# Automated array technologies for gene expression profiling

Elmar Maier, Sebastian Meier-Ewert, David Bancroft and Hans Lehrach

Array-based technologies are poised to revolutionize the application of genomic research to drug discovery. These highly parallel techniques are able to simultaneously analyze up to hundreds of thousands of elements, each arrayed at high density across the surface of solid supports. This review provides a summary of some recently established array-based technologies for gene expression profiling. As an example, the authors highlight their experience with OligoFingerprinting – a technology platform that enables the generation of complete profiles of gene expression patterns – and compare it with conventional high-throughput cDNA sequencing.

Virtually all changes in health, regardless of their genetic or environmental causes, are mediated by changes in gene expression. Indeed it is believed that many common diseases are caused by changes in the expression pattern of genes, such as those coding for transcription factors or cell-surface receptors. To detect and quantify any changes in tissue-specific expression of such genes requires a technique with high sensitivity and large dynamic range. In principle, random sequencing of primary cDNA libraries generated from tissues of interest can fulfill these requirements, but only if between 50,000 and 500,000 cDNAs from each tissue-specific library are analyzed. Even with today's high-throughput sequencing technology, only a few thousand cDNAs per tissue can be sequenced economically. Thus, generating a profile of complete gene expression at sufficient depth for the identification of important rare genes as potential drug targets is unfeasible by random sequencing.

In this review, we describe established robotic technologies for the parallel analysis of tens or hundreds of thousands of cDNA clones arrayed at high density on solid supports. As an example we use OligoFingerprinting, a technology platform that enables the generation of complete profiles of gene expression patterns. Using this 'production line' approach we have analyzed more than half a million cDNA clones in the past year, generating gene expression profiles from several disease areas. A complete profile of a gene expression pattern is a unique tool in the characterization of disease onset and progression or mode of drug action. Because this pattern reflects the current status of the organ or tissue, it is key to the identification of new and highly specific drug targets.

## Disease and the genome

There is an immense amount of information contained within the human genome. Yet, for each cell or tissue type, only a very small subset of this information is transcribed as a tissue-specific population of mRNA molecules. Many of these mRNAs are then translated into proteins, forming highly complex interactions within and between cells or tissues of the developing, diseased or fully formed organisms. This combination of vast tracts of genomic DNA, tissue-, disease- and development-specific transcription of small subsets of this information, plus complex interactions between gene products, presents a daunting scientific and logistic challenge if a global picture of the function of

**Elmar Maier***, **Sebastian Meier-Ewert**, **David Bancroft** and **Hans Lehrach**, Max-Planck-Institut für Molekulare Genetik, Ihnestraße 73, D-14195 Berlin-Dahlem, Germany. tel: +49 30 8413 1232, fax: +49 30 8413 1380, e-mail: maier_e@mpimg-berlin-dahlem.mpg.de

  1359-6446/97/$17 00. PII: S1359-6446(97)01054-4 

gene products in the health of an individual is to be understood.

The ability to read and understand the biological function of information encoded within DNA sequences offers a completely new approach to biological research that has enormous consequences for basic research, medicine and especially the pharmaceutical industry. The identification and characterization of genes and gene products will play a major role in the understanding of the causes of many human diseases, and will provide many new drug targets for the pharmaceutical industry. Biological research must identify all genes, so that the signals governing cell specific expression and, ultimately, the interactions between genes and gene products on a functional level may be understood and the role of gene products in different disease states fully appreciated.

The analysis of gene products from man will be a fundamental step in the functional understanding of human disease. However, in many cases complementary organism-based model systems will also be required. Examples of these model systems are animal strains with mutated, extra or eliminated genes (e.g. mouse strains), organisms that offer better genetics and more easily accessible embryology (e.g. zebrafish) and organisms that offer suitable disease models or compound screening procedures (e.g. rat and rabbit). Although it will be necessary to analyze human genes, parallel investigations using the tissues of such model organisms will greatly enhance our understanding of human disease[1] and ultimately improve the efficiency of drug development.

## Role of bioinformatics

Vast amounts of biological information are rapidly becoming available; there are now more than 100 DNA sequence, genetic, structural, biochemical and biological databases available on the World Wide Web[2]. The use of computer-based information systems to analyze and integrate these resources with local data is commonly referred to as 'bioinformatics'. The ultimate goal of bioinformatics in this context is to discover the relationship between DNA sequences and their function. For example, if a locally analyzed DNA fragment appears to be a gene, a search for identifiable functional motifs may uncover a similarity to other genes of known sequence, structure, biochemistry and ultimately function. The sheer potential of bioinformatics to expand the range of biological information available to scientists is set to revolutionize the efficiency of drug target discovery.

Only through integration of biological science and bioinformatics will it be possible to identify and understand the biological function of almost all genes. There are more than 200 tissue types of interest in humans, subject to countless disease states and stages, each expressing tens of thousands of different genes, at levels that range across approximately five orders of magnitude. The consequent amount of biological data generated from human samples, combined with significant contributions from animal models, mean that highly automated technologies feeding directly into bioinformatic processes become an absolute necessity for a comprehensive analysis of complete gene expression patterns in disease.

