Microarray technology: beyond transcript profiling and genotype analysis

Jörg D. Hoheisel

Abstract | Understanding complex functional mechanisms requires the global and parallel analysis of different cellular processes. DNA microarrays have become synonymous with this kind of study and, in many cases, are the obvious platform to achieve this aim. They have already made important contributions, most notably to gene-expression studies, although the true potential of this technology is far greater. Whereas some assays, such as transcript profiling and genotyping, are becoming routine, others are still in the early phases of development, and new areas of application, such as genome-wide epigenetic analysis and on-chip synthesis, continue to emerge.

No other methodological approach has transformed molecular biology more in recent years than the use of microarrays. Microarray technology has led the way from studies of the individual biological functions of a few related genes, proteins or, at best, pathways towards more global investigations of cellular activity. The development of this technology immediately yielded new and interesting information, and has produced more data than can be currently dealt with. It has also helped us to realize that even a 'horizontally exhaustive' molecular analysis is insufficient. A simultaneous 'vertical' molecular and structural analysis is a prerequisite for the elucidation of the complex and interrelated processes that occur in biological systems.

From its origin as a new technique for large-scale DNA mapping[1] and sequencing[2] and its initial success as a tool for transcript-level analyses[3], microarray technology has spread into many areas by adaptating the basic concept and combining it with other techniques. For example, a more detailed analysis of genomic DNA has become possible with respect to sequence, copy number, identification and characterization of protein binding sites, structural variations and nucleotide modifications. At the same time, microarrays also have the potential to become a device for manufacturing purposes. However, although more mature applications are on the verge of becoming routine, others still have to pass the initial test of time (TABLE 1).

Here I focus on the range of applications that have been developed using the microarray platform and discuss their potential for functional analyses. In addition, I contemplate future directions for microarray technology, which are no longer determined by technical advances but by experimental requirements that derive from the objective of the assay.

Towards routine

Transcriptional profiling. To many, the term microarray analysis is equivalent to transcript analysis. Although transcriptional profiling is unquestionably the most widely used application at present, it might become less important in future because it focuses on a biological intermediate. RNA might be too volatile and prone to producing artefacts to be used for diagnostic purposes; furthermore, for identifying therapeutic approaches, RNA might be too far removed from the actual cellular effectors, most of which are proteins. However, this might be different for regulatory RNAs, such as microRNAs[4,5].

Although RNA profiling represents an advanced microarray application, many issues remain to be dealt with. Recent analyses have demonstrated that good reproducibility can be achieved across laboratories and platforms. The main factors that influence variation are the biological samples and human factors, rather than technical diversity[6–8]. For example, as long as relatively brief hybridizations to array platforms with spots that differ in size by several orders of magnitude are expected to yield similar results — which they cannot, for good biophysical reasons that relate to kinetics and mass transport — scientists are insufficiently prepared for the routine use of the technology.

Division of Functional Genome Analysis, Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 580, 69120 Heidelberg, Germany.
e-mail: J.Hoheisel@dkfz.de
doi:10.1038/nrg1809

Table 1 | **Status of microarray-based processes**

Process	Status*
Transcriptional profiling	Mature, but still to be improved
Genotyping	Mature, but still to be improved
Splice-variant analysis	In progress
Identification of unknown exons	Early stages
DNA-structure analysis	Pilot phase
ChIP-on-chip	In progress
Protein binding	Under development
Protein–RNA interaction	Idea
Chip-based CGH	In progress
Epigenetic studies	Under development
DNA mapping	Mature
Resequencing	In progress
Large-scale sequencing	Under development
Gene/genome synthesis	Early stages
RNA/RNAi synthesis	Pilot phase
Protein–DNA interaction	Under development
On-chip translation	Under development
Universal microarray	Under development

*From most to least developed: mature, in progress, under development, early stages, pilot phase, idea. CGH, comparative genomic hybridization; ChIP-on-chip, on-chip chromatin immunoprecipitation.

Another issue relates to the analysis of results from microarray experiments. The assumption that changes of more than twofold are significant is still surprisingly widespread. This threshold is based on initial publications by the Stanford group who found, from concordance analysis, that a more than twofold variation was significant for a particular set of experiments[9]. This factor of two was subsequently referred to by others as a universal significance threshold, without realizing how it came about. Limitations to data analysis (BOX 1) continue to be frustrating, although basic standards have been established[10]. However, even generally accepted analysis tools have not necessarily been firmly validated. For example, LOWESS[11] — a normalization process — is used to correct a frequently observed curvature in regression plots that is attributed to an assumed dependence of the normalization factor on signal intensity. However, there is no proof that this assumption is correct.

There is much to be improved with regard to data interpretation. A lack of change in the level of a transcript is just as important as no variation, especially in view of a systemic analysis. Today, this information is largely disregarded in data interpretation.

Genotyping. Genotyping technology has been instrumental to producing a high-resolution genetic map of the human genome (and others), yielding DNA blocks (haplotypes) that could help in the localization of disease genes by identifying those blocks that are associated with phenotypic traits (see the International HapMap Project web site)[14]. It can also be used to identify microrganisms, a process that is important in medicine, food production or screening of the microbial content of waste-water treatment plants.

The use of microarrays for genotyping is technically further advanced than for transcript profiling. One reason for this is that a qualitative analysis is usually adequate for genotyping: simple discrimination is sufficient to detect individual base differences (SNPs). Genotyping experiments also generally contain direct internal controls, which enable better data interpretation. For example, if detection is done by hybridization of a sample DNA to an oligonucleotide array, three of the four oligonucleotides that represent a SNP function as mismatch controls. Compared with an assay that is based on the analysis of signal intensities after mere hybridization, better discrimination can be achieved by a primer-extension reaction on oligonucleotides that bind adjacent to the base in question. Labelled dideoxynucleotides that are complementary to the base on the opposite strand are incorporated during the extension reaction. The combination of the discriminative effect of

Box 1 | **Data analysis and interpretation**

Techniques that are involved in microarray production have advanced considerably and have achieved a sufficient quality that they are no longer a limiting factor. However, data analysis is still lagging behind, although it is gathering momentum. Although basic image analysis is well advanced, there are no standards at the level of filtering, which is done according to the researcher's experience, leading to discrepancies at this early stage and preventing a high degree of reproducibility. As a consequence, data comparison is difficult. Normalization, a process that adjusts microarray data for effects that arise from variation in the technology rather than from biological differences, is another important early step in the analysis process. Fortunately, continuous progress in normalization issues is being made[12,13]. Only with the establishment of commonly accepted protocols and routines will a better cross-evaluation become feasible. The Microarray Gene Expression Data (MGED) Society is pushing towards common protocols for transcript analysis. This effort needs to be expanded to other areas of microarray-based analysis.

Apart from quality issues, data interpretation is currently the main bottleneck in microarray analyses. In particular, the automated integration of complementary information in analysis algorithms is not yet well established. In part, this is because a common nomenclature has been lacking and data are not stored in a format that can be easily queried. The Gene Ontology Consortium and similar initiatives have taken on the crucial task of providing such a common framework. Although not intrinsic to microarray technology, these efforts are fundamental for the success of the technology. Open-source initiatives such as Bioconductor provide a means for developing, testing and disseminating new algorithms. Comprehensive expert systems that carry out data interpretation automatically are under development but are unlikely to be available in the near future, even for the most commonly used techniques.

hybridization with the base-pairing specificity of polymerases[15,16] leads to an improved base-calling accuracy. Probably the most precise currently available assay is based on a continuous detection of the hybridization process (dynamic allele-specific hybridization)[17]. The association and dissociation curves of the duplexes that form following hybridization allow more accurate SNP detection. Once such a dynamic measurement over time is possible with highly complex microarrays, accuracy and throughput will not be limiting; instead, sensitivity and sample preparation will. Both these limitations could be overcome by label-free detection; for example, this can be achieved by using secondary-ion mass spectrometry to identify the phosphates of hybridizing nucleic acids on microarrays that are made of peptide nucleic acid (PNA) oligomers[18], by detecting mass changes that introduce an alteration in the vibration frequency of arrayed cantilevers[19], which in turn translates immediately into electronic signals, or by using conductance and impedance measurements[20].

