J. Biochem. **124**, 251-258 (1998)

Large-Scale Expression Measurement by Hybridization Methods: From High-Density Membranes to "DNA Chips"[1]

Bertrand R. Jordan[2]

TAGC Group, ICIM, Centre d'Immunologie INSERM/CNRS, Case 906, 13288 Marseille Cedex 9, France

Received for publication, February 12, 1998

The vast amount of sequence information becoming available on genes from man and from other species calls for corresponding increases in the rate of collection for data of a more functional nature. Expression measurements often constitute a first step in this direction, and can be performed on a reasonably large scale using highly parallel hybridization methods. Large sets of targets (clones, inserts, oligonucleotides) are hybridized with labeled complex probes prepared from total cell or organ mRNA; under the proper conditions, signals measure the relative abundance of each sequence species, and can be acquired quantitatively. These techniques are presently available in three formats: high-density membranes to be hybridized with radioactive complex probes, microarrays of DNA spots (a miniaturized version of the former technique) using fluorescent complex probes, and oligonucleotide chips that, although developed originally for mutation detection, can be adapted to perform expression measurements. The miniaturized formats clearly represent the future, since they allow higher sensitivity, assay of large numbers of entities and hopefully provide the opportunity to use small amounts of starting material.

Key words: arrays, expression, hybridization, genes, large-scale.

Knowledge on genes at the sequence level has grown explosively over the last few years. Complete genome sequences have been obtained for a number of prokaryotes, for the unicellular eukaryote *Saccharomyces cerevisiae* and, soon, for the metazoan *Caenorhabditis elegans*. Total sequencing of the human genome has begun in earnest and is planned to be essentially complete by 2005. In addition, a number of public and private projects have undertaken extensive tag sequencing, in which a single sequencing run yielding a few hundred nucleotides of sequence at the 5′ or 3′ end is performed for large numbers of randomly picked cDNA clones from various libraries. The resulting expressed sequence tags (ESTs) now number more than 1,000,000 for the human system in the (public) dbEST database (*1*). Of course these entities are highly redundant; various analyses indicate however that they probably represent more than half of the estimated human complement of 100,000 genes. Private databases, established by INCYTE and Human Genome Science and kept confidential, are claimed to contain each more than 2,000,000 human ESTs.

This fast increase in sequence information calls for corresponding improvements in the following steps of genetic analysis, loosely called "functional genomics." Unfortunately, however, most existing approaches (full-length cloning and sequencing, chromosomal mapping, protein expression, interaction studies in the yeast two-hybrid system...) focus on individual analysis of particular genes and do not lend themselves readily to high-throughput implementation. Expression studies are a fortunate exception: they can be performed in a highly parallel and effective format using hybridization of complex probes to large arrays of DNA fragments representing many genes (*2*).

Of course, there are numerous other methods to detect and quantify differential gene expression. Differential display (*3*) and its numerous variants, if set up in the correct fashion, provide an efficient way to identify genes that are differentially expressed between two tissues or two cell types, even when their transcripts are very rare. It does not, however, give quantitative expression information. Systematic cDNA sequencing can provide such data, especially if performed with libraries and methods designed for this purpose (*4*); however, quantification of low expression levels involves inordinate amounts of sequencing: statistically acceptable measurement of mRNAs at the 1:10,000 abundance level requires the sequencing of 50 to 100,000 clones for each library. The SAGE method (*5*) diminishes sequencing requirements by reducing each cDNA to a 9 nucleotide tag incorporated into a concatemer, but the work-up of the method is complex and statistical limitations still apply. On the other hand, hybridization methods, using complex probes and large arrays of targets, derive their power from the fact that each individual experiment provides a very large amount of information; they appear unrivaled for large-scale measurement of gene expression. This paper will review the present status of the field, that is evolving rapidly towards miniaturized formats and non-

[1] This work was supported in part by grants from the French Muscular Dystrophy Foundation (AFM) and from the French Genome Project (GREG) as well as by institutional grants from INSERM (Institut National de la Santé et de la Recherche Médicale) and CNRS (Centre National de la Recherche Scientifique).

[2] Phone: +33 4 91 26 94 96, Fax: +33 4 91 26 94 30, E-mail: jordan@ciml.univ-mrs.fr

radioactive detection, with large improvements of sensitivity and throughput.

Principle of the method, pitfalls, and performance requirements

Complex probes and arrayed targets. In essence, "hybridization signature" methods use a labeled "complex probe" prepared from the total mRNA of a given cell line, tissue or surgical sample by reverse transcription and labeling. The probe contains many different messenger RNA species, from a few thousand to as many as 30,000 for brain tissue, in widely different amounts. Typical figures suggest that a single mammalian cell contains approximately 300,000 individual mRNA molecules. A few of them are present at abundances of one or several percent, *i.e.* at several thousand copies per cell; the majority are rare, with levels ranging down to one molecule per cell or even less (an mRNA species present only once every ten cells may still have biological significance). These levels result from the combination of transcription rate, processing efficiency, and speed of degradation, and provide a reasonable estimate of the activity of each of the corresponding genes in the tissue of origin.

The strength of hybridization signature methods lies in the fact that thousands of expression levels can be measured in a single experiment. Briefly, a labeled complex probe is hybridized with an array consisting of many DNA targets, each representing a particular gene.[3] The targets can be either processed bacterial colonies, PCR products from cDNA clones, or sets of synthetic oligonucleotides designed to assay a particular gene, and we will discuss the corresponding implementations. The important point is that the combination of a complex probe containing many different mRNA species with a large array of targets allows highly parallel collection of information.