## cDNA analyses

Analysis of cDNA libraries is a fundamental approach for the study of gene expression and has become an established method for the discovery and analysis of expressed genes[3–11]. There are two types of information that can be obtained from cDNA analyses:

- The spectrum of genes that are transcribed in the source tissue.
- The relative abundance of the transcripts in the source tissue when a primary cDNA library is used (a primary cDNA library is one in which the relative abundance of cDNA species closely represents that of the corresponding mRNA species).

The relative abundance of mRNA molecules transcribed from a given gene (and therefore also the resulting cDNAs) varies from 1 in 100 to 1 in 10,000,000 of the total cellular mRNA pool. It is therefore necessary to analyze at least $10^5$ cDNAs per tissue to have a significant chance of identifying most transcripts at least once or to obtain meaningful information on the abundance of transcripts. Even at this depth, some very rare transcripts will be missed. Therefore, in total, it will be necessary to analyze in the order of $10^7$ cDNA clones for *Homo sapiens* in order to obtain a comprehensive understanding of the flow of genetic information from the genome, via the messenger RNA, to the gene products, and may ultimately affect the development, physiology and pathology of the individual.

## What information is needed?

Key issues to be addressed at the outset of any investigation are how much of what kind of information should be

obtained to achieve the required goals, and what the most efficient means to obtain the information is, both in terms of time and cost.

As far as cDNA analyses are concerned, the solution to the information needs lies in determining the abundance of as many different species as possible, and in establishing which of the observed clones have been observed before. These are very powerful data to obtain, because they should facilitate a very large reduction in complexity of gene transcripts of particular biological relevance to the system under investigation.

Precisely this kind of information is generated by tag-sequencing large numbers of cDNA clones, as has been performed most notably by researchers in the USA[8,12–18]. Each tag generates enough information to identify homologies to known sequences stored in databases, permitting either an assignment, or categorization based on motif similarities. In order to truly characterize any given gene, a full-length clone must be isolated and sequenced several times, to ensure that the correct sequence has been determined. However, because many genes need to be completely sequenced only once, it is valuable to eliminate thoroughly characterized genes from further analyses. This aspect becomes increasingly pervasive as the proportion of human genes that have been identified, at least once, increases. This is not to say that once a gene has been sequenced in its entirety, no further characterization is necessary. Apart from the sequence, some of the most powerful pointers to gene function are its temporal and spatial expression patterns. This of course means that its occurrence in all tissues, disease and differentiation stages must be established.

Tag-sequencing of cDNAs is highly appropriate for the generation of the kind of information that allows the identification of gene sequences. However, in its present form, the technology is simply too expensive to be applied on a scale of multiple millions of clones.

## Library considerations

The construction of cDNA libraries is one of the most critical steps when considering their use in expression studies. Many months or even years of characterization will be invested in a library to ensure its highest possible quality. There are several choices to be made at the outset of generating a cDNA library. These include issues such as phage or plasmid cloning systems, directionality of cloning, priming of first strand synthesis, overexpression vectors and, not least, which reverse transcriptase to use.

At the outset of any project designed to study gene expression patterns, it is very important to decide what level of gene expression should be detected. Given the choice, any investigator would want to detect even the rarest transcripts. However, the benefits of the information gained should always be weighed against the experimental costs, both in terms of time and money. The probability of detecting transcripts of different expression levels as a function of library size is indicated in Figure 1. Nevertheless, if a 90%

