Bertrand Jordan *Editor*

Microarrays in Diagnostics and Biomarker Development

Current and Future Applications

Springer

Microarrays in Diagnostics and Biomarker Development

Bertrand Jordan
Editor

Microarrays in Diagnostics and Biomarker Development

Current and Future Applications

Springer

Editor
Bertrand Jordan
CoReBio PACA
Luminy Science Park
13288 Marseille Cedex 9
France

ISBN 978-3-642-28202-7 ISBN 978-3-642-28203-4 (eBook)
DOI 10.1007/978-3-642-28203-4
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2012936972

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This is the second time I assemble a book on microarrays for Springer—almost exactly 10 years after the first one (*DNA microarrays: Gene expression applications*. Springer Berlin Heidelberg 2001). During this interval, we have seen tremendous changes in the landscape of genomics and molecular biology. The first human genome sequence, with its revelations on gene number and relatively low human genetic diversity, was shortly followed by extensive investigation of single nucleotide polymorphisms in our DNA and, with the help of more and more complex "SNP arrays", by extensive whole-genome association studies (GWAS). These have uncovered many gene/phenotype correlations, but have also revealed how limited our understanding of the functional genome still is. The importance of non-coding transcripts and of epigenetic modifications of our DNA has been realised. And very recently, DNA sequencing technology has made tremendous progress, to the point that the "1,000-dollar genome" is in sight and that clinical use of whole-genome or exome sequencing is becoming significant (and is in fact competing with some uses of arrays). Meanwhile, clinical applications of DNA microarrays have developed, but not exactly in the directions anticipated 10 years ago and not quite at the level predicted by some analysts. Other microarrays using proteins, antibodies, peptide, or aptamer molecules as probes have made significant advances in spite of technical problems, and now have an impact not only in research but also in the clinic and in biomarker development.

This new book includes 13 main chapters, and covers essentially all types of arrays. It is focussed on entities that are in actual clinical use, or quite close to it—thus it does not discuss, for example, complex SNP arrays that remain essentially a research tool. It does present some very recent developments, such as peptide or aptamer arrays, or also miniaturisation towards "nanoarrays", that I see as having great potential in medicine even though their current presence is still limited. It discusses in detail very important issues in bioinformatics and in statistical analysis of array data, as well as the hurdles faced in the commercialisation of array-based tests and the vexing IP issues associated with these activities.

I believe this book will be useful to current array users who wish to have a complete view of the field, to newcomers who attempt to make the best choice between different technologies, and to academic scientists who engage in technology transfer activities and need to evaluate the hurdles involved in this process. I am especially grateful to the authors, my colleagues, who kindly accepted to undertake the absorbing task of writing a chapter, especially at a time where tight funding means that they are extremely busy preparing grant applications. And, of course, the organisational support from Springer to overcome the 1,001 small problems encountered in the production of this book is gratefully acknowledged.

Marseille, France Bertrand Jordan

Contents

Chapter 1
The Microarray Paradigm and Its Various Implementations

Bertrand Jordan

1.1 The Beginnings

The essence of the microarray approach is its "highly parallel" nature, that is, the fact that information on a large number of entities is obtained in a single experiment, thanks to a regular arrangement of "probes" (whatever their molecular nature may be) on some kind of support and on assessment of signals detected after incubation with a sample (usually labelled). The origins of this technology can be traced to the widely practised method of "colony screening" that was used, once the cloning revolution had begun in the early 1970s, to identify a few bacterial or phage clones of interest among the many thousands present in a given library (Fig. 1.1). It is true that the arrangement of clones in such experiments was random and therefore unrepeatable, but it was only one step from there to "ordered libraries" in which individual genomic or cDNA clones had been picked from Petri dishes, then distributed into microtiter plates, and could subsequently be arrayed on a suitable support (usually a nylon or nitrocellulose membrane). The essential feature was that detection of a positive (radioactive) signal at a given position on the membrane immediately identified specific clone(s) that could then be recovered from storage and used for further experiments (Fig. 1.2). This scheme was largely pioneered by Hans Lehrach's group in Germany, who adapted or developed the necessary instruments to perform these sophisticated (for the time) manipulations (Michiels et al. 1987; Craig et al. 1990; Lennon and Lehrach 1991). It was just a further step from there to use such "high-density filters" to assess gene expression, using arrays of cDNA clones (or their DNA) and labelled cDNA prepared from total mRNA of a given cell line or tissue (Gress et al. 1992), as well as the newly appeared imaging plate systems (Amemiya and Miyahara 1988) that allowed quantification of the signals. This approach was used by several groups to perform studies measuring simultaneously the