Kinetic considerations. It is important at this point to realize that hybridization conditions in expression measurements are quite different from those commonly used for Northern or Southern blotting. In those applications, there is a huge molar excess of the probe. Accordingly, the reaction goes to completion, *i.e.* the target is fully hybridized at the end of the incubation. For expression measurements, instead, each individual sequence species in the complex probe is present in small amounts relative to its target(s) on the array. Typical figures determined by us (*6*) for high-density membranes are 30 ng of a specific, 1 kb long target represented by a single colony on the membrane, and less than a tenth of a nanogram for the corresponding mRNA (more exactly, cDNA copy of the mRNA) assuming a relative abundance of one in a thousand with respect to total mRNA. Under these conditions, the kinetics of hybridization are linear (see Ref. *6* for more detail), and the amount of probe hybridized to a given target will be proportional to the abundance of the corresponding sequence in the complex probe—therefore to the expression level of the relevant gene in the cell or tissue from which the probe was derived.

As a consequence, the signals to be measured are fairly small, as targets are only hybridized to an extent of 0.1 or 1%, typically; also, since the measurement is actually a time point on a kinetic curve, ensuring equal speed of reaction over the whole array is essential, otherwise artefactual differences in signal will occur.

Artefacts and pitfalls. Although the principle of using arrayed clones and complex probes for high-throughput expression measurement was discussed as early as 1991 (*2*), most quantitative implementations of this approach were only published four years later (*6-8*). This stems from the need to resolve a number of experimental difficulties, from reliable manufacture of arrays to the definition of probe labeling and hybridization conditions ensuring reproducible and valid results. One of the major artefacts encountered, discussed in detail by us (*6, 9*) is the so-called "polyA effect," whereby labeled probe molecules containing a dT stretch after reverse transcription of mRNA can bind to any target molecule on the array that contains a polyA sequence. This can be eliminated by various means, including specific probe preparation and hybridization procedures, but must be monitored by inclusion in the array of negative control targets containing solely polyA sequences. Other difficulties include the contribution of repeat sequences to the signal and the existence of closely related gene families that may not be resolved at the stringencies used. Data acquisition, whether from radioactive or fluorescent probes, involves other issues such as spot detection algorithms, background subtraction procedures, assignment of the measured hybridization signatures to the correct entities (*10*) and, generally, efficient procedures for handling the thousands of data points generated in each individual experiment.

Performance requirements. Performance requirements for an ideal system are primarily the ability to quantify expression levels down to the level of one molecule per cell (1/300,000); linear response over the whole abundance range (from 1/300,000 to 1/100); a capability of simultaneous measurement for many targets, ideally the whole human complement of 100,000 genes; and the use of small amounts of starting material for the probe so that measurements can be done on hard-to-get cell populations or on clinical biopsies. This must be achieved with good reproducibility, such that a difference of a factor of 2 between two hybridization signatures should be highly significant. Finally, of course, the apparatus required should be compact, reliable, affordable and easy to use! Needless to say, none of the existing methods fulfills all of these criteria...

A short history of the field. The systematic use of arrayed clone libraries and high-density membranes was introduced by Hans Lehrach's group, first at EMBL, then at ICRF and now at the Max Planck Institute in Berlin. However, these were mostly used for applications such as access to libraries, where qualitative (and often manual) scoring of clones as positive or negative suffices. Quantitative use of this format for expression measurement was pioneered by this group (*11*); full implementation of the method was published later by several others (*6-8*). In these applications bacterial cDNA clones (or their PCR products) are arrayed onto Nylon filters; ^{32}P or ^{33}P-labeled complex probes are hybridized with them, and the results are acquired quantitatively using imaging plate systems.

The use of fluorescent labeled probes (that allow mini-

[3] I am keeping here the conventional terminology, where the immobilized DNA is considered as the target. Others propose to call the (known) immobilized DNA segments "probes" and the (unknown) labeled mRNA "sample."

aturization because of their much better intrinsic resolution) had been discussed for some time, and the first experimental implementation appeared in 1995 (*12*); this has been further developed to the point where "microarrays" containing several thousand DNA samples in surfaces of the order of 1 square centimeter have been produced and used by a few academic laboratories (*13-16*) and by several firms. Meanwhile a different type of device, the oligonucleotide chip, had been developed by three groups (*17-19*). In current industrial versions (marketed by Affymetrix, Palo Alto, USA), 64,000 different oligonucleotides are synthesized on a surface of less than two square centimeters (*20*). The resulting "DNA chip" is then used for "quasi-sequencing" applications, *e.g.* the detection of mutations in a predetermined gene such as p53. While primarily developed for such applications, the oligonucleotide chip also has capability for expression measurement, as first shown in late 1996 (*21*). Because of its tremendous potential for miniaturization, the oligo chip may in the long term represent the best avenue for systematic expression measurement.

High-density membranes and radioactive probes

I will discuss this version of the technology (Fig. 1) in some detail as it is the most developed and has been rather extensively used; also, many of the technical issues apply as well to other versions of hybridization signature measurement.

Making HD membranes. High-density membranes ("HD membranes") can be prepared by spotting bacterial colonies onto Nylon filters; the colonies are then grown and the membranes treated by standard procedures to lyse the cells, denature and bind the DNA to the support. Alternatively, DNA obtained by vector-PCR amplification of the cDNA inserts can be directly spotted onto the membranes. This version avoids the somewhat critical step of bacterial growth on the membranes, and gives better signal to background ratios; on the other hand the logistics of performing thousands of PCR reactions is cumbersome, and large differences in the amounts amplified between different clones are common. PCR products can be spotted "as they come," with resulting loss of sensitivity for those that are present in small amounts; alternatively, all concentrations can be measured and then adjusted to the same value, a procedure that is obviously better but also much more demanding.