Sophisticated adaptations

Expanding RNA studies. An important realization that came from genome-sequencing studies was that it is the molecular interpretation of the encoded message, and not so much the number of genes, that is responsible for the variation between organisms. At the level of RNA, splicing differences contribute to inter-individual variation. The transcriptional representation of exons can be analysed by placing at least one representative (PCR fragment or oligonucleotide) of each onto an expression microarray. The ability to carry out such combined analysis of differences in splicing and transcript levels usually depends on the quality of the genome sequence annotation. Even in the rather simply structured genome of *Saccharomyces cerevisiae*, several hundred ORFs were initially missed, and in *Drosophila melanogaster*, the initial underestimation of the number of genes was even greater[21]. Non-coding transcribed sequences are even more likely to be missed. Although the number of protein-coding genes in the human genome that have gone unnoticed could be relatively small, there might be many exons that are unaccounted for. Intron–exon junction sequences can be used as an alternative to probes that represent exon sequences[22].

Whereas the lack of hybridization signals at the corresponding oligomers easily identifies exons that are false positives from annotation efforts, the identification of unknown exons requires extra steps. One option is to synthesize oligonucleotides that correspond to the sequences at the exon–intron boundaries with their 5′ ends attached to the chip surface. A polymerase extension reaction is carried out on hybridization of the RNA or cDNA. The reaction also includes a mix of primers that are complementary to the next annotated exon, although it might also be possible to attach the second primer to the chip surface next to, or mixed with, the first one. For each RNA/cDNA–oligomer duplex, on-chip PCR yields a defined fragment that is attached to the support. The length of the fragment can be determined by scoring the signal intensities obtained in a secondary hybridization with a labelled heptamer library or by adding a DNA-specific dye (FIG. 1). Compared to given standards, the signal intensities will indicate DNA length, thereby revealing any extra DNA if there is an unknown exon between the two exonic sequences that were used for priming.

Instead of using annotation-based DNA fragments, the entire genome could be represented on a microarray. Whereas for a microrganism a tiling path of oligonucleotides can be sufficient for complete coverage, only portions of a mammalian genome can be covered[23]. By definition, a complete coverage would represent all the transcriptionally active units. For large genomes, a complete genomic representation is currently only feasible in the form of fragments in the kilobase range[24].

Apart from changes introduced by splicing variation, more subtle differences in sequence or even in the speed of transcription might also have considerable consequences on the folding of RNA, and therefore its activity. Based on analyses — initiated by Edwin Southern and colleagues[25] — of the influence of RNA structure on its

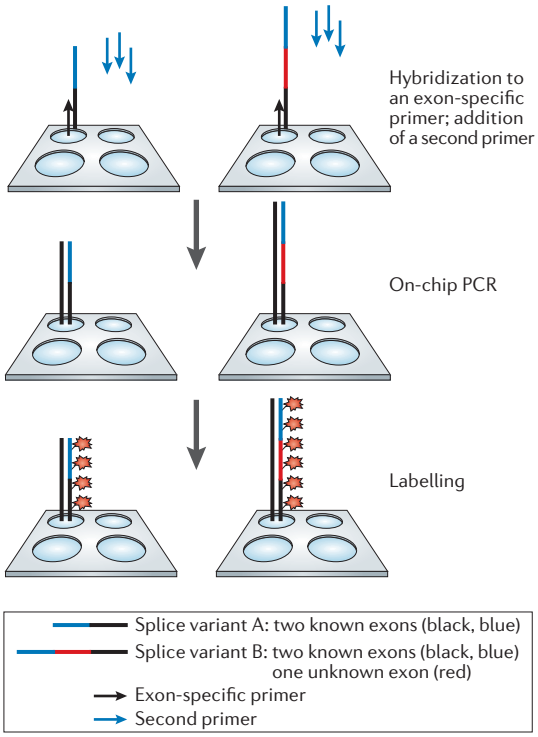

Figure 1 | **Scheme for identifying unknown exons.** RNA or cDNA molecules are hybridized to an array-bound exon-specific primer oligomer. A second primer that is known to bind to the next annotated exon is added and on-chip PCR is carried out. In an alternative approach, the second primer is also present on the microarray surface. The binding of a dye to the dsDNA or ssDNA created by the amplification process indicates, by the signal intensity, whether an unexpected piece of DNA (an unknown exon) is contained in the newly synthesized molecule.

Secondary-ion mass spectrometry
A technique in which a focused ion beam is directed to a solid surface, removing material in the form of neutral and ionized atoms and molecules. The secondary ions are then accelerated into a mass spectrometer and separated according to their mass-to-charge ratio.

Tiling path
The coverage of a given genomic region by a set of overlapping DNA fragments.

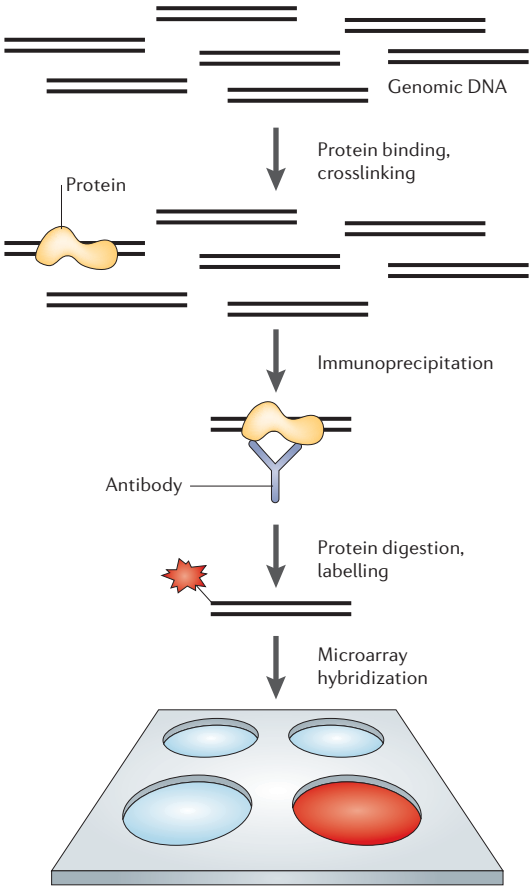

Figure 2 | **On-chip chromatin immunoprecipitation (ChIP-on-chip).** Genomic DNA is incubated in the presence of protein (yellow). DNA–protein complexes are stabilized by crosslinking. The protein is then precipitated using an appropriate antibody. Subsequently, all protein is removed by digestion and the DNA molecules are labelled. On hybridization to a representation of the genome (represented on a microarray), the origin of the precipitated DNA-fragments can be deduced from their binding positions (indicated by the red spot on the array).

binding to oligonucleotides that represent a tiling path of the entire RNA sequence, structural variants of RNA molecules (which are themselves active, for example, as ribozymes or as part of structural cellular components) could be scored. Any change in structure is likely to bring about a transformation in activity or functionality.

Probing with genomic sequences. The interaction between proteins, such as transcription factors, and DNA is important in many analyses of transcription. In on-chip chromatin immunoprecipitation[26] (ChIP-on-chip) (FIG. 2), a protein is crosslinked to DNA on binding. The protein is then used as a tag to pull out the bound DNA by means of an antibody. Hybridization of the isolated DNA to a genomic tiling path of DNA fragments identifies its position, and therefore the genomic position of the protein binding site. The specificity of the protein–DNA interaction and the epitope recognition by the antibody are crucial factors in ChIP-on-chip, although many other factors also influence its outcome, such as the crosslinking reagent[27].