B. Jordan (✉)
CoReBio PACA, Luminy Science Park, 13288 Marseille, Cedex 9, France
e-mail: bertrand.jordan@univmed.fr

B. Jordan (ed.), *Microarrays in Diagnostics and Biomarker Development*,
DOI 10.1007/978-3-642-28203-4_1,

Fig. 1.1 A "colony filter" as used *circa* 1980 to find genes in clone libraries. Bacterial clones containing human DNA inserts are spread on nutrient agar in a Petri dish (ca 3,000 per dish). After growth of colonies, a nitrocellulose filter is placed on the plate, then removed and treated to lyse the bacteria and denature the DNA. The filter is then incubated with a radioactive probe corresponding to the gene of interest. After exposure to X-ray film, the "positive" clones can be seen and can be picked from the original Petri dish using the orientation marks drawn in radioactive ink

expression of hundreds or thousands of genes, an impressive feat at the time (Nguyen et al. 1995; Zhao et al. 1995) and, together with systematic tag sequencing of cDNA clones as pioneered by Craig Venter (Adams et al. 1991), gave the first general outlook on the human transcriptome as well as on that of a number of model organisms.

1.2 The Essential Miniaturisation Step

The detection system used in these experiments was radioactivity, that is, the cDNA prepared from the mRNA of the sample was labelled by incorporation of radioactive nucleotides containing 32P or 33P phosphate. Although this provided high sensitivity and a wide dynamic range, it severely limited the possible resolution and was also poorly adapted to many working environments. Thus, the first demonstration of an expression array using fluorescent labelling of the sample and detection with a confocal system (Schena et al. 1995; 1996) represented an important advance and is generally considered as the beginning of the microarray era. Meanwhile,

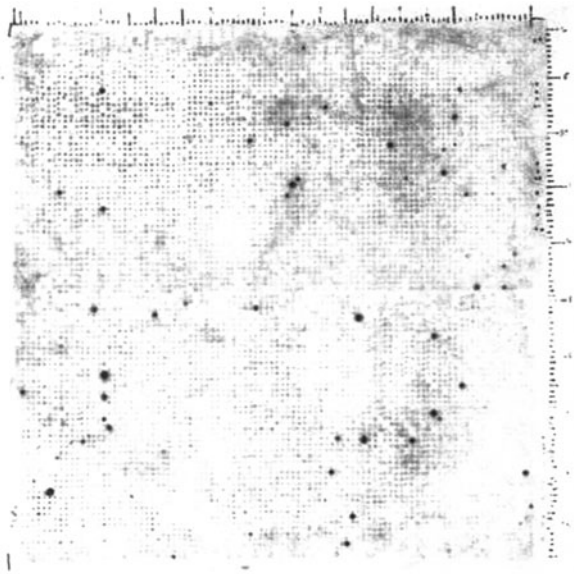

Fig. 1.2 A "high-density filter" used to access clone libraries (*circa* 1990). cDNA clones that have been picked from a library into microtiter plates are arrayed (100 × 100) on a large nylon filter. Hybridisation of this filter with a radioactive probe specific for the gene of interest detects several positive clones that can be directly recovered from storage for further study. Successive hybridisations of the same filter (or of replicas thereof) accumulate information on the 10,000 clones represented

the use, as probes, of oligonucleotides synthesised in situ had been described by Ed Southern as early as 1989 (see Fig. 1.3; Southern et al. 1992) and later by Steve Fodor (Pease et al. 1994). Rapid development followed, with commercial suppliers progressively taking over from research laboratories, and expression profiling using DNA arrays became a major approach in the late 1990s and early 2000s, with progressively more complex arrays that generally used oligonucleotides (pre-synthesised or synthesised in situ) as probes, and relied on fluorescent detection, although radioactivity kept some supporters and colorimetric methods were also implemented (Chen et al. 1998). By 2001, it had become possible to assess all (or most) of the human genes using a small set of commercial arrays, and by 2004, a single chip gave access of the whole human transcriptome.