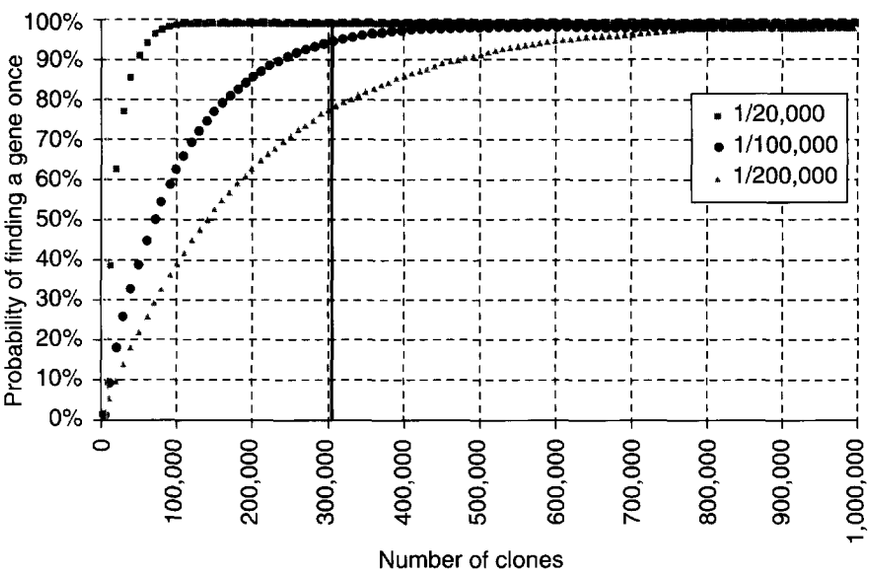


***Figure 1.*** *Probability of finding medium and rarely expressed genes in a specific cDNA library using random sequencing (gray vertical line) or OligoFingerprinting (thick black vertical line). This calculation assumes that there is no cloning bias in the gene of interest. This may not always be the case, because certain genes are not easily propagated in* E. coli. *The probability of finding moderately and rarely expressed genes using random gene sequencing at a conventional scale is very low (<10–20%). The OligoFingerprinting approach allows the analysis of over 300,000 cDNA clones on an economic scale. Here, the probability of finding medium and rarely expressed genes is 80–100%.*

probability of detecting a transcript that is represented once in $10^5$ mRNA molecules (a typical representation for many transcription factors) is required for instance, then a library of 300,000 clones would be necessary. Handling this number of clones as unique entities becomes a significant logistical problem that can be overcome only by large-scale automated processes and the use of arrayed libraries.

### Arrayed libraries and high-density grids

In recent years, the use of arrayed clone libraries has become an established tool in most genome analysis laboratories[4,6,7,19–25]. Many high-quality libraries are also available through commercial companies, and these often represent a valuable resource. Generally, the use of arrayed libraries has two main advantages. First, each clone that is analyzed (whether interesting in the context of a particular experiment or not) exists as a permanent reference in form of a frozen stock, meaning that usually no secondary rounds of screening are required, once an address in a micro-well plate has been identified. Second, an infinite number of copies of a library can be made, so that the same biological resource can be distributed to, and shared by, other investigators. Thus data generated in many laboratories in the world, using quite different experimental strategies, can be linked via a common factor, namely the clone library and associated clone addresses. Based upon this principle, several centers have been established that specialize in the distribution of arrayed libraries[9,22]. The generation and analysis of large arrayed libraries requires a high level of automation[26], and this is described in more detail below.

### Hybridization-based analysis of gridded cDNA libraries

The advantage of a hybridization-based approach is the high degree of parallelism that is achieved, because tens of thousands of samples can be hybridized simultaneously on high-density clone grids[23]. There are essentially two approaches to large-scale cDNA analysis by hybridization where cDNA clones are gridded in a high-density format on a suitable carrier.

In one approach, first-strand cDNA is generated from the total cellular mRNA extracted from a given tissue. This complex mix of cDNA probes is then labeled and hybridized to the gridded clones[7,11,27–31]. The representative levels of transcripts in the probe are derived from the signal intensity of the corresponding clone in the array. This approach is applicable to comparing the gene expression profiles of a few genes across multiple tissue and patient samples. However, the method suffers from some technical limitations. For example, because the sensitivity of this method is limited, especially where fluorescent labeled probes are used, up to 10 μg of poly-A mRNA may be required. Obtaining this amount of mRNA is possible only from sizeable tissue samples or cell culture models, and precludes the use of precious biopsy samples without disturbing the original mRNA abundances through *in vitro* amplification.

A second approach, OligoFingerprinting (Figure 2), is the use of short synthetic oligonucleotides as hybridization probes to profile complete gene expression patterns of large cDNA libraries[6]. Creation of these cDNA libraries requires approximately two orders of magnitude less mRNA as source material than would otherwise be necessary. OligoFingerprinting can be used as a very attractive alternative to classical gel sequencing techniques. Combining the hybridization results from several hundred different oligonucleotide probes that are each positive with about 2% of target clones can generate sufficient information for discrimination of all human genes[6,24]. The information generated by an oligonucleotide fingerprinting approach is of comparable utility to tag-sequencing, in that it allows the matching of clones to one another and to the theoretical fingerprints of database sequences (Figure 3). For clones with average inserts of 1,500 base pairs, approximately 200 hybridizations with specially selected octamer probes will allow the discrimination of over $10^5$ sequences at a likelihood ratio of 1 in 1,000,000. Because this approximates to the number of genes predicted to be in the human genome, 200 hybridizations should theoretically allow the discrimination of all sequences found in any human cDNA library.

In practice, the set of oligonucleotides used has undergone constant evaluation and improvement based on the hybridization frequencies and efficiencies of the individual oligonucleotides. As few as 130 specific and highly informative oligonucleotide hybridizations can provide enough information to identify positively clusters of cDNA clones and match them to databases at more than 90% confidence. No homology to any database entry can be found for 10–20% of human cDNA clones, and this figure increases to 75% if expressed sequence tag databases[18] are not included.

