High density synthetic oligonucleotide arrays

Robert J. Lipshutz, Stephen P.A. Fodor, Thomas R. Gingeras & David J. Lockhart

Affymetrix, Inc. 3380 Central Expressway, Santa Clara, California 95051, USA. e-mail: rob_lipshutz@affymetrix.com

Experimental genomics involves taking advantage of sequence information to investigate and understand the workings of genes, cells and organisms. We have developed an approach in which sequence information is used directly to design high-density, two-dimensional arrays of synthetic oligonucleotides. The GeneChip® probe arrays are made using spatially patterned, light-directed combinatorial chemical synthesis, and contain up to hundreds of thousands of different oligonucleotides on a small glass surface. The arrays have been designed and used for quantitative and highly parallel measurements of gene expression, to discover polymorphic loci and to detect the presence of thousands of alternative alleles. Here, we describe the fabrication of the arrays, their design and some specific applications to high-throughput genetic and cellular analysis.

Biological systems read, store and modify genetic information using the rules of molecular recognition. Every nucleic acid strand carries the capacity to recognize complementary sequences through base pairing. The process of recognition, or hybridization, can be highly parallel; every sequence in a complex mixture can, in principle, be interrogated simultaneously. We have used these simple principles to develop powerful new experimental tools designed to collect and analyse vast amounts of genetic and cellular information. The introduction, development and integration of two key technologies[1–5] form the cornerstone of the new methods. The first is the fabrication of hundreds of thousands of polynucleotides at high spatial resolution in precise locations on a surface. The second, laser confocal fluorescence scanning, facilitates the measurement of molecular binding events on the array. These technologies and some variants have been adopted in both the commercial and academic sectors (see pages 25 (ref. 6), 10 (ref. 7) and 15 (ref. 8) of this issue).

At Affymetrix, we have focused on light-directed synthesis for the construction of high-density DNA probe arrays using two techniques: photolithography and solid-phase DNA synthesis. We attach synthetic linkers modified with photochemically removable protecting groups to a glass substrate and direct light through a photolithographic mask to specific areas on the surface to produce localized photodeprotection (Fig. 1). The first of a series of chemical building blocks, hydroxyl-protected deoxynucleosides, is incubated with the surface, and chemical coupling occurs at those sites that have been illuminated in the preceding step. Next, light is directed to different regions of the substrate by a new mask, and the chemical cycle is repeated[9,10]. Highly efficient strategies can be used to synthesize arbitrary polynucleotides at specified locations on the array in a minimum number of chemical steps[1]. For example, the complete set of 4^N polydeoxynucleotides of length N, or any subset, can be synthesized in only 4×N cycles. Thus, given a reference sequence, a DNA probe array can be designed that consists of a highly dense collection of complementary probes with virtually no constraints on design parameters. The amount of nucleic acid information encoded on the array in the form of different probes is limited only by the physical size of the array and the achievable lithographic resolution. Current large scale commercial manufacturing methods allow for approximately 300,000 polydeoxynucleotides to be synthesized on small 1.28×1.28 cm arrays—experimental versions now exceed one million probes per array.

Photolithography allows the construction of arrays with extremely high information content. Because the arrays are constructed on a rigid material (glass), they can be inverted and mounted in a temperature-controlled hybridization chamber. A fluorescently tagged nucleic acid sample injected into the chamber hybridizes to complementary oligonucleotides on the array. Laser excitation enters through the back of the glass support, focused at the interface of the array surface and the target solution. Fluorescence emission is collected by a lens and passes through a series of optical filters to a sensitive detector. By simply scanning the laser beam or translating the array, or a combination of both, a quantitative two-dimensional fluorescence image of hybridization intensity is quickly obtained[1,2].

Gene expression monitoring

Once sequence information (partial or complete) for a gene is obtained, the next question is generally: "what does its product do?". To understand gene function, it is helpful to know when and where it is expressed, and under what circumstances the expression level is affected. Beyond questions of individual gene