A recent study in yeast demonstrated how ChIP-on-chip can yield functional information[28]. The study revealed that the general transcription machinery is present but largely inactive in quiescent cells. RNA polymerase II is bound upstream of numerous early response genes that become activated as soon as the cell is released from the stationary phase.

The interest in the interactions between proteins and RNA is likely to increase, given the growing realization that RNAs have more functions than have until now been appreciated. Technically, all analyses that are carried out with DNA are feasible at the level of RNA also.

Another assay format that uses probes derived from genomic sequences is comparative genomic hybridization (CGH) (FIG. 3), a method that is used to analyse variations in DNA copy number, which can vary significantly on a regional basis. This variation can be measured by hybridizing a labelled genomic DNA sample to a genomic array. An increase or decrease of the signal indicates areas of deletion or amplification, respectively. Cancer-related variations were the first to prove the direct connection between such changes and disease[29] and are used in diagnosis[30]. Although it is likely that copy-number variation is used as another mechanism of regulation in normal cells, there is little evidence to support this hypothesis. Many changes are likely to be too locally confined to be picked up on metaphase chromosomes, but clone-based microarray equivalents significantly improve the resolution and ease of such assays[31]. BAC-based arrays have proved highly effective in defining the location of copy-number changes, frequently within a range of about a megabase. More precise positioning is possible with microarrays that contain BAC contigs that are tiled across a locus of interest. Such continuous coverage has even been achieved for the entire human genome[32]. cDNA microarrays can also be used for this purpose, although the density of genes varies significantly across a genome[33]. Labelling techniques and the quality of the arrayed oligomers have improved markedly over the past few years, so that even short oligonucleotides can be sufficient for high resolution in complex genomes[34,35]. For kinetic reasons, however, a reduction of the genome complexity is required before hybridization to such arrays[36].

Epigenetic studies. About 4% of the cytosine residues in the human genome are modified by methylation at their C5 position. This epigenetic programme provides an important level of regulation of gene expression, adding another dynamic feature to the frequently presumed static DNA that regulates the interpretation of genetic information. It is generally assumed, for example, that epigenetic misregulation of cancer-related genes has an important role in cellular transformation. Owing to its exceptional biochemical stability, DNA methylation is a particularly attractive biomarker. Microarray-based technologies provide an opportunity for a genome-wide determination of methylation

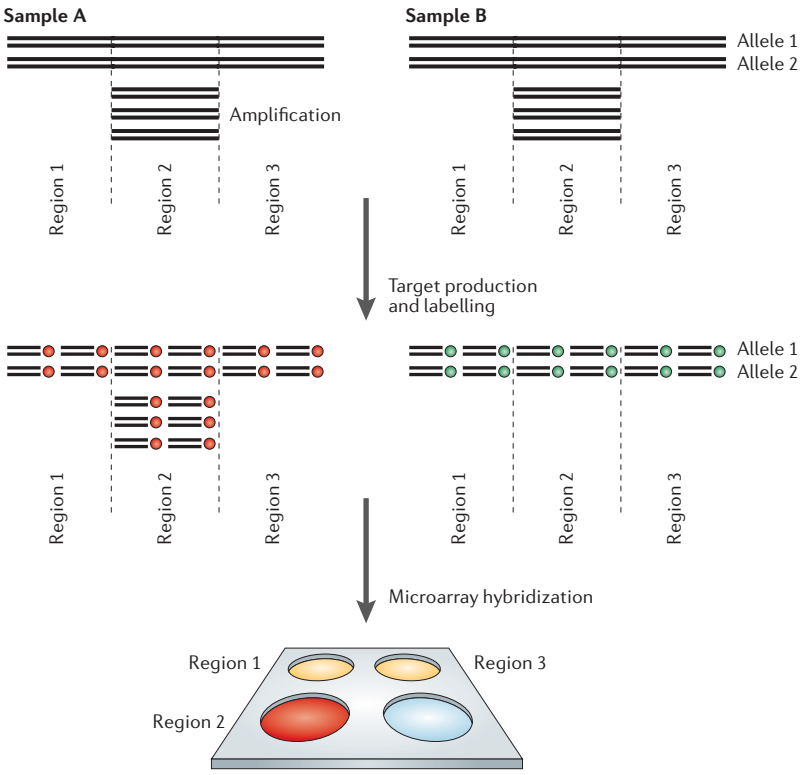

Figure 3 | **Comparative genomic hybridization (CGH).** Two genomic DNA samples (sample A and B) are labelled with different fluorophores (red and green). If a particular region is present in more copies in one of the two samples, equivalently more labelled fragments are generated. The difference in copy number can be detected by a competitive hybridization of both samples to an array of genomic fragments. In this example, a stronger red signal component is yielded at region 2.

patterns with a locus-specific resolution. Based on the bisulphite conversion of unmethylated cytosine to uracil and, following PCR, eventually to thymine, methylation patterns represent chemically introduced 'SNPs' and can be analysed accordingly using different microarray-based approaches. In one approach, DNA is labelled and hybridized directly to arrays of either CpG-island fragments[37,38] or oligonucleotides[39,40]; the former assumes that all CpG dinucleotides within an island have the same methylation status, which might not always be the case[40]. The latter requires a careful selection of oligonucleotides, because the complexity of the genomic sequence is reduced to essentially three bases, with most cytosines being converted into thymines. In addition, the CpG dinucleotides at which almost all methylation occurs are frequently packed so densely that a given oligonucleotide might correspond to more than one CpG dinucleotide. As an alternative, enzymatic primer-extension reactions could improve the analysis in terms of throughput and resolution, as discussed above for standard SNP typing.

'Old' techniques revisited

Microarray technology was initially developed for DNA-mapping[41] and sequencing-by-hybridization (SBH)[42–44] applications.

In DNA-mapping applications, arrayed DNA fragments were hybridized to several different oligonucleotides. The fragments could be ordered on the basis of their individual hybridization fingerprints. Initial functional information could also be assigned at the same time if it was encoded in the DNA sequence (for example, retroviral sequences, *Alu* repeats and exon–intron boundaries)[45]. With so many organisms (especially miroorganisms) being sequenced, and with the ability to produce complex mixtures of oligonucleotides in small quantities and at low cost, but at concentrations that are sufficient for molecular biology (see below), this kind of analysis could be expanded and refined to the metagenome level[46]. Instead of being sequenced, DNA fragments from environmental isolates (for example, soil[47] or water[48]) can be placed on microarrays for a comparative analysis[49]. By hybridizing mixtures of short oligomers, individual fragments can be attributed to a particular microbial family or their cellular function can be identified by the degree of their homology to sequences from characterized genomes. At the same time, new microorganisms can quickly be related to the phylogenetic tree by such means. Because most fragments would have come from microorganisms, the arrayed DNA would mainly consist of coding sequence, allowing a global transcriptional analysis and comparison of metagenomes without the need to have previously identified its components. Ultimately, the sequence of each fragment can be inferred from hybridization information[50].

(Re)sequencing. Sequencing can be thought of as an extended form of SNP typing. Technically, the only difference is that for sequencing comprehensive coverage of a DNA fragment is required. Initially, a comprehensive octamer library of some 65,000 oligonucleotides was thought to be optimal for this purpose, especially if each oligonucleotide was arranged in a tandem of tetramer oligonucleotides, split by a dimer of nonspecific base composition[51]. A read-length of more than two kilobases was possible in principle. Currently, however, arrays that contain oligonucleotides of some 20 nucleotides in length are used[52,53]. With the advent of techniques that allow the flexible production of complex oligonucleotide microarrays, such as micromirror-based, photocontrolled *in situ* synthesis[54,55] (BOX 2), any region can be resequenced in comparison to a given standard.

As with genotyping, primer extension is another option in resequencing. Arrayed oligonucleotides are extended on hybridization to the target DNA that serves as a template. Nucleotides that carry a base-specific fluorescent label cannot be extended as long as the label is attached. After scoring the type of fluorophore that is incorporated in one round of extension, the label is removed and the molecules are extended by another nucleotide[56]. By analysing many individual molecules, a large amount of data could be generated in a single experiment[57]. The resulting sequence reads would be about 25-nucleotides long and could be assembled according to existing sequence information for the identification of differences, such as mutations.