1.3 Why Such Rapid Success?

The adoption of microarray technology (essentially for expression profiling in this first phase) was extremely rapid, both in terms of publications (hundreds in 2000, thousands in 2005) and of market size. It is clearly a technology that came at the right time: knowledge on genes and genomes had been accumulating for some time,

SYNTHESIS OF OLIGONUCLEOTIDES TETHERED TO A GLASS SURFACE: APPLICATIONS IN THE ANALYSIS OF NUCLEIC ACID SEQUENCES.
E M Southern and U Maskos; Department of Biochemistry, University of Oxford, U.K.

We have developed a new linker for oligonucleotide synthesis which leaves the oligonucleotide tethered to the glass surface after the deblocking step. The support bound oligonucleotide can be used for molecular hybridisation. The glass support can be in the form of microspheres, in which case synthesis and hybridisation can be conveniently carried out in columns.

We have also synthesised oligonucleotides in patches on the surface of microscope slides. This has the advantage that several different sequences can be synthesised on the same slide and the hybridisation signal detected by autoradiography. Applications to the analysis of nucleic acid sequences will be discussed.

Fig. 1.3 Abstract of the talk presented by Ed Southern at the 1989 "Genome mapping and sequencing" meeting at Cold Spring Harbor Laboratory (reproduced from the abstract book provided at the meeting). This is the first recorded example of oligonucleotide arrays, later published by the same authors (Southern et al. 1992)

and large-scale sequencing projects, after some initial difficulties, were getting on stream, one of the early landmarks being the sequence of yeast chromosome III (Oliver et al. 1992), all of 315 kb to be followed by much larger sequences in the following years, culminating with the first human sequence in 2001/2003. It was clear that there would be a pressing need to bring some functional significance to all these sequences, and expression profiling with microarrays appeared able to do so on the required scale. By the late 1990s, technology had advanced to the stage where this was feasible, with liquid handling and "spotting" robots to manufacture the arrays, and high-end confocal scanning devices for data acquisition. Informatics had also progressed, with the advent of the personal computer during the 1980s and of practical Internet at the beginning of the 1990s, so the required information processing and bioinformatics, while largely underestimated at the beginning, appeared feasible.

1.4 Generalisation of the DNA Array Approach

DNA arrays did not remain limited to expression profiling. Indeed, they had initially been thought of as sequencing devices (Khrapko et al. 1991; Southern et al. 1992), but this application did not succeed in providing a viable alternative to Sanger sequencing and was (provisionally) abandoned. However, another application progressively gained huge acceptance, the scoring of SNPs (single nucleotide

polymorphisms or "snips") in human (or other) DNA. Large-scale genetic mapping, after the early eras of RFLPs and then of microsatellites, became based on the much more abundant snips (International HapMap Consortium 2005) and dependent on evermore complex oligonucleotide arrays that are able to assess first 10,000, then 500,000 and today several million snips in a person's DNA in a single experiment. These were intensively used in GWAS (genome-wide association studies) that revealed many genes influencing the risk for common, complex genetic diseases and also in DTC (direct to consumer) profiling of dubious predictive value but of strong commercial interest. In parallel, the realisation in the mid-2000s of the frequent occurrence in human DNA of insertions, deletions, duplications and inversions, collectively called CNVs, copy number variations (Redon et al. 2006), some of which have clinical relevance, motivated the development of specialised arrays (variably termed CGH arrays, CytoChips or CNV arrays) that today represent the major clinical use of DNA arrays.