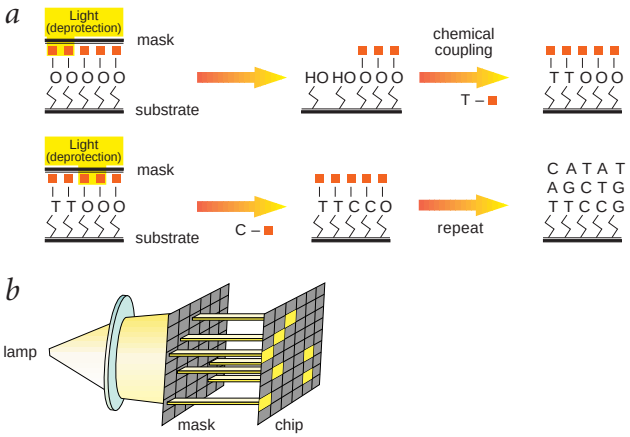

Fig. 1 ***a***, Light directed oligonucleotide synthesis. A solid support is derivatized with a covalent linker molecule terminated with a photolabile protecting group. Light is directed through a mask to deprotect and activate selected sites, and protected nucleotides couple to the activated sites. The process is repeated, activating different sets of sites and coupling different bases allowing arbitrary DNA probes to be constructed at each site. ***b***, Schematic representation of the lamp, mask and array.

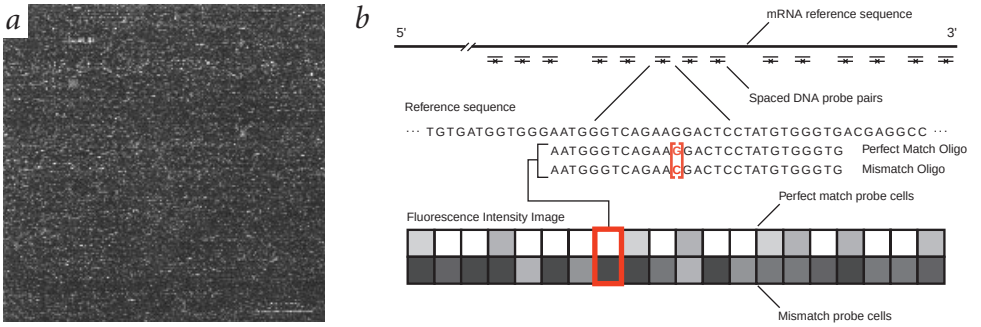

Fig. 2 Gene expression monitoring with oligonucleotide arrays. ***a***, A single 1.28×1.28 cm array containing probe sets for approximately 40,000 human genes and ESTs. This array contains features smaller than 22×22 μm and only four probe pairs per gene or EST. ***b***, Expression probe and array design. Oligonucleotide probes are chosen based on uniqueness criteria and composition design rules. For eukaryotic organisms, probes are chosen typically from the 3′ end of the gene or transcript (nearer to the poly(A) tail) to reduce problems that may arise from the use of partially degraded mRNA. The use of the PM minus MM differences averaged across a set of probes greatly reduces the contribution of background and cross-hybridization and increases the quantitative accuracy and reproducibility of the measurements.

function are also questions concerning functional pathways and how cellular components work together to regulate and carry out cellular processes. Addressing these questions requires the quantitative monitoring of the expression levels of very large numbers of genes repeatedly, routinely and reproducibly, while starting with a reasonable number of cells from a variety of sources and under the influences of genetic, biochemical and chemical perturbations. High-density oligonucleotide arrays have been shown to be well suited to this task[11–14], allowing the simultaneous monitoring of all yeast genes[15–17], all *Escherichia coli* genes, tens of thousands of human and mouse genes and selective subsets of genes from a wide range of organisms.

Oligonucleotide arrays for expression monitoring are designed and synthesized based on sequence information alone, without the need for physical intermediates such as clones, PCR products, cDNAs and so on. The key to their use is the targeted design of sets of probes to specifically monitor the expression levels of as many genes as possible. Using as little as 200 to 300 bases of gene, cDNA or EST sequence, independent 25-mer oligonucleotides are selected (non-overlapping if possible, or minimally overapping if necessary) to serve as sensitive, unique, sequence-specific detectors. Probe design is based upon complementarity to the selected gene or EST reference sequence, uniqueness relative to family members and other genes, and an absence of near-complementarity to other RNAs that may be highly abundant in the sample (for example, rRNAs, tRNAs, alu-like sequences, actin mRNA; Fig. 2). As a further filter, probes are chosen based on a set of empirically derived, composition-dependent design rules[11,15]. These rules are based on array hybridization data and help improve the odds of choosing oligonucleotides that will hybridize with high affinity and specificity (because of the probe redundancy, the rules do not have to be perfect). The arrays are