Serial analysis of gene expression
A method for analysing transcription patterns. A short cDNA tag sequence of 10 to 14 bp is isolated for each transcript. They are linked at random to form long concatemeric molecules that can be sequenced to determine the frequency of each tag sequence, and therefore the respective RNA, in the entire population.

Solid-phase synthesis
A chemical synthesis reaction during which the synthesized molecules are continuously attached to a solid support medium.

Phosphoramidite chemistry
The chemistry of choice for oligonucleotide synthesis; the stable tri-coordinated phosphorous function of one nucleoside phosphoramidite is activated by a weak acid and reacts with the hydroxyl moiety of another nucleoside.

RNAi
RNA-mediated, sequence-specific transcriptional silencing of gene expression.

Box 2 | **Microarray production**

Consistent quality in array technology is now possible and ensures that a high degree of reproducibility can be achieved within and between platforms. Two basic processes are used in microarray manufacture: *in situ* synthesis and spotting of pre-produced molecules; they can be used individually or in combination. *In situ* synthesis is either controlled by light, electro-chemistry or by a directed application of the monomer phosphoramidites (for example, by piezo elements). For large-scale production, the photolithographic approach is currently the gold standard in terms of reproducibility. No other currently established technique can compete, especially with respect to the number of oligonucleotides per chips. However, depending on the protection chemistry used, the achievable length of the oligonucleotide is limited. Photolithography also lacks the flexibility of creating new content designs, which is inherent to other processes. Using a micromirror device instead of masks introduces this flexibility into photo-controlled synthesis. Attachment of pre-existing molecules, usually long oligonucleotides or PCR products, is still a valid alternative for many purposes, and provides the freedom of operation that could be advantageous in some areas.

Massively parallel signature sequencing (MPSS)[58] is another option for obtaining sequence information in a highly parallel manner. Instead of cycles of nucleotide incorporation, it involves 4 to 5 cycles of cleavage of target DNA with a type II restriction enzyme that produces a 4-nucleotide protrusion. Sequence-specific ligation of a fluorescent linker follows, so that the identity of the overhang can be queried (see REF. 59 for a recent review of high-throughput sequence-analysis techniques).

All sequencing approaches can be adapted for other purposes. such as transcriptional profiling to count the presence of individual cDNA fragments in a process that is equivalent to serial analysis of gene expression (SAGE)[60].

On-demand synthesis

From probe production to gene synthesis. Automation of DNA synthesis has benefited from the combination of solid-phase synthesis and phosphoramidite chemistry[61]. Today, standard DNA synthesizers produce oligonucleotides in the nanomole to micromole range. Oligomer synthesis for biomedical purposes has been dominated by two diametrically opposite tendencies. For applications such as oligomer-based therapy, gram or even kilogram amounts of only a few oligonucleotides are required; in molecular biology, quantities in the picomole to femtomole range are usually adequate, but large numbers of different oligomers are needed. To meet this demand, procedures that are aimed at the simultaneous production of different oligonucleotides in small quantities have been developed[62,63]. Early experiments showed that oligonucleotides of sufficient quality and quantity could be synthesized on microarray surfaces and released subsequently for applications in molecular biology[64]. Recent work has extended this technique to synthesizing whole genes and optimizing their translation efficiency by sequence alterations[65–67]. Programmable light-controlled *in situ* synthesis in microfluidic devices seems to be most suited to this application[54]. The technique is expanding, because near-quantitative yields can be achieved[68] and the direction of chemical synthesis can be selected[69]. Therefore, either the 5′ or 3′ ends of the produced oligonucleotides are defined accurately, because synthesis starts there. If synthesis proceeds from the 5′ end to the 3′ end, the 3′ termini of all shorter derivatives, which result from the low proportion of unextended molecules in each synthesis cycle, can be blocked selectively so that only full-length molecules remain reactive in subsequent polymerase reactions, thereby defining a common 3′ end for all reactive molecules.

Yielding RNAi. An increasing number of eukaryotic genes are being found to have naturally occurring antisense transcripts[70]. RNAi is a useful tool for functional analysis[71,72]. Unfortunately, the degree of transcriptional silencing varies significantly and does not rise beyond 95%. Therefore, individual molecules are frequently insufficient for effective interference. Moreover, owing to molecular compensation, even complete removal of gene function does not

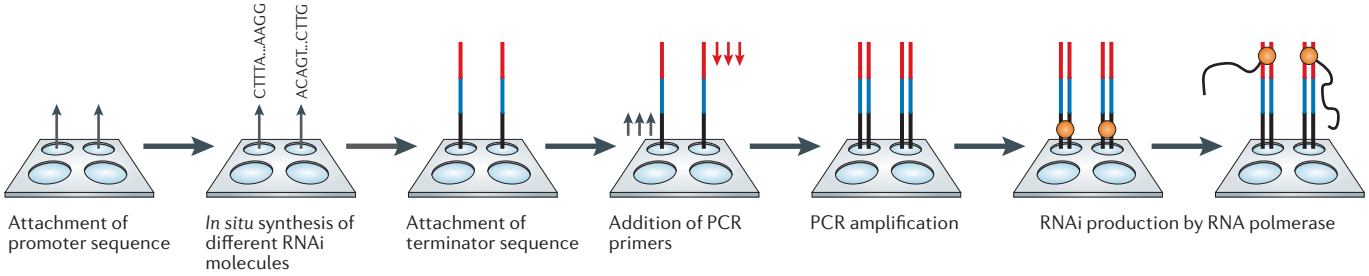

Figure 4 | ***In situ* synthesis of RNAi libraries.** Templates for RNA polymerase are synthesized on microarrays. Instead of using a primer-extension reaction, dsDNA is obtained by a back-folding reaction of the synthesized oligonucleotides, length permitting. The RNA-polymerase reaction takes place on the chip or on PCR fragments that are eluted from the microarray.