An additional attraction of OligoFingerprinting is that it is possible to identify subpopulations of clones that contain certain motifs of interest. Thus, an analysis can be focused rapidly on clones of particular interest, such as the presence of CAG repeats[32]. Using motif oligonucleotides, a directed

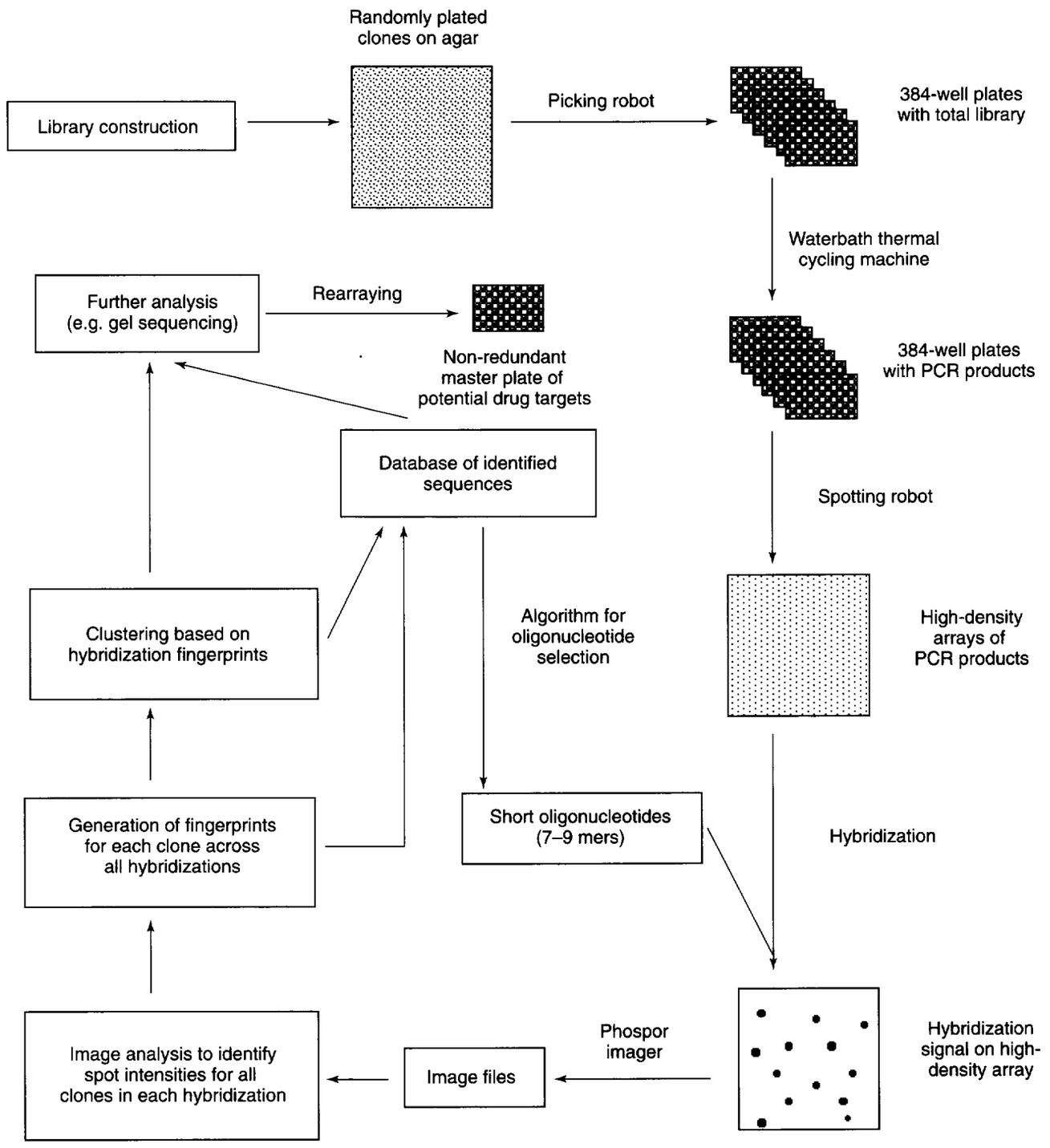


***Figure 2.*** *Steps involved in complete profiling of gene expression patterns by the OligoFingerprinting approach. The plated cDNA library is picked into 384-well microtiter plates, replicated and then PCR amplified in 384-well plates using a large waterbath thermal cycling machine. The PCR products are then arrayed on high-density grids with a spotting robot. The high-density grids are hybridized with a total of approximately 200 short oligonucleotides. The individual hybridization patterns are captured, for example, on a phosphor imager. Image files are then analyzed using custom-written software and fingerprints for all clones across all oligonucleotide hybridizations are generated. The clustered fingerprints are compared with theoretical fingerprints from known database sequences and unknown genes are analyzed in more detail.*