1.5 Going Beyond DNA

The striking advantages of microarrays (parallel acquisition of many data points in controlled conditions and at moderate expense) did not escape the attention of scientists and clinicians in other fields, and extensive development efforts have been aimed for more than 10 years at protein, peptide, carbohydrate, tissue and even living cell arrays. Protein arrays in particular have been investigated in detail and under many different implementations (Joos and Bachmann 2009); they have great potential (and a number of actual uses) for clinical applications (Yu et al. 2010). Their development has nevertheless been relatively slow since many favourable features of nucleic acids are absent: procurement of the protein (antigen or antibody) "probes" to be spotted on the arrays is quite difficult (no easy synthesis as for oligonucleotides, no PCR for amplification), the protein/protein interaction on which detection is based is complex and very sensitive to many factors (unlike hybridisation of DNA and RNA, that is relatively well understood) and the data acquisition often requires complex two-step protocols ("sandwich" assays) that introduce further variability. Nevertheless, protein arrays are indeed being used in research and (usually in fairly low-plex form) in clinical diagnostics and can be expected to assume increasing importance. Peptide arrays have also been developed and successfully used, in particular for assays of kinase activities that have great importance in cancer drug development and use (Hilhorst et al. 2009). Carbohydrate arrays have also been reported (Chang et al. 2010) and could become important as glycosylation changes play a significant role in cancer. A particular and relatively low-tech application of arrays that has a major role in the clinic is represented by tissue arrays (Kononen et al. 1998). This consists of assembling several hundred slices of paraffin-fixed tissue from normal and pathological samples on a single glass slide. This slide can then be incubated with a given antibody, and the specificity of this reagent can be examined in detail (under the microscope) for each of the samples and correlated with the clinical information

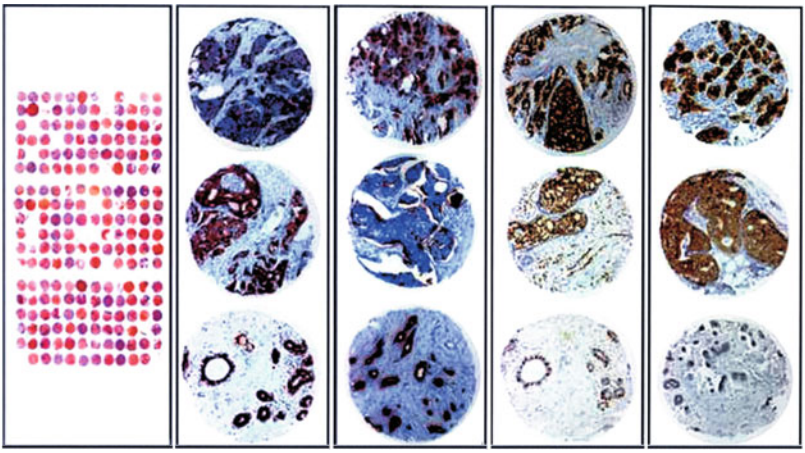

Fig. 1.4 Example of a tissue array. *Left*, general view of a microscope slide on which ~200 thin slices from as many pathological samples embedded in paraffin ("FFPE samples") have been assembled; *right*, blow-up of three of these slices after staining with four different antibodies. This allows precise characterisation of reagents on patient samples for which detailed clinical information is recorded

recorded. This simple application (Fig. 1.4) has great value and is a good example of the power of the microarray paradigm. Living cell arrays have also been developed for specialised applications such as large-scale RNAi (inhibitory RNA) screening (Wu et al. 2011).

1.6 Organisation of This Book

This book aims to provide an overview of current microarrays (DNA-based or otherwise), with emphasis on clinical and biomarker applications. Part I deals with DNA arrays in their various forms and uses; Part II with protein and peptide arrays, including also aptamer-based arrays that are protein-like in spite of their chemical structure; Part III with tissue arrays and miniaturised implementations of the microarray approach; and, finally, Part IV with data analysis and validation issues as well as commercial, regulatory and intellectual property aspects.

References

Adams MD, Kelley JM, Gocayne JD, Dubnick M, Polymeropoulos MH, Xiao H, Merril CR, Wu A, Olde B, Moreno RF et al (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. Science 252:1651–1656

Amemiya Y, Miyahara J (1988) Imaging plate illuminates many fields. Nature 336:89–90

Chang SH, Han JL, Tseng SY, Lee HY, Lin CW, Lin YC, Jeng WY, Wang AH, Wu CY, Wong CH (2010) Glycan array on aluminum oxide-coated glass slides through phosphonate chemistry. J Am Chem Soc 132:13371–13380

Chen JJW, Wu R, Yang PC, Huang JY, Sher YP, Han MH, Kao WC, Lee PJ, Chiu TF, Chang F, Chu YW, Wu CW, Peck K (1998) Profiling expression patterns and isolating differentially expressed genes by cDNA microarray system with colorimetry detection. Genomics 51:313–324

Craig AG, Nizetic D, Hoheisel JD, Zehetner G, Lehrach H (1990) Ordering of cosmid clones covering the herpes simplex virus type I (HSV-I) genome: a test case for fingerprinting by hybridisation. Nucleic Acids Res 18:2653–2660