Probe labeling and hybridization. Complex probes are prepared from mRNA or total RNA by reverse transcription and labeling. Various procedures have been used: cDNA synthesis primed by oligo-dT followed by labeling of the isolated cDNA by random priming, simultaneous synthesis and labeling by random priming on purified polyA$^+$ RNA (*6, 8*), simultaneous oligo-dT primed synthesis and labeling on total RNA (*9*). Elimination of the "polyA effect" is achieved in our case (*9*) through three steps: saturation of the polyA tail of mRNA with a large excess of dT25 (so that reverse transcription starts close to the beginning of the specific sequence), pre-annealing of the labeled probe with an excess of dA80 (to trap any probe molecules containing long dT stretches), and high-stringency washes after hybridization [to remove probe molecules attached through (short) polyT/polyA duplexes]. Labeling is done using ^{32}P or, preferably, ^{33}P that provides better resolution, allowing the use of higher spot densities.

Hybridization conditions are fairly standard; they must give low background (since signals are quite small) and ensure equivalent hybridization rate over the whole membrane. In our hands this eliminates formamide-based buffers (that do not consistently give low enough background) and requires comparatively large hybridization volumes (5 to 50 ml). Long hybridization times (48 to 72 h) are used to maximize the signals, that increase linearly

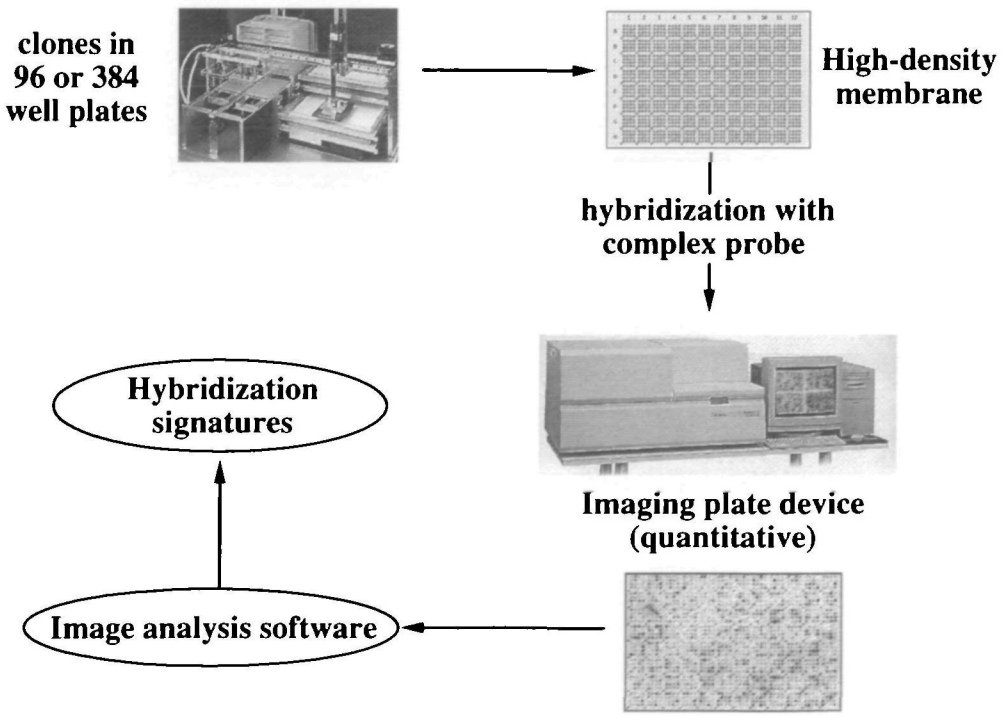

Fig. 1. **Outline of HD membrane system.** Bacterial clones containing cDNA plasmids (or PCR products prepared from them) are spotted from microtiter plates onto Nylon membranes. In the example illustrated, clones from seventeen 96-well plates (sixteen experimental plates and one set of controls) are spotted onto a single 8×12 cm membrane. Hybridization with a ^{33}P-labeled complex probe is followed by imaging plate exposure, scanning and image analysis to yield quantitative expression data.

with probe concentration and hybridization duration.

Acquisition of quantitative data. Acquisition of the hybridization signals must be done in a quantitative fashion. Imaging plates and their associated laser scanners provide this in a convenient format, since many membranes can be simultaneously exposed using several imaging plates, while the scanning occupies the instrument for just 5 or 10 min. These systems give linear, quantitative response over four or five orders of magnitude, with exposure times for this application ranging from a couple of days to a week. Real-time detection devices can also be used; they have the advantage of direct, more exact measurement and even wider dynamic range, but the instrument is tied up throughout the exposure that typically lasts at least several hours, thus reducing throughput.

Once acquired as a computer file, the image of the hybridization must be quantified, *i.e.* spots have to be detected (preferably with individual contour definition), quantified, background subtraction performed and, last but not least, each value must be attached to a correctly identified clone name for further analysis. The standard software provided with imaging plate machines (or other detection systems) does not handle this task effectively for HD membranes containing thousands or tens of thousands of spots. Specific software has been developed for this type of application and, at least in one instance, is commercially available (*10*).

Performance of HD systems. Present performance of HD membranes for hybridization signature measurement can be summarized as follows. Spot densities are of the order of 20 to 40 per cm^2, making it practical to assay thousands of clones on small (microplate-sized) membranes, and tens of thousands on large 20×20 cm^2 filters. Sensitivity is in the 1/10,000 range, *i.e.* expression levels can be quantified that correspond to mRNA species present at that level respective to total mRNA; dynamic range is around 1,000-fold, and reproducibility in properly controlled experiments is such that differences in expression level of a factor of two are highly significant. Probes are typically prepared from one or a few micrograms of polyA$^+$ RNA or, in our case, 25 μg of total RNA (as little as 4 μg can be used for an individual hybridization). These probe amounts limit the uses of the technology to situation where a relatively abundant source of cells or tissue is available, and prevent for example most uses of clinical material, particularly biopsies.