Table 1 • Gene expression oligonucleotide array performance characteristics

	Routine use	Current limit
starting material[a]	5 μg total RNA	0.5 μg total RNA
detection specificity[b]	1:100,000	$1:2\times10^6$
difference detection	twofold changes	10% changes
absolute quantitative accuracy[c]	±2×	±2×
false positives[d]	<2%	<0.1%
discrimination of related genes[e]	70–80% identity	93% identity
dynamic range (linear detection)[f]	~500-fold	$\sim10^4$-fold
number of probe pairs per gene or EST[g]	20	4
number of genes or ESTs per array	7,000	40,000

Performance characteristics for eukaryotic expression experiments using sets of 20 probe pairs per gene or EST, 24 μm synthesis features (more than 280,000 features per 1.28×1.28 cm array), overnight hybridizations of biotin-labelled, randomly fragmented cRNA, and standard washing, staining, detection and image analysis protocols. The typical time required for a high-resolution (3 μm pixels) fluorescence scan is less than 10 minutes. Labelled samples are typically hybridized to arrays between two and ten times without significant loss of performance (arrays are used for a single hybridization only).

[a]Total RNA is used directly without poly(A)$^+$ pre-purification steps. Messenger RNA is converted to cDNA using a dT-primed reverse transcription reaction. The cDNA is made double-stranded and then transcribed into cRNA in an in vitro transcription (IVT) reaction. The IVT reaction results in a linear, unbiased amplification (typically 30–100-fold) of the original mRNA population[11,15].

[b]Results obtained using recommended post-hybridization signal amplification protocols. Detection of spiked RNAs at a relative abundance of less than $1:10^6$ has been achieved for a variety of transcripts in the presence of both human and mouse complex RNA samples (H. Dong, D.J. Lockhart, unpublished).

[c]The hybridization signal intensities (PM minus MM values averaged over the probe pairs in the set) have been shown to be directly proportional to RNA concentration, and are predictive of absolute RNA concentration within a factor of two[11,15].

[d]False positives are defined based on experiments in which samples are split, hybridized to different arrays, and the results compared (performed with a wide range of human, mouse and yeast mRNA samples and arrays). A false positive is indicated if a probe set is scored qualitatively as an "Increase" or "Decrease" (based on an analysis of the overall patterns) and quantitatively as changing by at least twofold. The extremely low false positive rates of less than 0.1% are obtained using arrays synthesized on the same wafer and using simple repeated array scans and multiple-image data analysis methods.

[e]Probes are chosen from regions of sequence that have greatest divergence between family members, when known. Because of the targeted design of short oligonucleotides and the use of multiple probes per gene, it is possible to distinguish between very closely related sequences. For example, the yeast histone genes *HTA1* and *HTA2* are 93% identical at the DNA level (98% at the amino acid level). It was possible to design more than ten 25-mer oligonucleotides that were sufficiently different between the two sequences to allow unambiguous, independent detection of the two RNAs (L. Wodicka and D.J. Lockhart, unpublished). The histone genes *HTB1* and *HTB2* are also highly similar (87% identical at the DNA level) but were independently detected as well.

[f]The range of RNA abundance over which hybridization signal intensity is linearly related to concentration (linearity within approximately a factor of two at the extremes of concentration). The extended dynamic range is achieved by combining the results of repeated scans at different wavelengths and/or detection system gains.

[g]Most expression arrays currently use between 15 and 20 probe pairs per gene or EST to increase sensitivity and quantitative accuracy and reduce the rate of false calls. Probe sets containing as few as four probe pairs (chosen using standard *a priori* probe design methods and without direct empirical data) have been used to semi-quantitatively screen very large numbers of ESTs for expression changes.

Table 2 • Commercial GeneChip® probe arrays[a]

Application	Species	Information
expression	human[b]	~42,000 genes/ESTs
	mouse[b]	~30,000 genes/ESTs
	rat[c]	>11,000 genes/ESTs
	yeast (*S. cerevisiae*)	whole genome (all ORFs)
	Drosophila melanogaster[c]	>12,000 genes/ESTs
	Arabidopsis thaliana[c]	>12,000 genes/ESTs
	C. elegans[c]	whole genome (all ORFs)
	E. coli[c]	whole genome (all ORFs)
	other bacteria[d]	whole genome (all ORFs)
	targeted[e]	functionally selected gene sets
	custom[f]	any eukaryotic organism
genotyping	human	~2,000 SNPs
polymorphism screening	human	screening service
variant analysis	human	*CYP450* (2D6, 2C19)
	human	p53 (exons 2–11)
	HIV-1, Clade B	HIV (protease, rev. transcriptase)