Gress TM, Hoheisel JD, Lennon GG, Zehetner G, Lehrach H (1992) Hybridization fingerprinting of high-density cDNA-library arrays with cDNA pools derived from whole tissues. Mamm Genome 3:609–661

Hilhorst R, Houkes L, van den Berg A, Ruijtenbeek R (2009) Peptide microarrays for detailed, high-throughput substrate identification, kinetic characterization, and inhibition studies on protein kinase A. Anal Biochem 387:150–161

International HapMap Consortium (2005) A haplotype map of the human genome. Nature 437:1299–1320

Joos T, Bachmann J (2009) Protein microarrays: potentials and limitations. Front Biosci 14:4376–4385

Khrapko KR, Lysov YuP, Khorlin AA, Ivanov IB, Yershov GM, Vasilenko SK, Florentiev VL, Mirzabekov AD (1991) A method for DNA sequencing by hybridization with oligonucleotide matrix. DNA Seq 1:375–388

Kononen J, Bubendorf L, Kallioniemi A, Bärlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP (1998) Tissue microarrays for high-throughput molecular profiling of tumor specimens. Nat Med 4:844–847

Lennon GG, Lehrach H (1991) Hybridization analyses of arrayed cDNA libraries. Trends Genet 7:314–317

Michiels F, Craig AG, Zehetner G, Smith GP, Lehrach H (1987) Molecular approaches to genome analysis: a strategy for the construction of ordered overlapping clone libraries. Comput Appl Biosci 3:203–210

Nguyen C, Rocha D, Granjeaud S, Baldit M, Bernard K, Naquet P, Jordan BR (1995) Differential gene expression in the murine thymus assayed by quantitative hybridization of arrayed cDNA clones. Genomics 29:207–215

Oliver SG, van der Aart QJ, Agostoni-Carbone ML, Aigle M, Alberghina L, Alexandraki D, Antoine G, Anwar R, Ballesta JP, Benit P et al (1992) The complete DNA sequence of yeast chromosome III. Nature 357:38–46

Pease AC, Solas D, Sullivan EJ, Cronin MT, Holmes CP, Fodor SP (1994) Light-generated oligonucleotide arrays for rapid DNA sequence analysis. Proc Natl Acad Sci USA 91:5022–5026

Redon R, Ishikawa S, Fitch KR et al (2006) Global variation in copy number in the human genome. Nature 444:444–454

Schena M, Shalon D, Davis RW, Brown PO (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270:467–470

Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW (1996) Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 93:10614–10619

Southern EM, Maskos U, Elder JK (1992) Analyzing and comparing nucleic acid sequences by hybridization to arrays of oligonucleotides: evaluation using experimental models. Genomics 13:1008–1017

Wu N, Castel D, Debily MA, Vigano MA, Alibert O, Mantovani R, Iljin K, Romeo PH, Gidrol X (2011) Large scale RNAi screen reveals that the inhibitor of DNA binding 2 (ID2) protein is

repressed by p53 family member p63 and functions in human keratinocyte differentiation. J Biol Chem 286:20870–20879

Yu X, Schneiderhan-Marra N, Joos TO (2010) Protein microarrays for personalized medicine. Clin Chem 56:376–387

Zhao N, Hashida H, Takahashi N, Misumi Y, Sakaki Y (1995) High-density cDNA filter analysis: a novel approach for large-scale, quantitative analysis of gene expression. Gene 156:207–213