A quick tour of uses of HD membranes. The HD membrane approach has been used in two modes that can in fact overlap. For screening purposes, many (largely unknown) cDNA clones are hybridized with a series of probes designed to assay for expression patterns indicating genes likely to be relevant to the biological question studied (Fig. 2). The subset of clones meeting these criteria is then studied in more detail, *e.g.* tag-sequenced, assayed by *in situ* hybridization on tissues, mapped on the genome... This "quantitative differential screening" approach has been

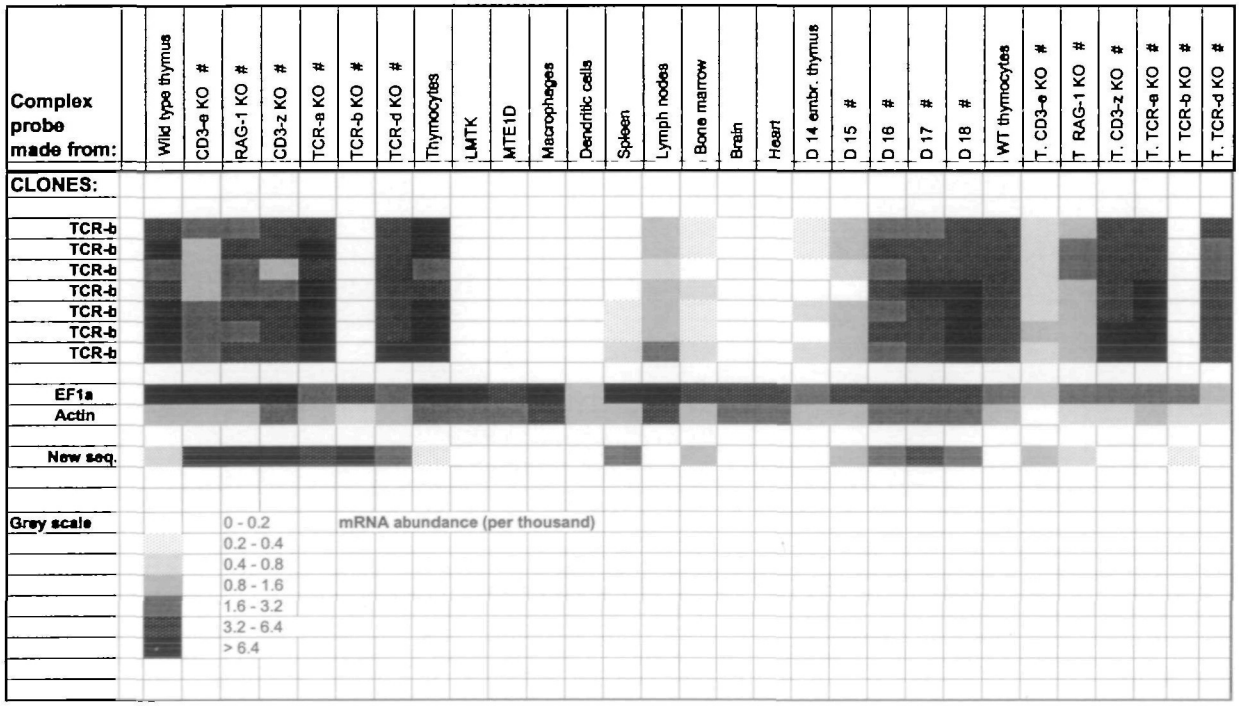

Fig. 2. **Example of data from multiple probings.** Arrays (whether in the form of HD membranes or in more miniaturized format) lend themselves readily to accumulation of data, as long as they include standardization procedures to allow quantitative comparison between successive experiments. In this example data for a few cDNA clones from mouse thymus after hybridization with a total of 29 different complex probes is shown using a grey scale system for intensity representation. Top, a set of different clones all containing inserts corresponding to the beta chain of the T cell receptor; middle, two "housekeeping" genes (that are expressed in all tissues tested, but do show variations in expression level); bottom, one of the clones corresponding to an "interesting" gene in the context of this project by virtue of its sequence ("new" on comparison with databases) and expression pattern: specific for lymphoid tissues, strongly modulated in several mouse knock-out mutants affecting T cell maturation, and strongly modulated during embryonic development. (Data from Dr. Alice Carrier in the author's laboratory).

applied to search for "new" genes differentially expressed in cancer cells and tissues (*7, 11*), to study genes specifically expressed in muscle (*8*), and for studies on events occurring during the maturation of the mouse thymus (*22*), to give but a few examples.

In a slightly different implementation, set of known genes expected to be relevant to a given process are assembled on a membrane: this is made easy by the freely available set of IMAGE cDNA clones (*23*).[4] The membrane is then used to assay expression levels for this chosen set of genes in a series of experimental or clinical situations. This approach, called by us "multiplex messenger assay" (MMA) is equivalent to a large series of Northern blots (*9*). Commercial suppliers (*24, 25*) now provide Nylon membranes on which a set of 588 human genes represented by PCR products from cDNA clones (*24*), or nearly 20,000 colonies corresponding to unique human genes (*25*) have been arrayed; these can then be hybridized by the investigator and analyzed to provide quantitative expression information.

The two approaches will in fact coalesce when it becomes practical to assemble a complete set of human genes on a single membrane, providing a universal tool for both screening and assay applications. Because of physical limitations, this is unlikely to happen with HD membranes and radioactive detection but is in sight with more compact versions of the technology using microarrays or oligonucleotide DNA chips and detection by fluorescence.

[4] Unfortunately this very useful resource is not completely error-free: discrepancies in the correspondence between physical clones and tag sequences range from 5 to 10%, imposing expensive and time-consuming verification when these entities are used as reagents.