[a]The GeneChip® System is required to run the arrays. The complete system is priced at $175,000 in the U.S. and includes the GeneChip® Fluidics Station, Hewlett Packard GeneArray™ Scanner, GeneChip® workstation and GeneChip® 3.1 analysis software. [b]Array update including additional gene and EST information expected in 1999. [c]Expected to be available mid-1999. [d]Five bacterial species planned for late 1999. [e]Specific subsets of genes selected based on their relevance to a given biological, clinical or disease area. [f]Gene or EST sets (public or proprietary) chosen by external users.

designed *in silico*, and as a result, it is not necessary to prepare, verify, quantitate and catalogue a large number of cDNAs, PCR products and clones, and there is no risk of a misidentified tube, clone, cDNA or spot.

Key to this approach is the use of probe redundancy. By this, we do not mean the deposition of the same piece of DNA in multiple locations on an array, but rather the use of multiple oligonucleotides of different sequence designed to hybridize to different regions of the same RNA. The use of multiple independent detectors for the same molecule greatly improves signal-to-noise ratios (due to averaging over the intensities of multiple array features), improves the accuracy of RNA quantitation (averaging and outlier rejection), increases the dynamic range, mitigates effects due to cross-hybridization, and drastically reduces the rate of false positives and miscalls (Fig. 2; Table 1; see also, page 43 of this issue (ref. 18)). An additional level of redundancy comes from the use of mismatch (MM) control probes that are identical to their perfect match (PM) partners except for a single base difference in a central position. The MM probes act as specificity controls that allow the direct subtraction of both background and cross-hybridization signals, and allow discrimination between 'real' signals and those due to non-specific or semi-specific hybridization (hybridization of the intended RNA molecules produces more signal for the PM probes than for the MM probes resulting in consistent patterns that are highly unlikely to occur by chance—the pattern recognition rules are codified in the analysis software). In the presence of even low concentrations of RNA, hybridization to the PM/MM pairs produces recognizable and quantitative fluorescent patterns. The strength of these patterns directly relates to the concentration of the RNA molecules in the complex sample (even without a competitive hybridization or two-colour comparison[11,15]). In short, redundant PM/MM probe sets allow one to determine with high confidence whether signal is generated by hybridization of the intended RNA molecule—a determination that cannot, in general, be made using a single spot with a given intensity.

It is important to stress that the oligonucleotide probes are designed using the same set of composition rules regardless of the organism or set of RNAs interrogated. This consistency of design methods (and consistency of probe length and chemical synthesis) makes it possible to use identical hybridization and analysis protocols for a wide range of different expression array types, obviating the need for extensive assay development with each new array design. The flow from sequence information to experimentation is thus rather direct, making oligonucleotide arrays both comprehensive and highly versatile tools for experimental genomics. While this approach requires knowledge of sequence in order to choose probes for a given gene or EST, and in that sense is not useful for *de novo* gene discovery, this is becoming less important as the body of sequence information continues to grow (Table 2).

A common misperception is that our approach is inflexible because changing a set of monitored genes requires a new design and a completely new set of photolithographic masks. While it is true that new masks are required, this has not proven to be a significant limitation. In fact, custom arrays (expression arrays as well as other types) are our fastest growing category of synthesized array by a large margin. In the last year, we have made more than 40 custom expression designs for biotechnology and pharmaceutical partners and anticipate that more than ten new custom designs per month will be made in the the coming year (Table 2; http://www.genechip.com). The next generation of expression arrays will be designed according to emerging sequence information from the human, mouse and rat. The technical aspects involved with the design, synthesis and use of oligonucleotide arrays for gene expression experiments are now fairly routine. New breakthroughs will require careful consideration of the types of experimental designs and analyses that best enable the extraction of biologically,