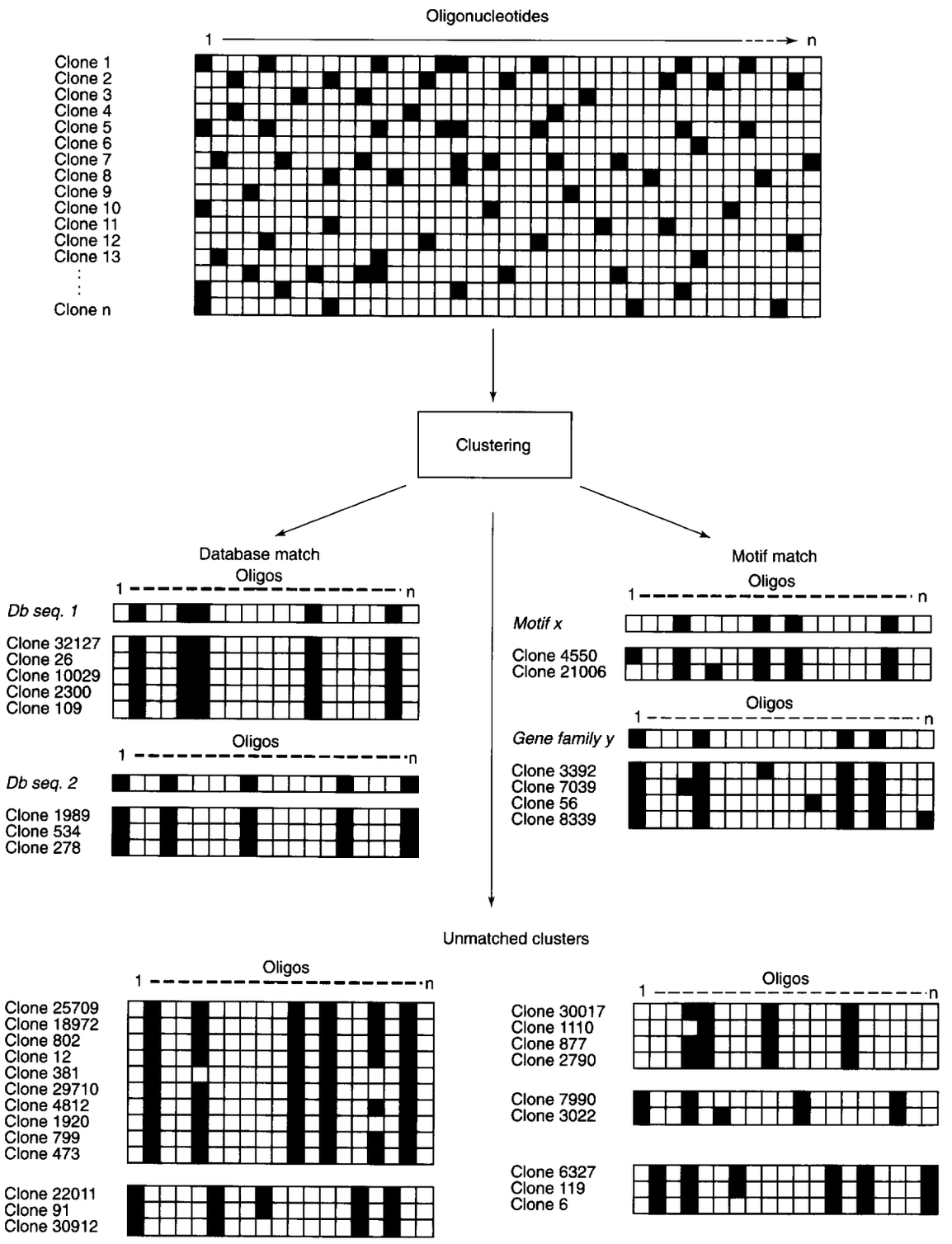


***Figure 3.*** *Clustering scheme for hybridization data generated by the OligoFingerprinting approach. The oligonucleotide fingerprints of the cDNA clones are clustered by pair-wise comparisons of the hybridization data for each clone. The fingerprints of the gene clusters are then compared with theoretical fingerprints generated from known gene sequences available in public databases. In general, unmatched clusters represent previously unknown genes.*

search for certain gene families and functional motifs can also be performed. Some motifs are not suitable for oligonucleotide probe design, because they are too short or variable, yet many of the important motifs for drug discovery can be defined, such as tyrosine kinase motifs, G-protein-coupled receptors and the TGF-β family. By designing oligonucleotides that recognize certain motif sequences characteristic of a given class of gene, such as DNA-binding domains, a highly enriched set of clones that contain features of particular interest can be created rapidly. Using such probes, clones can be classified according to motif classes or gene families, even in the absence of further sequence information. Such an approach is not possible in a classic gel sequencing project.

Generating oligonucleotide fingerprints is a not a trivial task and requires the integration of many experimental steps. The real challenge lies in implementing the procedure efficiently on a large scale. Only then can really significant savings be achieved, relative to more serial technologies, through the enormous scaling efficiency of highly parallel hybridization systems. We have integrated such systems and proven the speed and efficiency of the OligoFingerprinting approach[6] by building a 'production line' that, within an academic environment, enables a very large cDNA library (of the order of over 100,000 clones) to be analyzed. The individual experimental steps involved in the Oligo-Fingerprinting technology platform for complete profiling of gene expression patterns to be achieved are illustrated in Figure 3.