Microarrays

Microarrays *versus* "DNA chips". I will use the term "microarray" to refer to sets of DNA targets, usually purified PCR products from cDNA or genomic clones, with sizes of the order of 0.5 to 1 kb, deposited on a solid support (generally a glass microscope slide) with a spot density of several hundred individual spots per cm^2, and keep the term "DNA chip" (more precisely, oligonucleotide chip) for large sets of oligonucleotides synthesized *in situ*, usually at much higher densities. For expression profiling, the microarrays are hybridized with complex probes in which the cDNA copy of the mRNA mixture has been labeled by inclusion of a nucleotide derivative containing a fluorochrome, and the hybridization intensity is acquired quantitatively by laser scanning and CCD camera detection (Fig. 3).

Microarrays thus correspond to the transposition on a much smaller scale of the HD membrane approach. Probe detection by fluorescence provides much better resolution (of the order of 10 μm instead of 100 μm with ^{33}P) and

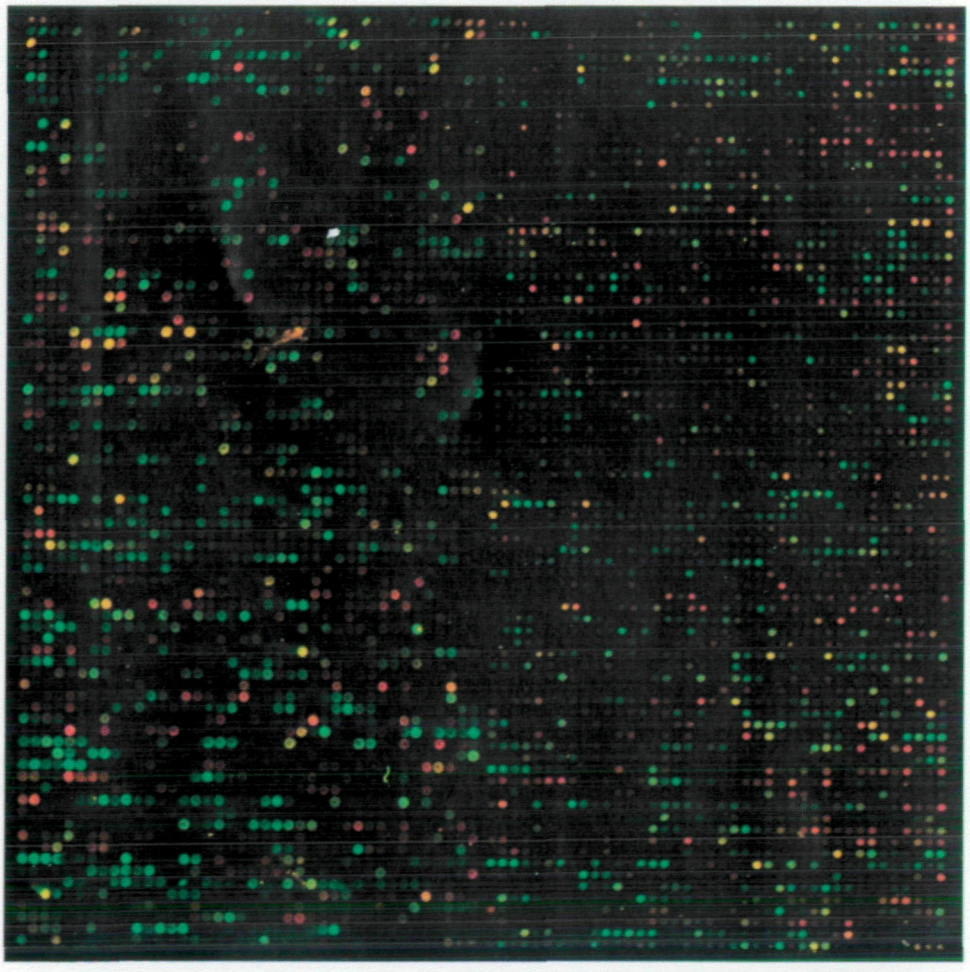

Fig. 3. **A microarray representing the whole *Saccharomyces cerevisiae* genome (*16*).** This microarray, whose physical dimensions are 1.8×1.8 cm, contains 6,400 spots representing more than 6,000 genes plus a number of controls—*i.e.* the almost complete set of genes revealed by sequencing of the yeast genome. The DNAs are PCR products spotted by a home-made split-pin robot (*26*) onto a glass slide. In the example shown, the microarray has been simultaneously hybridized with two fluorescent probes: one prepared from 1.25 μg of poly(A)$^+$ RNA from yeast cells shortly after inoculation in rich medium, labeled with Cy3-dUTP (green), the other from yeast cells after 19 h of growth in this medium, by which time the glucose has been exhausted and the shift to aerobic growth has occurred, labeled with Cy5-dUTP (red). Expression ratios can be directly estimated from the color of the spots (yellow = similar expression levels in the two conditions) and quantified by scanning in a laser confocal system. (Reprinted with permission from Ref. *16*. Copyright 1997. American Association for the Advancement of Science).

allows higher spot densities. A typical array has a surface of one or a few square centimeters and hybridization volumes can be as low as 2 μl, although 10 to 50 μl are more common. The high probe concentrations achievable in this way lead to good sensitivity. In the first published implementation of this principle (*12*), a set of 48 DNA targets (PCR products from *Arabidopsis thaliana* cDNA clones) was arrayed in a 3.5 mm by 5.5 mm area on glass microscope slides and hybridized in a total volume of 2 μl with fluorescein and/or lissamine-labeled cDNA prepared from 2 μg of total mRNA. Detection of (spiked) mRNA demonstrated a sensitivity of 1/50,000. Later implementations involving the same authors demonstrated progressive scaling-up of the technique, with an array of more than 1,000 elements to study gene expression in human cancer (*13*) and, recently, 6,400 elements representing the complete set of *Saccharomyces cerevisiae* genes (Fig. 3) used to explore the metabolic control of expression in this organism (*16*). This approach has also been implemented by firms, either for specific in-house projects or, as in the case of one of them (*26*), to offer large-scale expression measurement on a contract basis to (mostly) pharmaceutical industries.