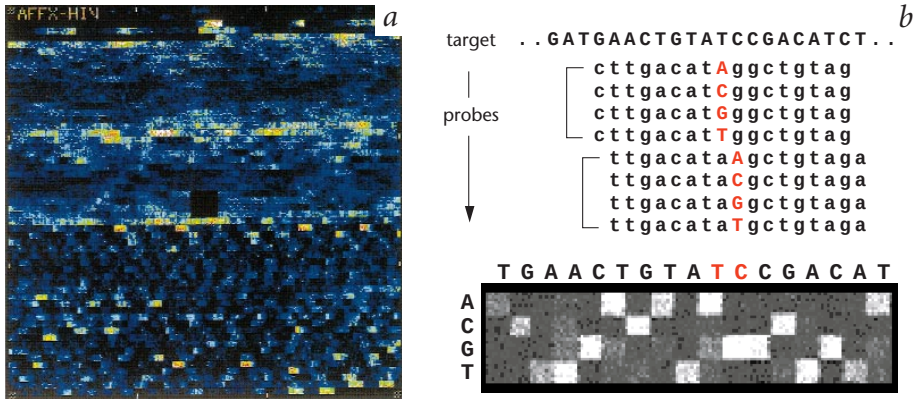

Fig. 3 Sequence analysis arrays. ***a***, Image of an HIV-1 genotyping array (HIV PRT) hybridized to labelled PCR-generated DNA copies of the protease (PR) and reverse transcriptase (RT) genes (0.8×0.8 cm array). Each base on both strands covering a total of 1515 base pairs is interrogated in the HIV-1 PR (codons 1–99) and RT (codons 1–400) genes. Additionally, to ensure detection in the event of multiple mutations occurring in close proximity, most common drug-resistance conferring mutations are encoded on the array as specialized tilings. ***b***, General tiling strategy. Detection of mutations or polymorphisms in a sequence is accomplished by using a four-probe interrogation strategy. In this illustration, four 17-mer oligonucleotide probes are used to determine the identity of the base in the middle of the probe sequence. The probe that forms the most stable duplex will provide the highest fluorescent signal among the four probes assigned to interrogate the central base. The next nucleotide in the target sequence is interrogated in the same manner, using another set of four oligonucleotide probes. Probes with interrogation positions other than the central position, or probes of different lengths can also be used to query the targeted base. Analysis of both strands of a target can be carried out on the same array to increase the confidence of the base determination.

Table 3 • Array capacity and feature size

Feature Size	Expression[a]	Sequence Analysis[b]	Genotyping[c]
50 μm	1600–6400 genes	8–16 kb	2,000–4,000 markers
20 μm	10,000–50,000 genes	50–100 kb	12,000–25,000 markers
2 μm	>1 million genes	500–1,000 kb	1.2–2.5×10^6 markers

All numbers calculated for 1.28×1.28 cm arrays. [a]Assuming 4–20 probe pairs per gene. [b]Assuming 4–8 probes per basepair. [c]Assuming 6–32 probes per marker.

clinically and pharmaceutically relevant information from massively parallel, genome-wide measurements of this type.

It should be noted that arrays can be used for purposes other than those for which they were designed. For example, Raymond Cho and colleagues have 'read' the results of yeast two-hybrid screens using a yeast expression array[19]; Elizabeth Winzeler and colleagues used the same type of arrays to carry out genome-wide, direct allelic variation scanning experiments in order to identify, map and score variation between two yeast strains[20].

Genotypic analysis

Variation in DNA sequence underlies most of the differences we observe within and between species. Locating, identifying and cataloguing these genotypic differences represent the first steps in relating genetic variation to phenotypic variation in both normal and diseased states (see pages 56 (ref. 21) and 42 (ref. 18)). The design of arrays for this purpose is straightforward. Given a reference sequence for a region of DNA, four probes are designed to interrogate a single position. One is designed to be perfectly complementary to a short stretch of the reference sequence; the other three are identical to the first, except at the interrogation position where one of the other three bases are substituted (Fig. 3; see also, Fig. 2 on page 43). Upon incubation with the reference sequence, the probe complementary to the reference sequence will obtain the highest fluorescence intensity. In the presence of a sample with a different base at the interrogation position (a substitution variant), the probe containing the complementary variant base will obtain the highest fluorescence intensity. One thousand sets of four probes are used, to interrogate one thousand bases of sequence, in parallel. This strategy works extremely well for the detection of closely related variants of the reference sequence(s), indicating both the position, and in the case of substitutions, the identity of the differences. It is not, however, well-suited to the detection of length polymorphisms or large deletions and insertions.

