, U.E. Properties and composition of the isolated ribosomal DNA satellite of *Xenopus laevis*. *Nature* **219**, 454–463 (1968).
4. Grunstein, M. & Hogness, D.S. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl Acad. Sci.* USA **72,** 3961–3965 (1975).
5. Southern, E.M. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503–517 (1975).
6. Kafatos, F.C., Jones, C.W. & Efstratiadis, A. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. *Nucleic Acids Res.* **24**, 1541–1552 (1979).
7. Lennon, G.G. & Lehrach, H. Hybridization analyses of arrayed cDNA libraries. *Trends Genet.* **7**, 314–317 (1991).
8. Khrapko, K.R. *et al.* An oligonucleotide hybridization approach to DNA sequencing. *FEBS Lett.* **256**, 118–122 (1989).
9. Livshits, M.A. & Mirzabekov, A.D. Theoretical analysis of the kinetics of DNA hybridization with gel-immobilised oligonucleotides. *Biophys. J.* **71**, 2795–2801 (1996).
10. Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J. Expression profiling using cDNA microarrays. *Nature Genet.* **21**, 10–14 (1999).
11. Chakravarti, A. Population genetics—making sense out of sequence. *Nature Genet.* **21**, 56–60 (1999).
12. Milner, N., Mir, K.U. & Southern, E.M. Selecting effective antisense reagents on combinatorial. *Nature Biotechnol.* **15**, 537–541 (1997).
13. Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R. & Lockhart, D.J. High density synthetic oligonucleotide arrays. *Nature Genet.* **21**, 20–24 (1999).
14. Blanchard, A.P., Kaiser, R.J. & Hood, L.E. Synthetic DNA arrays. *Biosensors and Bioelectronics* **11**, 687–690 (1996).
15. Maskos, U. & Southern, E.M. A novel method for the analysis of multiple sequence variants by hybridisation to oligonucleotide arrays. *Nucleic Acid Res.* **21**, 2267–2268 (1993).
16. Southern, E.M., Maskos, U. & Elder, J.K. Analyzing and comparing nucleic acid sequences by hybridization toarrays of oligonucleotides: evaluation using experimental models. *Genomics* **13,** 1008–1017 (1992).
17. Maskos, U. & Southern, E.M. A study of oligonucleotide reassociation using large arrays of oligonucleotides synthesized on a glass support. *Nucleic Acids Res.* **21**, 4663–4669 (1993).
18. Southern, E.M. *et al.* Arrays of complementary oligonucleotides for analysing the hybridisation behaviour of nucleic acids. *Nucleic Acids Res.* **22**, 1368–1373 (1994).
19. Gray, D.E., Case-Green, S.C., Fell, T.S., Dobson, P.J. & Southern, E.M. Ellipsometric and interferometric characterization of DNA probes immobilised on a combinatorial array. *Langmuir* **13**, 2833–2842 (1997).
20. Guo, Z., Guilfoyle, R.A., Thiel, A.J., Wang, R. & Smith, L.M. Direct fluorescence analysis of genetic polymorphisms by hybridization with oligonucleotide arrays on glass supports. *Nucleic Acids Res.* **22**, 5456–5465 (1994).
21. Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J. Expression profiling using cDNA microarrays. *Nature Genet.* **21**, 10–14 (1999).
22. Cheung, V.G. *et al.* Making and reading microarrays. *Nature Genet.* **21**, 15–19 (1999).
23. Maskos, U. & Southern, E.M. Oligonucleotide hybridizations on glass supports: a novel linker for oligonucleotide synthesis and hybridization properties of oligonucleotides synthesized *in situ*. *Nucleic Acids Res.* **20**, 1679–1684 (1992).
24. Matson, R.S., Rampal, J., Pentoney, S.L. Jr., Anderson, P.D. & Coassin, P. Biopolymer synthesis on polypropylene supports: oligonucleotide arrays. *Anal. Biochem* **224**, 110–106 (1995).
25. Shchepinov, M.S., Case-Green, S.C. & Southern, E.M. Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays. *Nucleic Acid Res.* **25**, 1155–1161 (1997).
26. Williams, J.C., Case-Green, S.C., Mir, K.U. & Southern, E.M. Studies of oligonucleotide interactions by hybridisation to arrays: the influence of dangling ends on duplex yield. *Nucleic Acids Res.* **22**, 1365–1367 (1994).
27. Maskos, U. & Southern, E.M. Parallel analysis of oligodeoxyribonucleotide (oligonucleotide) interactions. I. Analysis of factors influencing oligonucleotide duplex formation. *Nucleic Acids Res.* **20**, 1675–1678 (1992).
28. Wood, W.I., Gitschier, J., Laskey, L.A. & Lawn, R.M. Base composition-independent hybridization in tetramethylammonium chloride: A method for oligonucleotide screening of highly complex gene libraries. *Proc. Natl Acad. Sci. USA.* **82**, 1585–1588 (1985).
29. Jacobs, K.A. *et al.* The thermal stability of oligonucleotide duplexes is sequence independent in tetraalkylammonium salt solutions: application to identifying recombinant DNA clones. *Nucleic Acids Res.* **16**, 4637–4650 (1988).
30. Mir, K.U. Novel approaches for the analysis of nucleic acids. (D. Phil. thesis, Oxford University 1995).
31. Wetmur, J.G. & Davidson, N. Kinetics of renaturation of DNA. *J. Mol. Biol.* **31**, 349–370 (1968).
32. Nikiforov, T.T. *et al.* Genetic bit analysis: a solid phase method for typing single nucleotide polymorphisms. *Nucleic Acids Res.* **22**, 4167–4175 (1994).
33. Shchepinov, M.S., Udalova, I.A., Bridgman, A.J. & Southern, E.M. Oligonucleotide dendrimers: synthesis and use as polylabelled DNA probes. *Nucleic Acid Res.* **25**, 4447–4454 (1997).
34. Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L. & Syvanen, A.C. Minisequencing: a specific tool for DNA analysis and diagnostics on oligonucleotide arrays. *Genome Res.* **7**, 606–614 (1997).
35. Nickerson, D.A., Kaiser, R., Lappin, S., Stewart, J. & Hood, L. Automated DNA diagnostics using an ELISA-based oligonucleotide ligation assay. *Proc. Natl Acad. Sci.* USA **87**, 8923–8927 (1990).
36. Landegren, U., Kaiser, R., Sanders, J. & Hood, L. A ligase-mediated gene detection technique. *Science* **241**, 1077–1080 (1988).
37. Fotin, A.V., Drobyshev, A.L., Proudnikov, D.Y., Perov, A.N. & Mirzabekov, A.D. Parallel thermodynamic analysis of duplexes on oligodeoxyribonucleotide microchips. *Nucleic Acids Res.* **26**, 1515–1521 (1998).