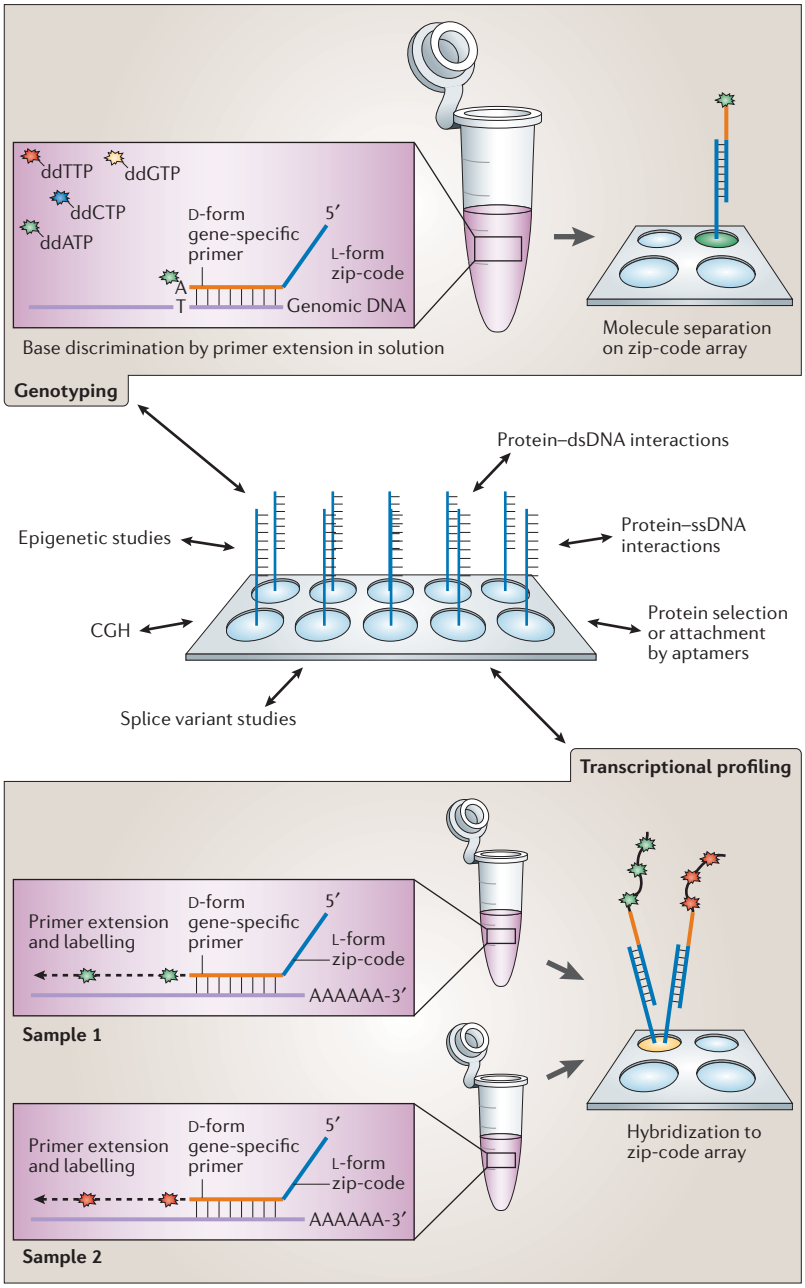

Figure 5 | **A universal chip that is based on L-DNA.** A single microarray design is used for all applications (central panel). Rather than using a solid support, appropriately designed molecule mixtures are incubated with the analyte in a homogenous solution. It is only subsequently that they are physically separated by a tag sequence, which is distinct for each oligomer and has its complement on the microarray. This experimental design allows for combinations of different analysis types to be carried out simultaneously. Some typical applications are listed but more are possible. The use of L-DNA for the 'zip-code' tag avoids background. Also, L-DNA is more stable than the natural D-form owing to the inability of degrading enzymes to use it as a substrate. More details are shown for the performance of genotyping (top panel) and transcriptional profiling (bottom panel). In both applications, chimeric oligonucleotides that consist of an L-DNA tag and a primer portion are incubated with the sample DNA or RNA in a homogenous solution. In genotyping, the complementary dideoxynucleotide is incorporated, which carries a nucleotide-specific label. For transcript profiling, standard cDNA labelling reactions are carried out on two RNA samples, incorporating either of two labels. On application to the array, the molecules are physically separated by hybridization of their respective tags and can be analysed individually. CGH, comparative genomic hybridization.

necessarily lead to a phenotypic effect . So, for increased effectiveness, the simultaneous inhibition of several transcripts can be advantageous, although inclusion of more than 100 small interfering RNA (siRNA) molecules makes the solution difficult to handle owing to its viscosity. Synthesis of comprehensive RNAi libraries is extremely expensive and requires frequent resyntheses to replace molecules that turn out to be ineffective. Instead, RNAi templates can be synthesized on microfluidic microarrays for each experiment[73]. All oligodioxyribonucleotides that are synthesized contain an identical promoter sequence that is adjacent to the complement of an RNAi molecule. In fact, the promoter fragment can be attached chemically in a single step; in this case DNA synthesis proceeds from this common fragment. At the end of the synthesis, an identical T7-RNAi-hairpinloop-RNAi-terminator-adaptor is coupled to each oligomer. Subsequently, double-stranded molecules are produced by an extension of added primers that are complementary to the terminal sequences shared between all molecules (FIG. 4). RNA is produced by an enzymatic reaction either on the chip or after elution of the PCR products. Because of the high activity of RNA polymerases, relatively large amounts of RNA can be produced. As a positive potential side effect, the molar ratios of the different RNAi molecules could be manipulated. With miniaturization of genomic and proteomic approaches proceeding at an ever faster rate[74], the yield from such RNAi synthesis could suffice for several assays and even for combinatorial approaches. Although combinatorial techniques are well established in chemistry they are not yet a firm part of biological sciences.

On-chip translation. Many proteins cannot be expressed well from *Escherichia coli* or other cell-based expression systems. Examples include proteins with intra-chain disulphide bonds, such as antibodies, that tend to denature and form inclusion bodies, and proteins that are toxic to the bacterial host. For these, and for individual protein domains, production by cell-free synthesis directly from the DNA can be advantageous. One technique that addresses these problems is protein *in situ* synthesis[75,76]. PCR products that have appropriate initiation sequences and stop codons are used to make the protein in a cell-free transcription/translation system. The proteins include a tag sequence for immobilization, and translation is carried out on surfaces that are coated with agents that bind the tag. This approach facilitates a single-step production of protein arrays from DNA without the need for protein handling or purification.

Approaching new horizons

Protein binding. Protein binding can be analysed directly using epitope-tagged proteins (obtained, for example, from fusion-protein libraries)[77] as an alternative to using ChIP-on-chip. The proteins are incubated with microarrays that contain spots of dsDNA. Once attached to their binding sites, the proteins are

detected with a labelled antibody. Many transcription factors, for example, have unknown DNA-binding specificities and regulatory roles. In combination with the flexibility of *in situ* synthesis of dsDNA molecules[78], the reaction of transcription factors on variations on a given sequence could be analysed in great detail, as the effect of many sequence permutations can be analysed quickly and at low cost.

Universal chip platform. Most microarray platforms are designed to address a specific set of questions in a specific organism. This means that a specific microarray platform needs to be established and produced for each application. Moreover, many assays that are carried out on microarrays would work even better in a homogenous solution rather than on a solid support. The establishment of 'zip-code' arrays can address these problems by separating the actual assay from the microarray hybridization[79]. Such microarrays contain a set of unique and distinct oligonucleotides that are immobilized at known locations. Because they should not be complementary to any sequence in any organism and are made solely to identify the 'address' of a particular location on the microarray, they are called zip-code sequences (FIG. 5). The oligonucleotides are designed to have similar thermodynamic properties: therefore hybridization can be carried out at one temperature and under defined stringency conditions. Instead of having to produce many different microarrays, a single design can be used for various assays. The target molecules are composed of assay-specific sequences that are linked to a zip-code complement. Reactions that involve target molecules take place in solution. Only subsequently are the molecules physically separated and therefore made available to scoring by hybridization to the zip-code microarray.

The first array of this kind was used to identify yeast mutants[80]. Instead of a deletion, each mutant contained two 20-nucleotide sequence tags. Following simultaneous PCR amplification of the tags from all the mutants that were grown in a given culture, the relative abundance of each mutant in the mixture was determined by hybridization to a microarray containing the complementary oligonucleotides. In principle, any kind of assay can be carried out along these lines, as long as the probe contains a zip-code sequence adjacent to its assay-specific part. For example, a primer extension can be carried out in solution for genotyping. Each primer terminates one base before the polymorphic site, and has a zip-code portion that is attached at the 5′ end. After incorporation of dideoxynucleotides, which is carried out with complex primer mixtures in homogenous solution, the sample is hybridized to the zip-code microarray, separating the individual molecules to known positions on the chip. This means that the incorporated label can be scored for each primer individually[81].

In another setting, padlock probes were used to identify SNPs without the need for amplifying the genomic regions of interest beforehand[82]. Only zip-code sequences of padlock probes that found a complementary sequence in as little as 1 ng of genomic DNA were amplified after binding and discrimination, and could therefore be identified on hybridization to the array.