Mapping the approximate location—in the three billion nucleotide human genome—of a genetic variation that correlates with a phenotype requires the identification and use of genetic 'landmarks'. Single-nucleotide polymorphisms (SNPs) are the most frequent type of variation in the human genome; this, and the ease with which they can be identified, recommend them for this purpose. Illustrating this, is a study by David Wang and colleagues[22], who identified 3,241 candidate SNPs contained in STSs collected at the Whitehead Institute and Sanger Center, and mapped 2,227 of them. They screened a total of 2.3 million bases of sequence for variations among eight individuals using conventional gel-based sequencing and high-density oligonucleotide arrays to identify candidate SNPs and create a third generation genetic map for the human genome. More than two thousand of these SNPs have been selected, and sets of appropriate probes synthesized on a high density array. The 'SNP chips' are intended for commercial distribution (Table 2) and can be used for linkage, linkage disequilibrium and loss of heterozygosity studies. Arrays can also be used to scan the genome for new SNPs. Shrinking the area occupied by each of the interrogating oligonucleotide probes (or 'synthesis features') to approximately 15×20 μm, a total of 50,000 nucleotides (on both strands) can be screened for the presence of polymorphisms on a single array.

For organisms with smaller, sequenced genomes, identifying the genetic causes of heritable variation is more straightforward. Yeast expression arrays provide a means of determining genome-wide, allelic variation[20]. More than 3,700 biallelic markers have been identified through a straightforward set of whole genome hybridizations. These markers were then used to map five different loci with high resolution. Similarly, a single high-density oligonucleotide array has permitted the characterization of every position of the entire 16,500-bp human mitochondrial genome in several dozen individuals[23]. SNP markers are currently sought in eukaryotic and prokaryotic genomes for constructing genetic maps—as the first step in locating variations that correlate with important phenotypic conditions. Arrays of this type have also been commercialized (Table 2; Fig. 3*a*) are now being applied in a clinical setting[24,25].

A frequently overlooked but important advantage of hybridization-based approaches to variant detection is that the samples can be rather complex and do not have to contain only a single (or small number of) DNA species. For example, eleven different exons are analysed in parallel on the p53 chips, more than 300 short STSs are interrogated at once on polymorphism screening arrays, as many as 3000 markers are scored on a single SNP chip, and nearly 20% of the entire yeast genome can be screened for variations based on a hybridization of complete genomic DNA (ref. 20). With respect to sample preparation, amplification of markers, STSs, exons and whole genes does not have to produce pure, single species. It is of course necessary that the desired

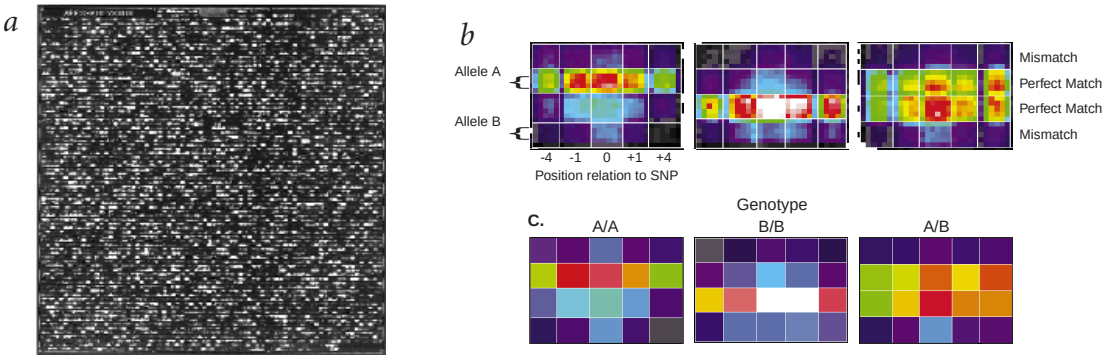

Fig. 4 Genotyping arrays. ***a***, A single array with over 120,000 probes designed to determine the genotype of a sample at over 3,000 biallelic loci. ***b***, The fluorescence intensity pattern for a set of probes designed to interrogate a single locus showing the presence of an AA homozygote, a BB heterozygote, and a BB homozygote. The upper and lower halves of the probe blocks interrogate the A and B alleles, respectively. Each half consists of pairs of probes centered on the polymorphic position and offset one and four bases to either side. The probe pairs consist of a perfect match and single base mismatch to the reference sequence for the specific allele. For each locus, interrogation blocks are included for both the sense and anti-sense strands.

sequence be amplified and labelled, but additional sequences can also be present in considerable amounts without significant loss of detection accuracy—except in cases in which highly similar pseudogenes or other related sequences are present that can cross-hybridize and complicate the analysis. Ultimately, it may be possible to hybridize complete human genomic DNA samples to arrays designed to interrogate any desired set of markers, genes or otherwise interesting regions.