Part I
DNA Arrays

Chapter 2
DNA Arrays in Many Guises

Bertrand Jordan

2.1 The Birth and Development of DNA Arrays

The first generally recognised publication on DNA microarrays appeared in 1995 (Schena et al. 1995), although a number of DNA array studies had been published previously, generally using nitrocellulose or nylon supports and radioactive labelling of the sample, options that were not compatible with extensive miniaturisation (Lennon and Lehrach 1991; Gress et al. 1992; Zhao et al. 1995; Nguyen et al. 1995). In-house production of microarrays was widespread in the first years of the technology (DeRisi and Iyer 1999), but industry gradually took over and provided great advances in quality control and, ultimately, real cost. The DNA "probes" used to construct arrays were initially amplification products from cDNA clones, but these were gradually superseded by oligonucleotides: management of large cDNA libraries is very cumbersome and error-prone and, contrary to some expectations, relatively short oligonucleotides turned out to provide satisfactory specificity on hybridisation. Currently, essentially all DNA microarrays consist of oligonucleotides arrayed on a solid support, usually glass, but also occasionally plastic or silicon. Where they differ is in the length of oligonucleotides. These may be short (20–25 bases) and synthesised in situ, as for Affymetrix, in which case a number of overlapping sequences are used to assess the expression of every gene, or long (50–70 bases), synthesised in situ (Agilent, Roche Nimblegen) or pre-synthesised (Illumina), in which case a single oligonucleotide may be sufficient to measure the expression of one gene. Each of the four major approaches has its strong and weak points. The microelectronics-like light-directed synthesis process developed at Affymetrix allows the efficient production of extremely complex chips but is relatively inflexible as any change in content involves expensive retooling (manufacture of a new set of very high precision masks); the Nimblegen process is also based on light-directed synthesis but uses a programmable digital

B. Jordan (✉)
CoReBio PACA, Luminy Science Park, 13288 Marseille, Cedex 9, France
e-mail: bertrand.jordan@univmed.fr

B. Jordan (ed.), *Microarrays in Diagnostics and Biomarker Development*,
DOI 10.1007/978-3-642-28203-4_2,

micromirror device[1] instead of masks, which makes it much more flexible. Agilent's in situ manufacturing system, based on successive delivery of synthesis reagents to each spot, is readily programmable and extremely flexible, well adapted to production of custom arrays, but complexity is somewhat lower and economies of scale less evident. Finally, the Illumina process relies on pre-synthesised long oligonucleotides and is extremely miniaturised but, again, not very flexible (see Table 2.1).

Progress in miniaturisation has resulted in arrays that can monitor the expression of all known human genes or score up to one million single nucleotide polymorphisms. Spot or "feature" size is now typically below 50 μm and can be as low as a few micrometres. In research applications, the sample is labelled with a fluorophore (usually Cy3 and/or Cy5), and data acquisition is achieved by laser scanning and confocal detection. This provides for very high resolution and sensitivity, allowing the use of complex arrays, but requires expensive apparatus and relatively cumbersome procedures. For clinical purposes, where the array is often less complex, labelling may involve a fluorophore, but also chemiluminescence or even colorimetry, and detection with a very affordable CCD unit may be adequate. In some cases, electrical detection systems allow the assay of unlabelled samples. In terms of cost, expression arrays covering the whole human transcriptome can now be obtained (in quantity) for as little as €100, down from several thousand at the time of their introduction in the early 2000s.[2] Diagnostic applications often call for moderately complex arrays, but the cost of an array is now more dependent on the size of the production run than on its complexity; thus, relatively simple custom arrays may be as expensive as highly complex standard chips. Of course, in diagnostics, cost issues are essential; thus, the compromises made and the choice of manufacturer can be quite different from those that apply in research. In most cases, manufacturers now provide their products as multi-array slides (e.g. 12 arrays of 135,000 oligonucleotides each for Nimblegen), a format that helps to bring down the per-array cost and helps with reproducibility issues but introduces some operational constraints.

2.2 Trends in Research

During the initial years, DNA arrays were essentially used for expression studies, and expression profiling became one of the major exercises in large-scale genome analysis, especially as, at that time, sequencing technology was still based on the Sanger technique, very slow and expensive by present-day standards. Of course, compared to previous methods (essentially Northern blotting), microarrays were incredibly powerful, and accordingly, their uptake by the research community was

[1] Similar to the devices found in video projectors.

[2] In March 2001, the list price, in Europe, for a set of five Affymetrix arrays (U95 A to E) representing the whole human transcriptome (as far as it was known at the time) was €6,400.