## Oligonucleotide arrays

An alternative school of thought proposes that, instead of depositing cDNA clones or gene fragments as an array, genetic material can be constructed *de novo* and *in situ* as high-density oligonucleotide arrays. The various applications of such oligonucleotide arrays have been reviewed recently by two of the main inventors of the technology[33,34]. Oligonucleotide arrays were initially proposed as high-throughput methods of sequencing by hybridization[35,36] or DNA diagnostics[37,38]. More recently, their use has also been demonstrated for simple cases of gene expression monitoring[27]. The use of complex cDNA hybridization mixtures as probes for gene expression profiling is associated with the same sample considerations as those of the first approach described in the previous section. There are also specific considerations of the manufacturing process for oligonucleotide arrays as well as the hybridization kinetics of all of the different oligonucleotides on the array[36,39–42]. The DNA sequence of the genes of interest must be known and used for the considerable task of designing highly specific oligonucleotide sequences for a great number of genes, all of which must have identical or similar hybridization kinetics.

The reproducible synthesis of equal amounts of many thousand oligonucleotides (20–25 mers) on high-density arrays necessary for gene expression profiling is very difficult using current photochemistry for the coupling reactions. For all approaches that depend on signal intensity from hybridization to predict gene expression, normalization of the target DNA is important. It is very difficult to apply quality control to these oligonucleotide arrays to normalize for the variations in oligonucleotide concentration on every element of the array. Repeated hybridizations of an individual array with various probes provide a reliable measure of the actual amount of target available for hybridization at each element of an array. Repeated hybridizations of DNA chips or oligonucleotide chips on glass or silicon surfaces have not yet proved reliable.

## Automation

### *Automated clone picking*

The first step towards the analysis of large libraries is the arraying of the clones into microtiter plates for long-term storage, analysis and subsequent individual retrieval. In recent years, several clone picking systems have been developed that rely on the use of a single- or a four-pin picking head[43,44]. Our clone-picking robot differs in several key aspects. The clone-picking feature has been integrated into a flat-bed robot system and a picking manifold with 96 spring-loaded pins, where each pin can be individually extended into a colony using a pneumatic actuator. Plate inoculation and pin sterilization are the two steps most limiting to the picking rate. The system is several times faster than previous devices, because all 96 pins are inoculated and sterilized at once. The system is capable of picking approximately 3,500 clones per hour into 384-well microtiter plates.

Hit rates of approximately 99% are being achieved even with very small colonies. To increase the success and rate of picking, calibration is undertaken automatically and variation in background intensity is corrected across the agar plate. A video camera scans a frame to identify colonies, the head of the picking device is then positioned and a pin is extended to pick the individual clone. System calibration typically takes 10 min after power-up and, once completed,

clone libraries can be picked off agar plates quickly, reliably and without any user intervention.

*Large-scale thermocycling robot*

A large-scale thermocycling robot for DNA amplifications of the cDNA libraries is used in our laboratories. Hybridization of short oligonucleotides to DNA targets requires purified insert DNA in order to avoid high background signals caused by the presence of cloning vector and host cell DNA. The most feasible way to prepare DNA from whole cDNA libraries is to perform a DNA amplification reaction on each clone separately. We have been able to adapt our reaction conditions in such a way that amplification can be performed directly in 384-well microtiter plates with a custom robotic thermocycler. Using three heated 225-litre waterbaths, up to 135 plates (51,840 reactions) can be cycled at a time. The basket of plates is moved from bath to bath using a pneumatic X–Z sliding configuration. Visual Basic software is used to control the whole system including robot motion, temperature probes and water-level sensors. The 384-well polypropylene plates used for thermocycling are heat sealed with two-sided plastic film and a heat sealer. The plastic film can easily be removed after the amplification step. After the amplification reaction, the DNA product is sufficiently pure and concentrated to be spotted directly onto nylon membranes, in preparation for hybridization.

*Automated clone gridding at high density*

Since the introduction of the 'high-density grid' concept for genome analysis by hybridization[23], several clone gridding devices have been built in various genome research centres. We have also built several of these gridding robots that are able to transfer clones stored in microtiter plates onto nylon membranes in high-density grids. Already, machines designed in our facility are used in several genetic screening or drug discovery centres where rapid, reliable and high-density screening of small quantities of biological or biologically active material is required. Various spotting heads can be accommodated by the system. Using spring-loaded devices with 384 pins of various tip diameters, a variety of biological materials can be spotted at different densities. Routinely, 15 hybridization filters with 57,600 clones are gridded in a duplicate pattern in around 3.5 h.