Once an array is designed based on a set of reference sequences, the hybridization patterns can be used to classify samples into groups without determining the exact sequence. For example, oligonucleotide arrays designed to query *rpoB*, which encodes the β subunit of the RNA polymerase and RNA, the 16S ribosomal gene in *Mycobacterium tuberculosis*, accurately classified unknown samples from different species of the *Mycobacterium* genus[26,27].

Generic arrays

In addition to arrays designed with gene-specific or marker-specific probes, we have made arrays with large collections of probes that can be used for multiple purposes. An example of more 'generic' arrays are ones that contain all possible oligonucleotides of 8 or 9 nucleotides (either alone or synthesized at the ends of short common sequences). Complete 8-mer and 9-mer arrays have been used to survey various DNAs ranging in length from a few hundred bases up to as many as 16.5 kb for sequence differences relative to a reference[28]. The data are analysed using the idea of 'virtual tilings', which are conceptually similar to the physical tilings of overlapping probes on conventional sequence analysis arrays. The difference in this case is that the appropriate probe sets, as defined by a designated reference sequence, are extracted *analytically* from the set of all possible N-mers. Based on the observed hybridization (or combined hybridization/ligation) patterns, it is possible to detect sequence differences (including substitutions, short insertions and deletions) in almost any sequence up to about 2.5 kb in length with high accuracy and a low rate of false positives[28]. Similar arrays have also been used to fingerprint and physically map cosmids[29].

Another example of a generic array is one that contains approximately 16,500 different 20-mer 'DNA tags' (tag probes represented in both orientations and accompanied by a mismatch partner). The 20-mer tags are selected so that they have similar compositions but very different sequences relative to each other (for similar affinity but high specificity). Daniel Shoemaker *et al*. used an earlier version of this type of array that contained only 4,500 tag sequences to read out the results of competitive growth experiments with tagged or 'bar-coded' yeast deletion strains[30]. The diverse 20-mer probe arrays are well-suited to other types of highly parallel tagging and analysis experiments, as well as for reading out the results of, for example, allele-specific amplification and extension reactions for high-throughput scoring of SNPs (X.C. Huang and colleagues, pers. comm.).

Data management, visualization and analysis

Experiments with high density arrays generate unprecedented amounts of genetic and cellular information. In expression monitoring, for example, a single hybridization experiment can produce quantitative results for as many as 40,000 human genes at once. Software tools have been developed to manage these huge data sets across many experiments; to query, sort, cluster and visualize by time, behaviour, function, chromosomal position, pathway and other phenotypic or experimental parameters; and to link the results to other information databases (http://www.genechip.com). Other analytical methods and software tools are currently being developed by Affymetrix and other companies and academic groups; a database architecture for the management of expression information should encourage the development of a broad set of mutually compatible tools

High-density DNA probe arrays are powerful tools for a broad set of applications including gene expression monitoring, sequence analysis, and genotyping. As the feature size shrinks, the information capacity of the array increases (Table 3). With recent adaptation of semiconductor-like photoresist processes, it is possible to synthesize arrays with features as small as 2 μm (refs 31, 32; Beecher, pers. comm.). At this resolution, one hundred million non-overlapping 30-mer probes spanning the entire human genome would fit on a 2×2 cm array—indicating that in some cases at least, smaller can be better.