Another option involves using the L-DNA enantiomer[83] — the mirror image form of 'normal' D-form DNA — for the zip-code oligomers (N.C. Hauser & S. Matysiak, personal communication). Because L-DNA forms a left-helical duplex, there is no cross-hybridization between L-DNA and D-DNA. However, chimeric molecules that are made of L-form and D-form stretches can be produced by standard chemistry. Therefore, D-DNA primers are produced with an L-DNA zip-code tag that binds to the L-DNA complementary oligonucleotide on the microarray. L-DNA microarrays are stable, because L-DNA is resistant to nuclease activities. Simultaneously, only the zip-code part of the molecules that is used in homogenous solution is able to hybridize to the array. Neither the D-formed primer portion nor the analyte (for example, genomic DNA or RNA preparations) will cross-hybridize with the array.

Alternative approaches to amplification — the sensitivity issue. For many applications, sensitivity is still a crucial issue. Apart from being able to detect rare molecules, the dynamic range increases with better sensitivity, allowing more accurate measurements. In many cases, amplification is required to achieve high sensitivity; therefore samples are frequently amplified before hybridization, which can introduce a bias. Carrying out amplification at the end of the process, once molecules are physically separated on the microarray, reduces competition and other effects. Optimally, a process should detect a single molecule and not involve any enzymatic step.

Rolling-circle DNA amplification (RCA)[84] is a well-established procedure. It results in sufficient sensitivity to detect a few hybridization events on glass surfaces when visualized by fluorescence. By using oligonucleotides that are tagged with a DNA primer that is complementary to a circular ssDNA, it is possible to generate (under isothermal conditions, with either linear or geometric kinetics) long ssDNA concatemers that contain many tandem repeats that are complementary to the original circle sequence. Up to 10^4 copies of the circle can be produced, which yield an equivalently enhanced hybridization signal.

Another option involves a self-assembly reaction of branched DNA molecules (Z. Li and J.H., unpublished observations). The branched DNA molecules are synthesized by standard chemical reactions, taking advantage of a symmetrical doubler phosphoramidite. These structures have a specific DNA sequence in their stem while the branches contain a common, repetitive sequence and a fluorophore at each terminus. The molecules are applied to microarrays in a second hybridization, following hybridization of the analyte. The initial binding to array-bound analyte molecules occurs through the specific stem sequence. Signal amplification occurs owing to the formation of

Padlock probes
Linear DNA molecules of 70–100 nucleotides that become circularized by DNA ligation in the presence of a target sequence that is complementary to both terminal sequences of the probe molecule.

a network of molecules through the self-assembly of the nonspecific branch sequences. No labelling of the analyte and no enzymatic reaction are required. The detection limit is a few hundred molecules.

Accurate quantification of the absolute amounts of target material that bind at the features of a microarray is crucial. At present, relative measurements are carried out in most cases. Although they are useful, determination of the absolute numbers makes an important difference to the basic understanding of biological processes. For example, it is important to know whether there are 5 and 2 copies or 5,000 and 2,000 copies of a particular molecule in diseased and normal tissue, respectively. For any antisense experiment, for example, the design depends crucially on this kind of information. Although initial algorithms for real quantification do exist, most depend on calibrated reference samples that need to be included in the analysis[85]. There are ways to avoid this[86], but they require further experimental refinement before they can be used in routine applications.

Structural analyses. By and large, structural variation remains an unexploited feature of DNA, although it could have an important effect on regulation. The twist angle of the double helix can vary from about 30° to 40° (REF. 87). Its precise structure depends on the sequence and is important to the recognition of target sites by DNA-binding proteins, for example.

DNA is also known to occur in conformations that are different from the typical right-handed double helix. The fact that alternating purine–pyrimidine sequences and especially CpG stretches — particularly methylated CpG dimers — are the DNA sequences that most easily form left-helical Z-DNA structures under physiological conditions[88] could of course be coincidence. No cellular function has been assigned to Z-DNA so far. However, only longer stretches of alternating purines and pyrimidines have been analysed for their possible function. It might be that a dinucleotide is sufficient for a functional effect of this transition in DNA structure.

DNA topology can be modified on a microarray, thereby providing a platform on which to screen for the effects of structural variations. Attaching both ends of a molecule to the microarray surface allows the introduction of turns in either direction. Even highly supercoiled DNA can be generated this way[88]. Given the capacity of DNA to memorize information in its structure rather than its base composition[89], such assays could yield new insights into DNA-based regulative effects.

Beyond nucleic acids. Many of the basic techniques that have been developed for the analysis of nucleic acids could be adapted for the analysis of other molecular entities, such as proteins[90,91], tissue samples[92] and even living cells[93] (see for example the MolTools web site). By being able to revisit individual spots[94] and re-analyse microarrays — for example, an optical detection followed by mass spectrometry — it will be possible to analyse more and more combinations of different biological components on the same chip platform.

Conclusions

Microarray technology has initiated an experimental approach that is based on unbiased sample screening and accumulation of data, preceding the formulation of hypotheses. To an extent, it has placed data production before intellectual concepts, although of course further and more detailed studies are required to confirm and refine the hypotheses that result from such studies. In this respect, biology is becoming more similar to physics. Although the value of this approach in biology is still a subject of debate, physics has clearly demonstrated its power. However, even those who are used to microarray technologies sometimes still need to dissociate themselves more fully from a hypothesis-driven view, as it is not data production but data interpretation that is still often biased by pre-existing ideas.

The global view obtained by microarray approaches might also lead to a more diffuse description of biological systems. This does not necessarily reflect an inability to produce quantitative data, but rather is a manifestation of a system's complexity. Likelihood values might describe a process better than exact numbers because complexity brings about variability, which is difficult to quantify because of its redundancy and ability to compensate.

Expansion towards experimentally complex systems will be an important direction of future development. Because the integration of information that is obtained in separate experiments at many molecular levels is crucial, this will eventually move beyond being carried out *in silico* only — experimental multiplexing by analysing different processes on a single system platform will become important. For example, in principle, all genes of an organism could be placed on a chip, including their promoter regions and other control units of transcription and translation. Many aspects of regulation, transcription, translation, protein modification and protein action, as well as their immediate effects on the other elements of the system (for example, protein activity on transcription), could be assayed *in vitro* in series or in parallel, in different combinations. Therefore, current *in silico* systems biology could translate into ever more complex experimental set-ups that allow the evaluation of a biological issue in a systemic experimentation. Given that surfaces have such a crucial role in biology, microarrays might be the archetypal platform for an eventual model of a cell. Systems-biology descriptions of both physical and biological surface effects, such as aspects of mass transport[95], will be crucial for this outcome.

Microarray technology has grown into a lively teenager, with all the features that are typical of this age. Although approaching adulthood, it is also still in a phase of trying to find its way and defining its goals. There is a lot of enthusiasm, which is typical for this age and which also reflects recent advances in performance. However, one should keep in mind that the technique's current behaviour is not always at its best, although the real blunders of the rough-and-tumble time are gone. At the same time, new and surprising sides become apparent and the complexity of the character continues to grow.