The microtiter plate stacking system has recently been upgraded to accommodate 72 plates so that the robot can run without any user intervention. A 'grabber' attached to the gridding head of the robot removes the individual plates from the microtiter plate rack and places them onto a plate holder, where the lid is automatically removed and the bar code is read. The bar-code reader supplies unique plate identifiers to a database of DNA source libraries and clones, making it an easy task to locate and retrieve colonies of interest following hybridization analysis of the selected grids. After the required number of spotting cycles, the lid is replaced and the grabber lifts the plate with lid back into the rack system and moves on to the next plate. Routinely, 57,600 clones are spotted on a 222 × 222 mm nylon filter, equivalent to 100 times the density of a standard microtiter plate. With the system described here we have now also achieved spotting densities of up to 230,400 clones per 222 × 222 mm nylon filter, equivalent to 400 times the density of a standard microtiter plate.

*Bioinformatics*

Efficient methods to analyze and handle the immense amount of data generated from hybridized grids are essential prerequisites for large-scale characterization of arrayed genetic libraries. The first step in this process is a reliable automated image-analysis system able to localize positive hybridization signals and subsequently quantify the results. Having tested several commercially available systems, none have been found to fulfill our data-handling requirements, and in-house software has therefore been developed for analysis of the hybridization patterns of positive signals at the high density at which cDNA clones are robotically gridded (Figure 4).

OligoFingerprinting data can be expanded by 'mining' public sources of biological data and integrating this information with locally generated data. In particular, it is possible to create theoretical OligoFingerprinting patterns from the vast resource of public gene sequences. These fingerprints can be used in two ways. First, the theoretical fingerprints can be used to link fingerprinted cDNA clones to a fully sequenced and annotated gene sequence – without the expense of directly sequencing these clones. Second, they can aid optimization of the design and use of specific oligonucleotide probes – in particular, probes that identify sequence motifs that characterize genes of common family or function. Sophisticated clustering algorithms have been developed to partition the vast number of random cDNA clones analyzed in our laboratories. These algorithms generate clusters of clones with identical or statistically similar fingerprint patterns. Clearly the number of clone members in a cluster is a measure of the expression level of that gene.

By matching the consensus fingerprint patterns of the clone clusters to the theoretical fingerprint derived from published gene sequences, the gene identity of many clusters can be determined. Importantly, there are many clusters, particularly of rare transcripts, with no matches to published gene sequences. Additionally, by matching the clusters to theoretical fingerprint patterns of known sequence motifs, other novel members of gene family of functional groups can be identified (Figure 3).

*Rearraying robot*

Bioinformatic processes can electronically identify many possible drug targets. However, there is a important distinction between identifying clones of interest and physically retrieving this subset of clones in a form amenable to further biological (and hence functional) analysis. In our system, clones of interest are retrieved robotically from large clone libraries as rearrayed subsets in new microtiter plates; these compressed clone libraries can be efficiently used in the verification of the drug targets across multiple patient samples.

Mother plates are handled automatically by the robot and the clones are selected automatically from bioinformatic output files to generate the subsets in the destination plate without any user intervention. The machine can rearray about 100 clones within 6 min, so that selections of a few thousand clones for further analysis can be undertaken very quickly.

***Figure 4.*** *Interface of an image analysis program for quantifying hybridization data of very high-density grids. The screen shows an image of an octamer hybridization on a high-density cDNA grid containing 57,600 clones. The regular hybridization pattern across the grid results from 'guide spots' for the image analysis software. These guide spots are always in the center of each block of 25 spots ('5 × 5 blocks'). After finding the 5 × 5 blocks of the spots, the program locates the 24 spots around each guide spot as shown on the top left hand corner. The hybridization signals of all clones on the high-density grid are quantified. For internal control and identification purposes, the cDNAs are spotted in special duplicate patterns as shown on the bottom left hand corner. Image analysis is only the first step in a complex software suite to analyze OligoFingerprinting data generated on a large number of cDNA clones. The quantified hybridization results of each cDNA clone across all oligonucleotides are analyzed by integrated bioinformatics systems in order to generate complete profiles of gene expression patterns.*

## Towards gene functional analysis

Gene discovery is just a small step towards drug development. Knowledge of a gene's sequence is almost meaningless in terms of its biological relevance or potential as a drug target, unless it can be determined in what context its information is utilized. The most important data to be subsequently gathered are measurements of the levels at which the gene is expressed in different cells, tissues and pathological states.

In order to facilitate the discovery of suitable drugs and drug targets against disease, knowing the identity of genes that alter their expression pattern during disease progression is invaluable. Classical gel sequencing has reached its zenith as a method to reveal gene expression profiles with the statistical meaning necessary for effective drug-target discovery. A cheaper and faster technology is required to carry the study of gene expression patterns to the next level – a level in which all active genes are analyzed in physiologically and pathologically interesting states. Because of the large number of samples that can be processed in automated systems, it becomes feasible to use highly parallel analysis of cDNA arrays, not only to discover new genes, but also to measure the levels at which genes are transcribed in different tissues or physiological states.