1. Fodor, S.P.A. *et al*. Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767–773 (1991).
2. Fodor, S.P.A. *et al*. Multiplexed biochemical assays with biological chips. *Nature* **364**, 555–556 (1993).
3. Pease, A.C. *et al*. Light-generated oligonucleotide arrays for rapid DNA sequence analysis. *Proc. Natl Acad. Sci. USA* **91**, 5022–5026 (1994).
4. Fodor, S.A. Massively parallel genomics. *Science* **277**, 393–395 (1997).
5. Southern, E., Maskos, U. & Elder, R. Hybridization with oligonucleotide arrays. *Genomics* **13**, 1008–1017 (1992).
6. Bowtell, D.L. Options available—from start to finish—for obtaining expression data by microarray. *Nature Genet.* **21**, 25–32 (1999).
7. Duggan, D.J., Bittner, M., Chen, Y., Meltzer, P. & Trent, J. Expression profiling using cDNA microarrays. *Nature Genet.* **21**, 10–14 (1999).
8. Cheung, V.G. *et al.* Making and reading microarrays. *Nature Genet.* **21**, 15–19 (1999).
9. McGall, G.H. *et al*. The efficiency of light-directed synthesis of DNA arrays on glass substrates. *J. Am. Chem. Soc.* **119**, 5081–5090 (1997).
10. Pirrung, M.C., Fallon, L. & McGall, G. Proofing of photolithographic DNA synthesis with 3′ 5′-dimethoxybenzoinyloxycarbonyl-protected deoxynucleoside phosphoramidites. *J. Organic Chem.* **63**, 241–246 (1998).
11. Lockhart, D.J. *et al*. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**, 1675–1680 (1996).
12. Mack, D.H. *et al*. Deciphering molecular circuitry using high-density DNA arrays. in *Biology of Tumors* (eds Hihich, E. & Croce, E.) 85–108 (Plenum Press, New York, 1998).
13. Zhu, H., Cong, J.-P., Mamtora, G., Gingeras, T. & Shenk, T. Cellular expression altered by human cytomegalovirus: global monitoring with oligonucleotide arrays. *Proc. Natl Acad. Sci. USA* (in press).
14. de Saizieu, A. *et al*. Bacterial transcript imaging by hybridization of total RNA to oligonucleotide arrays. *Nature Biotechnol.* **16**, 45–48 (1998).
15. Wodicka, L., Dong, H., Mittmann, M., M.-H. & Lockhart, D.J. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotechnol.* **15**, 1359–1367 (1997).
16. Cho, R.J. *et al.* A genome-wide transcriptional analysis of the mititoc cell cycle. *Mol. Cell* **2**, 65–73 (1998)
17. Gray, N.S. *et al.* Exploiting chemical libraries structure and genomics in the search for kinase inhibitors. *Science* **281**, 533–538 (1998).
18. Hacia, J. Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genet.* **21**, 42–47 (1999).
19. Cho, R.J. *et al*. Parallel analysis of genetic selections using whole genome oligonucleotide arrays. *Proc. Natl Acad. Sci. USA* **95**, 3752–3757 (1998).
20. Winzeler, E.A. *et al*. Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197 (1998).
21. Chakravarti, A. Population genetics—making sense out of sequence. *Nature Genet.* **21**, 56–60 (1999).
22. Wang, D.G. *et al*. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* **280**, 1077–1082 (1998).
23. Chee, M. *et al*. Accessing genetic information with high-density DNA arrays. *Science* **274**, 610–614 (1996).
24. Kozal, M. *et al*. Extensive polymorphisms observed in HIV-1 clade B protease gene using high density oligonucleotide arrays: implications for therapy. *Nature Med.* **7**, 753–759 (1996).
25. Gunthard, H.F., Wong, J.K., Ignacio, C.C., Havlir, D.V. & Richman, D.D. Comparative performance of high-density oligonucleotide sequencing and dideoxynucleotide sequencing of HIV type 1 pol from clinical samples. *AIDS Res. Hum. Retroviruses* **14**, 869–876 (1998).
26. Gingeras, T.R. *et al*. Simultaneous genotyping and species identification using hybridization pattern recognition analysis of generic *Mycobacterium* DNA arrays. *Genome Res.* **8**, 435–448 (1998).
27. Troesch, A. *et al*. *Mycobacterium* species identification and TB rifampin-resistance testing with high-density DNA probe arrays. *J. Clin. Microbiol.* (in press).
28. Gunderson, K.L. *et al*. Mutation detection by ligation to complete N-mer DNA arrays. *Genome Res.* (in press).
29. Sapolsky, R.J. & Lipshutz, R.J. Mapping genomic library clones using oligonucleotide arrays. *Genomics* **33**, 445–456 (1996).
30. 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).
31. Beecher, J.E., McGall, G.H. & Goldberg, M.J. Chemically amplified photolithography for the fabrication of high density oligonucleotide arrays. *Polymeric Materials Science & Engineering* **76**, 597–598 (1997).
32. McGall, G. *et al*. Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc. Natl Acad. Sci. USA* **93**, 13555–13560 (1996).