1. Poustka, A. *et al.* Molecular approaches to mammalian genetics. *Cold Spring Harb. Symp. Quant. Biol.* **51**, 131–139 (1986).
The first paper to describe the potential of array technology in genomics.
2. Cantor, C. R., Mirzabekov, A. & Southern, E. Report on the sequencing by hybridisation workshop. *Genomics* **13**, 1378–1383 (1992).
3. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470 (1995).
A publication that made the wider scientific community aware of the potential of array technology.
4. He, L. & Hannon, G. J. MicroRNAs: small RNAs with a big role in gene regulation. *Nature Rev. Genet.* **5**, 522–531 (2004).
5. Lu, J. *et al.* MicroRNA expression profiles classify human cancers. *Nature* **435**, 834–838 (2005).
6. Larkin, J. E., Frank, B. C., Gavras, H., Sultana, R. & Quackenbush, J. Independence and reproducibility across microarray platforms. *Nature Methods* **2**, 337–344 (2005).
7. Irizarry, R. A. *et al.* Multiple-laboratory comparison of microarray platforms. *Nature Methods* **2**, 345–350 (2005).
8. Petersen, D. *et al.* Three microarray platforms: an analysis of their concordance in profiling gene expression. *BMC Genomics* **6**, 63 (2005).
9. DeRisi, J. *et al.* Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.* **14**, 457–460 (1996).
10. Brazma, A. *et al.* Minimum information about a microarray experiment (MIAME) — toward standards for microarray data. *Nature Genet.* **29**, 365–371 (2001).
The first (successful) initiative to bring about at least some basic standardization to a microarray-based assay.
11. Cleveland, W. S. LOWESS: a program for smoothing scatterplots by robust locally weighted regression. *Am. Stat.* **35**, 54 (1981).
12. Huber, W., von Heydebreck, A., Sültmann, H., Poustka, A. & Vingron, M. Variance stabilisation applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18** (Suppl. 1), S96–S104 (2002).
13. Rocke, D. M. & Durbin, B. Approximate variance-stabilising transformations for gene-expression microarray data. *Bioinformatics* **19**, 966–972 (2003).
14. The International HapMap Consortium. A haplotype map of the human genome. *Nature* **437**, 1299–1320 (2005).
15. Syvanen, A. C. From gels to chips: 'minisequencing' primer extension for analysis of point mutations and single nucleotide polymorphisms. *Hum. Mutat.* **13**, 1–10 (1999).
16. Tõnisson, N. *et al.* Evaluating the arrayed primer extension resequencing assay of TP53 tumor suppressor gene. *Proc. Natl Acad. Sci. USA* **99**, 5503–5508 (2002).
17. Jobs, M., Howell, W. M., Strömqvist, L., Mayr, T. & Brookes, A. J. DASH-2: flexible, low-cost, and high-throughput SNP genotyping by dynamic allele-specific hybridization on membrane arrays. *Genome Res.* **13**, 916–924 (2003).
18. Brandt, O. *et al.* PNA-microarrays for hybridisation of unlabelled DNA-samples. *Nucleic Acids Res.,* **31**, e119 (2003).
19. Arntz, Y. *et al.* Label-free protein assay based on a nanomechanical cantilever array. *Nanotechnology* **14**, 86–90 (2003).
20. Hahm, J. & Lieber, C. M. Direct ultrasensitive electrical detection of DNA and DNA sequence variations using nanowire nanosensors. *Nano Lett.* **4**, 51–54 (2003).
21. Hild, M. *et al.* An integrated gene annotation and transcriptional profiling approach towards the full gene content of the *Drosophila* genome. *Genome Biol.* **5**, R3 (2004).
22. Johnson, J. M. *et al.* Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* **302**, 2141–2144 (2003).
23. Shoemaker, D. D. *et al.* Experimental annotation of the human genome using microarray technology. *Nature* **409**, 922–927 (2001).
24. Hollich, V. *et al.* Creation of a minimal tiling path of genomic clones for *Drosophila*: provision of a common resource. *Biotechniques* **37**, 282–284 (2004).
25. Milner, N., Mir, K. U. & Southern, E. M. Selecting effective antisense reagents on combinatorial oligonucleotide arrays. *Nature Biotechnol.* **15**, 537–541 (1997).
An elucidation of how the three-dimensional structure of nucleic acid influences its functionality.
26. Ren, B. *et al.* Genome-wide location and function of DNA binding proteins. *Science* **290**, 2306–2309 (2000).
27. Buck, M. J. & Lieb, J. D. ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments. *Genomics* **83**, 349–360 (2004).
28. Radonjic, M. *et al.* Genome-wide analyses reveal RNA polymerase II located upstream of genes poised for rapid response upon *S. cerevisiae* stationary phase exit. *Mol. Cell* **18**, 171–183 (2005).
29. Lichter, P. *et al.* High resolution mapping of human chromosome 11 by *in situ* hybridization with cosmid clones. *Science* **247**, 64–69 (1990).
30. Feuk, L., Carson, A. R. & Scherer, S. W. Structural variation in the human genome. *Nature Rev. Genet.* **7**, 85–97 (2006).
31. Solinas-Toldo, S. *et al.* Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer* **20**, 399–407 (1997).
32. Ishkanian, A. S. *et al.* A tiling resolution DNA microarray with complete coverage of the human genome. *Nature Genet.* **36**, 299–303 (2004).
33. Pollack, J. R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.* **23**, 41–46 (1999).
34. Bignell, G. R. *et al.* High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res.* **14,** 287–295 (2004).
35. Barrett, M. T. *et al.* Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc. Natl Acad. Sci. USA* **101**, 17765–17770 (2004).
36. Brennan, C. *et al.* High-resolution global profiling of genomic alterations with long oligonucleotide microarray. *Cancer Res.* **64**, 4744–4748 (2004).
37. Huang T. H., Perry M. R. & Laux D. E. Methylation profiling of CpG islands in human breast cancer cells. *Hum. Mol. Genet.* **8**, 459–470 (1999).
38. Yan, P. S. *et al.* Dissecting complex epigenetic alterations in breast cancer using CpG island microarrays. *Cancer Res.* **61**, 8375–8380 (2001).
39. Adorjan, P. *et al.* Tumour class prediction and discovery by microarray-based DNA methylation analysis. *Nucleic Acids Res.* **30**, e21 (2002).
40. Mund, C. *et al.* Array-based analysis of genomic DNA methylation patterns of the tumour suppressor gene p16 promoter in colon carcinoma cell lines. *Nucleic Acids Res.* **33**, e73 (2005).
41. Craig, A., Nizetic, D., Hoheisel, J. D., Zehetner, G. & Lehrach, H. Ordering of cosmid clones covering the herpes simplex virus type I (HSV-I) genome: a test case for fingerprinting by hybridisation. *Nucleic Acids Res.* **18**, 2653–2660 (1990).
42. Bains, W. & Smith, G. A novel method for nucleic acid sequence determination. *J. Theor. Biol.* **135**, 303–307 (1988).
The first publication on sequencing by hybridization.
43. Drmanac, R., Labat, I., Brukner, I. & Crkvenjakov, R. Sequencing of megabase plus DNA by hybridisation: theory of the method. *Genomics* **4**, 114–128 (1989).
44. Khrapko, K. *et al.* An oligonucleotide hybridization approach to DNA sequencing. *FEBS Lett.* **256**, 118–122 (1989).
45. Meier-Ewert, S., Maier, E., Ahmadi, A. R., Curtis, J. & Lehrach, H. An automated approach to generating expressed sequence catalogues. *Nature* **361**, 375–376 (1993).
46. Green Tringe, S & Rubin, E. Metagenomics: DNA sequencing of environmental samples. *Nature Rev. Genet.* **6**, 805–814 (2005).
47. Rondon, M. R. *et al.* Cloning the soil metagenome: a strategy for assessing the genetic and functional diversity of uncultured microorganisms. *Appl. Environ. Microbiol.* **66**, 2541–2547 (2000).
48. Venter, J. C. *et al.* Environmental genome shotgun sequencing of the Saragosso Sea. *Science* **304**, 66–74 (2004).
49. Sebat, J. L., Colwell, F. S. & Crawford, R. L. Metagenomic profiling: microarray analysis of an environmental library. *Appl. Environ. Microbiol.* **69**, 4927–4934 (2003).
50. Drmanac, R. *et al.* DNA sequencing by hybridisation with arrays of samples or probes. *Methods Mol. Biol.* **170**, 173–179 (2001).
51. Bains, W. Selection of oligonucleotide probes and experimental conditions for multiplex hybridization experiments. *GATA* **11**, 49–62 (1994).
52. Lipshutz, R. J. *et al.* Using oligonucleotide probe arrays to access genetic diversity. *Biotechniques* **19**, 442–447 (1995).
53. Maitra, A. *et al.* The human mitochip: a high-throughput sequencing microarray for mitochondrial mutation detection. *Genome Res.* **14**, 812–819 (2004).
54. Baum, M. *et al.* Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. *Nucleic Acids Res.* **31**, e151 (2003).
55. Singh-Gasson, S. *et al.* Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nature Biotechnol.* **17**, 974–978 (1999).
56. Canard, B., Cardona, B. & Sarfati, R. S. Catalytic editing properties of DNA polymerases. *Proc. Natl Acad. Sci. USA* **92**, 10859–10863 (1995).
57. Bennett, S. Array of hope for personalised medicine. *Curr. Drug Discov.* 15–19 (2004).
58. Brenner, S. *et al. In vitro* cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc. Natl Acad. Sci. USA* **97**, 1665–1670 (2000).
59. Shendure, J., Mitra, R. D., Varma, C. & Church, G. M. Advanced sequencing technologies: methods and goals. *Nature Rev. Genet.* **5**, 335–344 (2004).
A comprehensive coverage of high-throughput sequencing methods.
60. Velculescu, V. E., Zhang, L., Vogelstein, B. & Kinzler, K. Serial analysis of gene expression. *Science* **270**, 484–487 (1995).
61. Beaucage, S. L. & Caruthers, M. H. Deoxynucleoside phosphoramidites — a new class of key intermediates for deoxynucleotide synthesis. *Tetrahedron Lett.* **22**, 1859–1862 (1980).
62. Sindelar, L. E. & Jaklevic, J. M. High-throughput DNA synthesis in a multichannel format. *Nucleic Acids Res.* **23**, 982–987 (1995).
63. Lashkari, D. A., McCusker, J. H. & Davis, R. W. Whole genome analysis: experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR. *Proc. Natl Acad. Sci. USA*, **94**, 8945–8947 (1997).
64. Weiler, J. & Hoheisel, J. D. Combining the preparation of oligonucleotide arrays and synthesis of high quality primers. *Anal. Biochem.* **243**, 218–227 (1996).
65. Tian, J. *et al.* Accurate multiplex gene synthesis from programmable DNA microchips. *Nature* **432**, 1050–1054 (2004).
Describes the clever use of microarrays for gene synthesis and quality assessment.
66. Zhou, X *et al.* Microfluidic PicoArray synthesis of oligodeoxynucleotides and simultaneous assembling of multiple DNA sequences. *Nucleic Acids Res.* **32**, 5409–5417 (2004).
67. Richmond, K. E. *et al.* Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis. *Nucleic Acids Res.* **32**, 5011–5018 (2004).
68. Beier, M. & Hoheisel, J. D. Production by quantitative photolithographic synthesis of individually quality-checked DNA microarrays. *Nucleic Acids Res.* **28**, e11 (2000).
69. Beier, M., Stephan, A. & Hoheisel, J. D. Synthesis of photolabile 5'-O-phosphoramidites for the production of microarrays of inversely oriented oligonucleotides. *Hel. Chim. Acta* **84**, 2089–2095 (2001).
70. Yelin, R. *et al.* Widespread occurance of antisense transcription in the human genome. *Nature Biotechnol.* **21**, 379–386 (2003).
71. Paddison, P. J. *et al.* A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428**, 427–431 (2004).
72. Boutros, M. *et al.* Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* **303**, 382–385 (2004).
73. Arjomand, A. & Kumar, A. Integration of RNAi into drug discovery. *Curr. Drug Discov.* 1–4 (2003).
74. Sauer, S. *et al.*Miniaturisation in functional genomics and proteomics. *Nature Rev. Genet.* **6**, 465–476 (2005).
75. He, M. & Taussig, M. J. Single step generation of protein arrays from DNA by cell-free expression and *in situ* immobilisation (PISA method). *Nucleic Acids Res.* **29**, e73 (2001).
76. Ramachandran, N. *et al.* Self-assembling protein microarrays. *Science* **305**, 86–90 (2004).