We have developed high-throughput technologies to enable complete profiling of gene expression, and these are

cheaper and faster than established methods. The arraying technologies and bioinformatic processes make it possible to analyze economically millions of cDNA clones – a scale made possible by an integrated, automated and highly parallel approach. Together with advanced bioinformatics tools, these technologies allow gene functional analysis at very high throughput. We envisage a further increase in throughput within the next year by implementing new developments in microarraying, detection and bioinformatic technologies.

## REFERENCES


1 Mangiarini, L. *et al.* (1996) *Cell* 87, 493–506
2 Benton, D. (1996) *Trends Biotechnol.* 14, 261–272
3 Sikela, J.M. and Auffray, C. (1993) *Nat. Genet.* 3, 189–191
4 Lennon, G.G. and Lehrach, H. (1991) *Trends Genet* 7, 314–317
5 Höög, C. (1991) *Nucleic Acids Res.* 19, 6123–6127
6 Meier Ewert, S. *et al.* (1993) *Nature* 361, 375–376
7 Gress, T.M. *et al.* (1992) *Mamm. Genome* 3, 609–619
8 Adams, M.D. *et al.* (1995) *Nature* 377, 3S–174S
9 Lennon, G. *et al.* (1996) *Genomics* 33, 151–152
10 Pietu, G. *et al.* (1996) *Genome Research* 6, 492–503
11 Gress, T.M *et al.* (1996) *Oncogene* 13, 1819–1830
12 Adams, M.D. *et al* (1991) *Science* 252, 1651–1656
13 Kahn, A.S. *et al.* (1992) *Nat. Genet.* 2, 180–185
14 Okubo, K. *et al.* (1992) *Nat. Genet.* 2, 173–179
15 Adams, M.D. *et al* (1992) *Nature* 355, 632–634
16 Adams, M.D. *et al.* (1993) *Nat. Genet.* 4, 373–380
17 Adams, M.D. *et al.* (1993) *Nat. Genet.* 4, 256–267
18 Boguski, M.S., Lowe, T.M.J. and Tolstoshev, C.M. (1993) *Nat. Genet.* 4, 332–333
19 Hoheisel, J.D. *et al.* (1993) *Cell* 73, 109–120
20 Nizetic, D. *et al.* (1991) *Proc. Natl. Acad. Sci. U. S. A.* 88, 3233–3237
21 Maier, E. *et al.* (1992) *Nat. Genet.* 1, 273–277
22 Zehetner, G. and Lehrach, H. (1994) *Nature* 367, 489–491
23 Lehrach, H. *et al.* (1990) *Genome Analysis. Genetic and Physical Mapping* (Vol. 1) (Davis, K.E. and Tilghman, S.M., eds), pp. 39–81, Cold Spring Harbor Laboratory Press
24 Drmanac, R. *et al.* (1990) *The First International Conference on Electrophoresis, Supercomputing, and the Human Genome* (Camor, C. and Lim, H., eds), pp 60–75, World Scientific
25 Schalkwyk, L.C., Francis, F. and Lehrach, H. (1995) *Curr. Opin. Biotechnol.* 6, 37–43
26 Maier, E. *et al.* (1994) *J. Biotechnol.* 35, 191–203
27 Lockhart, D.J. *et al.* (1996) *Nat Biotechnol* 14, 1675–1680
28 Derisi, J. *et al.* (1996) *Nat. Genet.* 14, 457–460
29 Schena, M. *et al.* (1996) *Proc. Natl. Acad. Sci. U. S. A.* 93, 10614–10619
30 Shalon, D., Smith, S.J. and Brown, P.O. (1996) *Genome Research* 6, 639–645
31 Schena, M. *et al.* (1995) *Science* 270, 467–470
32 Neri, C *et al.* (1996) *Hum. Mol. Genet.* 5, 1001–1009
33 Mirzabekov, A.D. (1994) *Trends Biotechnol* 12, 27–32
34 Southern, E.M (1996) *Trends Genet.* 12, 110–115
35 Southern, E.M., Maskos, U. and Elder, J.K. (1992) *Genomics* 13, 1008–1017
36 Khrapko, K.R. *et al.* (1991) *DNA Sequence* 1, 375–388
37 Yershov, G. *et al* (1996) *Proc. Natl Acad Sci U S. A.* 93, 4913–4918
38 Hacia, J G *et al.* (1996) *Nat. Genet* 14, 441–447
39 Maskos, U. and Southern, E.M. (1992) *Nucleic Acids Res.* 20, 1675–1678
40 Maskos, U. and Southern, E.M. (1993) *Nucleic Acids Res.* 21, 2267–2268
41 Casegreen, S.C. and Southern, E.M. (1994) *Nucleic Acids Res.* 22, 131–136
42 Hoheisel, J.D. (1996) *Nucleic Acids Research* 24, 430–432
43 Jones, P. *et al.* (1992) *Nucleic Acids Research* 20, 4599–4606
44 Uber, D.C. *et al.* (1991) *BioTechniques* 11, 642–645