Compared to the more traditional high-density membrane method, microarrays offer a number of advantages. Sensitivity is higher; the compact format makes it possible to handle larger numbers of targets (although quality control issues and downstream processing become correspondingly harder); dual labeling has been established and makes possible simultaneous hybridization of two different complex probes and direct comparison of expression levels without having recourse to correction and normalization procedures (*16*).

One potential advantage that has not yet been demonstrated in practice is a reduction in the amount of probe and therefore starting material. Extrapolation from HD membrane data indicate that it should be possible, in microarray format, to achieve good sensitivity with amounts of probe corresponding to a fraction of a microgram of total RNA, yet all published reports use micrograms of polyA$^+$ RNA. This is an important issue since such a development would render accessible expression profiling from small populations of sorted cells or from clinical biopsies; it does not seem to have received the necessary attention so far.

The main drawback of the microarray approach is its present inaccessibility to the typical academic or even industrial laboratory. At this time there is no commercial, off the shelf equipment (*e.g.* robotic spotting system, associated scanning device, and software suite) available to perform these measurements; this situation will change before long since manufacturers are developing such instruments. In addition, details on the system developed at Stanford (*13*), including instrument design and parts lists, are readily available (*27*). Microarrays with similar densities have also been developed on Nylon membranes with colorimetric detection of hybridized probe, a method that also displays surprisingly good sensitivity and allows detection and quantification with very affordable instruments such as commercial flatbed scanners or digital cameras (*30*). In summary, microarrays are definitely the next logical step in high-throughput expression measurement—although they may end up being superseded by oligonucleotide chips that were originally developed for a different purpose.

Oligonucleotide chips

A powerful mutation detection tool. Oligonucleotide chips ("oligo chips," "DNA chips") were originally proposed in the context of sequencing applications by Ed Southern in 1989,[5] and have since been developed by several groups (*17-19*). They consist of a large number of oligonucleotides of predetermined sequence, synthesized *in situ* on a glass support. Fodor *et al.* (*18*), and the firm Affymetrix, developed photochemical methods to perform this synthesis and succeeded in miniaturizing these arrays to the point where 64,000 different oligonucleotides (from 10 to 25 nucleotides long) can be routinely synthesized on an 1.28 by 1.28 cm chip (*20*) (Fig. 4). Further miniaturization is in progress, arrays of more than 400,000 elements have been reported (*28*), and an improvement rate similar to that achieved by the microprocessor industry should allow chips containing several million oligonucleotides in the near future.

These oligo chips have mostly been used so far for "quasi-sequencing" applications, in which the set of oligonucleotides is designed to be complementary to a known sequence such as that of the p53 gene. Hybridization of a fluorescent labeled probe prepared from a normal p53 gene will then yield a characteristic intensity pattern; if one or several mutations are present, the pattern will change in a characteristic way that allows to specify the nature of the mutations and the positions in which they occur. Oligo chips have great potential for such applications; even though they cannot be reused and individual chip cost is in the thousand-dollar (US) range, they do provide a cost-effective alternative to complete sequencing of the gene to be assayed.

Adaptation to expression measurement. This tool can also be used to perform quantitative expression measurements, as first shown in 1996 (*21*). This is made possible by the fact that each of the 64,000 "features" of a typical oligo chip (each measuring 50 by 50 μm), containing a single type of oligonucleotide sequence, carries one to ten million actual copies of this molecule, allowing widely different levels of hybridization. In addition, for expression measurement, a series of different oligonucleotides, chosen to avoid regions that hybridize poorly or indiscriminately, is used for each mRNA to be assayed. In fact, most of the data reported in Ref. *21* was obtained with a 64,000 oligonucleotide array in which each gene was represented by a set of 300 different oligonucleotides, plus the same number containing a central mismatch. The signal was taken as the difference between matched and mismatched oligonucleotide, summed over the whole set. Good sensitivity (1/300,000) and dynamic range were demonstrated—but at the expense of using a whole 64,000 element oligo chip to assay just 118 genes. In such a format, the method (that involves significant tooling costs as well as semi-empirical design, using rules derived from probe performance data, of the oligonucleotide sets) does not appear practical compared to microarrays where relatively long target DNAs readily yield specific hybridization signals using a single spot for each gene to be assayed.

[5] Ed Southern's oral presentation at the 1989 Genome Mapping and Sequencing Meeting at Cold Spring Harbor was entitled "Synthesis of oligonucleotides tethered to a glass surface—applications to the analysis of nucleic acid sequences." His first formal paper on the topic appeared in 1992, after those of Fodor *et al.* (*18*) and Khrapko *et al.* (*17*).