Table 2.1 Four major approaches to DNA array manufacture

Company	Outline of method	Typical feature size (μM)	Oligonucleotide length	Number of oligonucleotides per gene (expression)	Number of features per array	Cost for large production runs	Flexibility
Affymetrix	In situ light-directed synthesis (mask lithography) of short oligonucleotides	10	15–20 nt	10–20	Up to 1 M	Very low	Limited (retooling necessary)
Roche Nimblegen	In situ light-directed synthesis (programmable micromirror device) of long oligonucleotides	13	60 nt	3	Up to 4 M	Low	Good (reprogramming of micromirror device)
Agilent	In situ synthesis of long oligonucleotides by programmed reagent delivery	50	60–70 nt	1–2	Up to 200 K	Low	Excellent (reprogramming, custom order on Internet)
Illumina	Pre-synthesised long oligonucleotides deposited in several microwells	3	60–70 nt	~100 identical in different wells	Up to 4 M	Very low	Relatively limited

This table shows the salient features of the approach adopted by four major manufacturers of arrays. Given the wide variety of arrays produced by these companies, the figures given can only be an approximation
http://media.affymetrix.com/support/technical/other/clinical_toolkit_product_spotlight.pdf
http://www.nimblegen.com/company/technology/
http://www.genomics.agilent.com/
http://www.illumina.com/applications/

extremely rapid: the number of publications on expression profiling rose from a few in 1995 to several hundred in 2000 and nearly 2,000 in 2005 (Lenoir and Giannella 2006). It quickly became apparent, however, that much of the early data was of dubious validity because of technical issues with the platforms used, insufficient use of replicates and over-optimistic interpretation. Serious as they are for research projects, these problems were an even stronger deterrent for clinical applications. However, with growing realisation of the importance of these issues (Tan et al. 2003; Jordan 2004), these problems were tackled and gradually resolved, thanks to industry standardisation of arrays and to the efforts of the FDA-sponsored MAQC Consortium (2006). Currently, however, profiling methods based on sequencing (RNA-seq) are becoming dominant in research applications, as they provide higher quality data (digital rather than analogue), with information on mutations and on "new" transcripts, at a cost that is now approaching that of an expression array experiment.

Meanwhile, the development of DNA arrays capable of scoring large numbers of single nucleotide polymorphisms (SNPs or snips) made great progress, and by the early 2000s, chips able to assess 10,000 snips in a single experiment were available and became widely used (Gunderson et al. 2006). Current snip arrays can now score 1,000,000 or more of these markers and have been essential for whole-genome association studies in which several thousands of patients and controls are genotyped to discover significant associations between loci in the genome and vulnerability to common diseases such as diabetes, Crohn's disease or even autism (Lam et al. 2010). While these studies have their problems, they do provide valid and useful results and have been extensively practised, providing a large market for these snip arrays that are now the dominant species in the DNA array world. Snip arrays are also used for direct to consumer (DTC) genetic profiles whose predictive value is however extremely limited.

Beyond snip genotyping, the growing recognition of the existence of numerous deletions, duplications and insertions (collectively called structural variations) in our DNA, some of which display clear clinical correlations, has motivated the development of arrays capable of detecting these events. Suitable probes may be included in snip arrays, or, alternatively, specialised "CGH arrays" may be developed to probe the whole genome for structural variations at various levels of resolution. This has become an important application of DNA arrays in research and, indeed, also in the clinic (see Chap. 5).

In summary, research use of DNA arrays has largely shifted from expression profiling to snip scoring and structural variation studies but remains very significant in spite of strong competition by next-generation sequencing for some applications.

2.3 Great Expectations for Clinical DNA Arrays

With the personalisation of medicine, made possible by our knowledge of the human genome and our growing, if still limited (Ashley et al. 2010), understanding of its influence on pathological conditions, molecular diagnostics are taking an increasingly prominent role. The size of the molecular diagnostic market

worldwide was estimated at 3–4 billion USD in 2010 (Medical News Today 2008), and it is expected to grow at an annual rate of 20%. While this figure is dwarfed by those of medical drugs, where cancer medicines alone account for 50 billion USD (Chapman 2010), diagnostics are nevertheless an increasingly essential part of medical care and are attractive from a business point of view since development schedules, regulatory issues and costs are generally more favourable than for drugs. Apart from clinical diagnostics *stricto sensu*, the drug development process relies more and more on various types of biomarkers to provide early assessment of toxicity or off-target effects of potential drugs and, increasingly, on selection for clinical trials of those patients more likely to respond to the molecule being tested by virtue of their particular genetic makeup. DNA arrays, in their various implementations, obviously have an important role to play in these developments.