77. Mukherjee, S. *et al.* Rapid analysis of the DNA-binding specificities of transcription factors with DNA microarrays. *Nature genet.* **36**, 1331–1339 (2004).
78. Bulyk, M. L., Gentalen, E., Lockhart, D. J. & Church, G. M. Quantifying DNA-protein interactions by double-stranded DNA-arrays. *Nature Biotechnol.* **17**, 573–577 (1999).
79. Gerry, N. P. *et al.* Universal DNA microarray method for multiplex detection of low abundance point mutations. *J. Mol. Biol.* **292**, 251–262 (1999).
80. Shoemaker, D. D., Lashkari, D. A., Morris, D., Mittmann, M. & Davis, R. W. Quantitative phenotypic analysis of yeast deletion mutants using a highly parallel molecular bar-coding strategy. *Nature Genet.* **14**, 450–456 (1996).
81. Fan, J. B. *et al.* Parallel genotyping of human SNPs using generic high-density oligonucleotide tag array. *Genome Res.* **10**, 853–860 (2000).
82. Hardenbol, P. *et al.* Multiplexed genotyping with sequence-tagged molecular inversion probes. *Nature Biotechnol.* **21**, 673–678 (2003).
83. Urata, H., Ogura, E., Shinohara, K., Ueda, Y. & Akagi, M. Synthesis and properties of mirror-image DNA. *Nucleic Acids Res.* **20**, 3325–3332 (1992). **This paper describes the interesting L-form derivative of nucleic acids that could be useful to microarray applications.**
84. Zhong, X. B. *et al.* Visualisation of oligonucleotide probes and point mutations in interphase nuclei and DNA fibers using rolling circle DNA amplification. *Proc. Natl Acad. Sci. USA* **98**, 3940–3945 (2001).
85. Dudley, A. M., Aach, J., Steffen, M. A. & Church, G. M. Measuring absolute expression with microarrays using a calibrated reference sample and an extended signal intensity range. *Proc. Natl Acad. Sci. USA* **99**, 7554–7559 (2002).
86. Brors, B., Hauser, N. C. & Vingron, M. Determination of absolute concentrations of biological molecules in complex samples; a real comparison of microarray experiments. European Patent PCT/EP03/03291 (2002).
87. Calladine, C. R. & Drew, H. R. *Understanding DNA; the Molecule and How it Works* (Academic Press, San Diego, 1997).
88. Rich, A. & Zhang, S. G. Z-DNA: the long road to biological function. *Nature Rev. Genet.* **4**, 566–572 (2003).
89. Pohl, F. M. Hysteretic behaviour of a z-DNA–antibody complex. *Biophys. Chem.* **26**, 385–390 (1987).
90. MacBeath, G. & Schreiber, S. L. Printing proteins as microarrays for high-throughput function determination. *Science* **289**, 1760–1763 (2000).
91. Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
92. Kononen, J. *et al.* Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nature Med.* **4**, 844–847 (1998).
93. Wheeler, D. B., Carpenter, A. E. & Sabatini, D. M. Cell microarrays and RNA interference chip away at gene function. *Nature Genet.* **37**, S25–S30 (2005).
94. Angenendt, P., Glöcker, J., Konthur, Z., Lehrach, H. & Cahill, D. J. 3D protein microarrays: performing multiplex immunoassays on a single chip. *Anal. Chem.* **75**, 4368–4372 (2003).
95. Kusnezow, W., Syagailo, Y. V., Goychuk, I., Hoheisel, J. D. & Wild, D. G. Antibody microarrays; the crucial impact of mass transport on assay kinetics and sensitivity. *Expert Rev. Mol. Diag.* **6**, 111–124 (2006).

Acknowledgements
The author thanks colleagues for sharing unpublished results and many interesting discussions. Relevant work in his laboratory was funded by grants from the Federal German Ministry of Education and Research (BMBF), the Deutsche Forschungsgemeinschaft and the European Commission.

Competing interests statement
The author declares no competing financial interests.

FURTHER INFORMATION
Bioconductor: http://www.bioconductor.org
Gene Ontology Consortium: http://www.geneontology.org
International HapMap Project: http://www.hapmap.org
The Microarray Gene Expression Data (MGED) Society: http://www.mged.org
MolTools — advanced molecular tools for array-based analyses of genomes: http://www.moltools.org
Access to this interactive links box is free online.