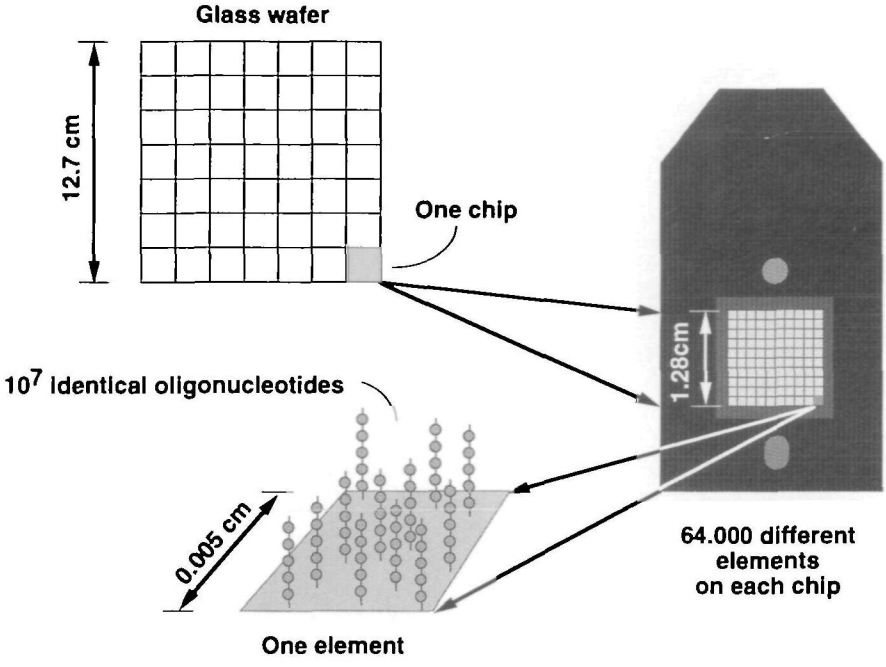

Fig. 4. **Manufacture and anatomy of an oligo chip (adapted from Ref. *18*).** Oligonucleotide chips are manufactured (top left) in batches of 49 (7×7) to 400 (20×20) by *in situ* synthesis on a glass wafer using photochemical methods that allow individual addressing of each element. The chips, each containing 64,000 to 400,000 elements, are then separated and mounted in a handling frame (right). Each element ("feature") contains millions of identical oligonucleotides; their present size 20×20 to 50×50 μm may be reduced in the future to less than 10×10 μm.

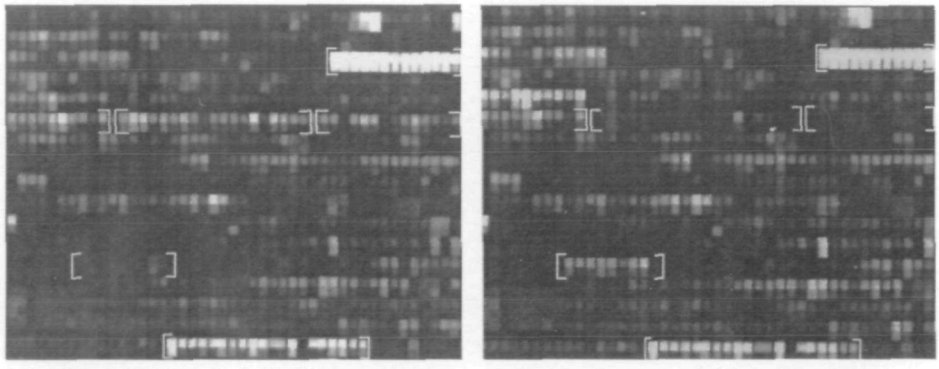

Fig. 5. **Use of oligo chips for expression measurement (adapted from Ref. *29*).** This figure shows an enlargement of a small region from a set of four 64,000 element chips designed to assay expression of the whole set of yeast genes. Expression measurement for each gene result from hybridization signals measured on a set of forty oligonucleotides (25-mers): 20 (top rows) that are chosen according to the sequence of the gene, and 20 (bottom rows) whose sequence is identical to the preceding set except for a central mismatch. The signal is taken as the background-subtracted sum of differences between matched and mismatched oligonucleotides. Automatic choice of the "best" set of oligos is clearly one of the critical features in the design of such "expression chips." The two panels shown correspond respectively to a probe from yeast grown in rich medium (left) and in minimal medium (right); chip areas corresponding to three "housekeeping" and three differentially expressed genes are indicated by white brackets. [Data provided by Dr. Lockhart, Affymetrix (*29*)].

It is however possible to reduce the number of oligonucleotides in the set to much smaller values, and still obtain quantitative results, as suggested in the paper referred to above. A recent publication from the same group demonstrates this with sets of 20 oligonucleotides (in fact, 20 pairs) per gene (*29*). This is applied to assay the whole complement of *S. cerevisiae* genes (Fig. 5), thus providing a very interesting comparison with the microarray constructed by DeRisi *et al.* (*16*) for the same purpose. Four oligo chips (256,000 oligonucleotides) are needed to perform this assay instead of a single microarray; performance appears to be very similar and in some respects better, although the study is less detailed in terms of yeast biology. The specificity of oligonucleotide chips in this application is enhanced by the avoidance of problematic sequences and the fact that non-specific hybridization is actually measured on the set of mismatched oligonucleotides.

Will oligo chips win over microarrays? Although this method appears *a priori* more involved than the straightforward microarray approach, it seems able to provide similar compacity and performance. It still has great potential for further miniaturization, as feature sizes of ten by ten micrometers square are possible (more than one million oligonucleotides per cm^2), while the potential of robotic spotting for drastic size reduction is more limited. Thus oligo chips may in the long term become the best solution to high-throughput expression measurement. Microarrays do at present provide the useful possibility of obtaining data for genes of unknown sequence, but as more and more sequencing is performed this advantage will become less significant.

Conclusions

Large-scale expression measurement is the logical next step after sequence determination. This is readily apparent to the yeast community now that the whole genome sequence has been determined, and it is no wonder that the two competing micro methods have each produced yeast

whole-genome chips (*16, 29*). Pharmaceutical companies, that have access to confidential EST collections containing more than two million entries and probably representing the majority of human genes, also invest heavily in this approach, either through in-house developments or by contracting the work out to specialized firms. Their aim is to detect genes that are differentially expressed between normal and diseased tissue and may therefore represent new targets for drug development. Basic research laboratories have somewhat lagged behind in using this approach, partly because of the equipment cost involved, but will have to invest in it as one of the necessary research tools of the near future.