Indeed, as soon as they appeared, DNA arrays were hailed as potentially revolutionary diagnostic devices. Schena et al. (1995) stated that "microarrays of cDNAs could thus provide a useful link between human gene sequences and clinical medicine". In the early 2000s, clinical microarrays were widely expected to represent a multibillion (USD) market in the near future (Moser 2001). DNA arrays have indeed been adapted to clinical use and, in some cases, have obtained FDA clearance in the USA: the most widely publicised examples are the CYP450 AmpliChip, developed by Affymetrix and Roche and aimed at characterising the genes for two enzymes involved in drug metabolism (Jain 2005), and the MammaPrint microarray test, proposed by the Dutch company Agendia and designed to analyse expression profiles in resected breast tumour material to indicate the likelihood of tumour recurrence (Glas et al. 2006). Both eventually obtained FDA clearance and were followed by several other array-based tests.

Nevertheless, the impact of microarrays in molecular diagnostics has remained relatively limited (Li et al. 2008; Jordan 2010): their market is currently estimated to be 300–400 million USD, i.e. less than 10% of the total molecular diagnostics market. In many discussions emphasising the rosy perspectives of clinical DNA arrays, important requisites for a diagnostic test were lost from sight, and the differences between the research environment and the clinical laboratory were sometimes overlooked. A diagnostic test must be very robust technically, even when performed under suboptimal conditions and by moderately qualified personnel. It must also provide a clear-cut result, a clear "yes", a definite "no" or a well-defined numerical score on which therapeutic decisions may be based. Regulatory approval is also essential. While this is a relatively minor issue for research equipment and products, it becomes an absolute requirement for a clinical test that aims to be eventually accepted for reimbursement by public or private health maintenance organisations. Various agencies and regulations are involved, and they may heavily influence technological choices, usually in favour of methods that are considered "established" (see Chap. 12).

The most important necessity, however, is real-world clinical utility: the test must provide solid input into therapeutic decisions, and its use must be shown to improve the outcome for the patient (Børresen-Dale 2003; Simon 2008). To take a widely discussed example, a major issue in cancer management is the avoidance of

unnecessary chemotherapy. Only a fraction of tumours turn out to have metastatic potential, and only a subset of those respond to treatment. A variety of clinical indexes based on tumour size, histochemical examination and clinical presentation have been developed, but the information they provide is insufficient, and as a result, many patients are subjected to an unpleasant, dangerous and expensive therapy that is actually useful only for a fraction of them. Expression profiling performed using a fragment of tumour tissue can provide indications on the aggressiveness of the cancer and thus help orient treatment but will be of real use only if the result allows, for example, definite identification of a class of patients whose risk is so low that additional therapy can be safely withheld. Otherwise, the test will neither be widely performed nor reimbursed by insurance companies, an essential issue given that its cost is usually quite high.[3]

2.4 Expression Arrays Versus Genotyping Arrays

The distinction formulated in the subtitle above is essential, for both technical and for regulatory reasons (see Table 2.2). *Expression arrays*, possibly the most familiar use of DNA arrays and historically the initial one, measure the expression level of a set of genes in a tumour, tissue or blood sample from a patient. *Genotyping DNA arrays,* on the other hand, are aimed at characterising a DNA (or, in some cases, RNA) sequence present in the sample. They may be designed for bacterial or viral identification or, in the case of human DNA, to find out if specific mutations are present, to define alleles for a number of sites where single nucleotide polymorphisms (snips) have been reported or to characterise copy number variation.

These two applications are technically quite dissimilar; they involve different sample handling procedures and entail diverse requirements on storage conditions since mRNA is much more labile than DNA. However, the most important difference between them lies in the solidity of clinical correlates. To use expression profiling in the clinic, the company introducing the test relies on new studies (performed in-house or with academic collaborators) showing that a given expression profile is prognostic (and possibly predictive[4] as well). While these studies have usually been published in reputable journals, they may not be generally accepted, and even if they are, the issue of clinical utility remains: it is up to the company (and the collaborating academic group) to show that using the test will measurably improve the outcome for patients. Indeed, for some of the major expression tests currently on the market, large clinical studies are still under way, and their final results will not be available for several years. Under these conditions, obtaining FDA clearance for clinical use, gaining physician acceptance and achieving

[3] The breast cancer prognostic tests discussed in Chap. 3 all cost at least €2,000, worthwhile only if their use can lead to avoidance of unnecessary chemotherapy.

[4] For the precise meaning of "prognostic" and "predictive", see Chaps. 3 and 11.