Some of the technical trends are clear. Expression profiling using completely uncharacterized clones will die out as more and more tag-sequenced libraries become available, with the dual advantage of eliminating redundancy and of providing immediate access to (at least partial) sequence information. High-density membranes and radioactive probes may keep a niche for relatively small-scale work in non-specialized laboratories, using ready-made membranes provided by commercial suppliers; large-scale work will be performed in miniaturized format with non-radioactive probes. It will be possible in the near future to use single arrays containing a significant fraction of the 100,000 human genes; whether these will be microarrays or oligo chips, and whether the detection will involve fluorescence, color development or some new purely electrical detection method remain to be seen. A very important issue that has only been touched so far is that of methods and software to interpret this massive amount of data, correlate it with other available information and extract as much biological significance as possible. In this field, as in many others, bioinformatics is going to be an indispensable tool.

I wish to thank my co-workers who have been instrumental in setting up our hybridization signature system, and particularly Catherine Nguyen and Samuel Granjeaud. Thanks also to Corinne Bèziers La Fosse and Jean Davoust for help with the artwork, and to Pat Brown (Stanford University) and David Lockhart (Affymetrix) who provided permission and material for Figs. 3 and 5.

REFERENCES

1. http://www.ncbi.nlm.nih.gov/dbEST/index.html
2. Lennon, G.G. and Lehrach, H. (1991) Hybridization analyses of arrayed cDNA libraries. *Trends. Genet.* **7**, 314-317
3. Liang, P. and Pardee, A.B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257**, 967-971
4. Okubo, K., Hori, N., Matoba, R., Niiyama, T., Fukushima, A., Kojima, Y., and Matsubara, K. (1992) *Nature Genet.* **2**, 173-179
5. Velculescu, V.E., Zhang, L., Vogelstein, B., and Kinzler, K.W. (1995) Serial analysis of gene expression. *Science* **270**, 484-487
6. Nguyen, C., Rocha, D., Granjeaud, S., Baldit, M., Bernard, K., Naquet, P., and Jordan, B.R. (1995) Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. *Genomics* **29**, 207-215
7. Zhao, N., Hashida, H., Takahashi, N., Misumi, Y., and Sakaki, Y. (1995) High-density cDNA filter analysis, a novel approach for large-scale, quantitative analysis of gene expression. *Gene* **156**, 207-213
8. Pietu, G., Alibert, O., Guichard, V., Lamy, B., Bois, F., Leroy, E., Mariage-Samson, R., Houlgatte, R., Soularue, P., and Auffray, C. (1996) Novel gene transcripts preferentially expressed in human muscles revealed by quantitative hybridization of a high density cDNA array. *Genome Res.* **6**, 492-503
9. Bernard, K., Auphan, N., Granjeaud, S., Victorero, G., Schmitt-Verhulst, A.M., Jordan, B.R., and Nguyen, C. (1996) Multiplex Messenger Assay: simultaneous, quantitative measurement of expression for many genes in the context of T cell activation. *Nucleic Acids Res.* **24**, 1435-1443
10. Granjeaud, S., Nguyen, C., Rocha, D., Luton, R., and Jordan, B.R. (1996) From hybridization image to numerical values: a practical, high throughput quantification system for high density filter hybridizations. *Genetic Anal. Biomol. Eng.* **12**, 151-162
11. Gress, T.M., Hoheisel, J.D., Lennon, G.G., Zehetner, G., and Lehrach, H. (1992) Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. *Mamm. Genome* **3**, 609-661
12. Schena, M., Shalon, D., Davis, R.W., and Brown, P.O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467-470
13. DeRisi, J., Penland, L., Brown, P.O., Bittner, M.L., Meltzer, P.S., Ray, M., Chen, Y., Su, Y.A., and Trent, J.M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.* **14**, 457-460
14. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P.O., and Davis, R.W. (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* **93**, 10614-10619
15. Shalon, D., Smith, S.J., and Brown, P.O. (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res* **6**, 639-645
16. DeRisi, J.L., Iyer, V.R., and Brown, P.O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680-686
17. Khrapko, K.R., Lysov, Yu, P., Khorlin, A.A., Ivanov, I.B., Yershov, G.M., Vasilenko, S.K., Florentiev, V.L., and Mirzabekov, A.D. (1991) A method for DNA sequencing by hybridization with oligonucleotide matrix. *DNA Seq.* **1**, 375-388
18. Fodor, S.P., Read, J.L., Pirrung, M.C., Stryer, L., Lu, A.T., and Solas, D. (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science* **251**, 767-773
19. Southern, E.M., Maskos, U., and Elder, J.K. (1992) Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. *Genomics* **13**, 1008-1017
20. Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X.C., Stern, D., Winkler, J., Lockhart, D.J., Morris, M.S., and Fodor, S.P. (1996) Accessing genetic information with high-density DNA arrays. *Science* **274**, 610-614
21. Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E.L. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**, 1675-1680
22. Rocha, D., Carrier, A., Naspetti, M., Victorero, G., Anderson, E., Botcherby, M., Nguyen, C., Naquet, P., and Jordan, B.R. (1997) Modulation of mRNA levels in the presence of thymocytes and genome mapping for a set of genes expressed in mouse thymic epithelial cells. *Immunogenetics* **46**, 142-151
23. IMAGE consortium: http://www-bio.llnl.gov/bbrp/image/image.html
24. Clontech "ATLAS array": http://www.clontech.com/clontech/Catalog/Hybridization/Atlas.html
25. Genome Systems "Gene discovery array": http://www.genome-systems.com/GDA/
26. Synteni: http://www.synteni.com/
27. http://cmgm.Stanford.EDU/pbrown/
28. Fodor, S.A. (1997) Massively parallel genomics. *Science* **277**, 393-395
29. Wodicka, L., Dong, H., Mittmann, M., Ho, M.-H., and Lockhart, D.J. (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotechnol.* **15**, 1359-1367
30. Chen, J.J.W., Wu, R., Yang, P.C., Huang, J.Y., Sher, Y.P., Han, M.H., Kao, W.C., Lee, P.J., Chiu, T.F., Chang, F., Chu, Y.W., Wu, C.W., and Peck, K. (1998) Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. *Genomics*, in press