Fundamentals of DNA-chip/array technology for comparative gene-expression analysis

Hans Peter Saluz*, Javeed Iqbal, Gino V. Limmon, Andre Ruryk and Zhihao Wu

Department of Cell and Molecular Biology, Hans Knoell Institute for Natural Product Research, Beutenbergstrasse 11, D-07745 Jena, Germany

In this article we review the state-of-the-art of DNA-chip/array technology and its application in biological and medical sciences. With many of the technological hurdles to overcome array fabrication, many academic groups and numerous specialized companies are developing new tools and strategies for exploiting this promising field. DNA-microarrays are fabricated either by *in situ* synthesis or by conventional means, e.g. amplification of specific target DNA, followed by immobilization on the surface of various miniaturized substrates. The immobilized DNA molecules are hybridized with labelled probes. One single experiment can encompass aspects of many thousands of genes simultaneously. The unprecedented speed and parallelism of this technology has begun to have profound impacts on biological, environmental, pharmacological and medical research.

IN the past decade, several eukaryotic and prokaryotic genomes have been analysed. The greatest effort concerned the analysis of the human genome through the Human Genome Project. The gigantic amount of data which resulted from these endeavours provide a completely new insight into the genetic information. Comparisons with sequences from databases allowed functions of a part of isolated genes to be predicted. However, to complete the search for pieces of the genomic puzzle and to gain far more detailed functional information concerning gene expression and protein analysis, a set of highly advanced technologies had to be developed. One of the most fascinating outcome is the chip technology. Besides its scientific potential, it opens a broad variety of novel markets with an enormous developmental potential for almost all key indications, e.g. early diagnosis, diagnostics and therapy monitoring of infectious and cancer diseases, prenatal diagnostics of inherited diseases, predisposition diagnostics or therapy monitoring of novel therapeutic principles, especially in gene therapy.

This article highlights recent technical developments concerning gene-expression analysis by means of microarray technologies and discoveries arising from the use thereof. In addition, it also discusses bottlenecks surrounding gene-expression analysis. Although a broad variety of approaches are available for gene-expression analysis, we focus on DNA arrays because they offer the high-throughput capacity and cost-effectiveness crucial for large-scale analysis of genomic data. DNA probes from hundreds to thousands of genes are positioned and immobilized on an active surface of a substrate. Hybridization reactions between the array and a cell/tissue sample indicate that certain genes represented on the array are active in the sample. This allows the elucidation of patterns of gene expression that may be associated with states of health and disease.

DNA-microarray technologies

The field of DNA-microarray technology has evolved from the Southern-blot analysis, i.e. Ed Southern's technology of detecting specific sequences among DNA fragments separated by gel electrophoresis[1]. Using this method, the Southern blot was the first array. The next historical steps were filter-based screening of clone libraries and then gridded libraries stored in microtiter plates and immobilized on filter membranes. Finally, two key innovations led to an explosion of interest in array technology. The first was pioneered by Pat Brown and colleagues[2,3]. It concerns the use of non-porous solid bodies, such as glass slides, on the activated surface on which the molecules of interest are positioned and immobilized. The immobilized molecules are reacted with labelled probe molecules. Labelling is usually performed by fluorescent dyes. Using this method, commonly up to 10,000 molecule groups per slide are robotically spotted. The great advantage of such procedures is the enormous flexibility, i.e. the array can be produced in most molecular-biology laboratories without involving dust-free rooms, thus allowing array arrangements according to the individual biological aims. In addition, their production is relatively cheap. Another interesting aspect, at least for technically advanced laboratories, concerns a variety of substrate-surface treatments, e.g. surfaces with hydrophilic and lipophilic areas[4] (Figure 1).

*For correspondence. (e-mail: saluz@pmail.hki-jena.de)

The second key innovation was the development of methods for a high-density *in situ* synthesis of oligonucleotides on support surfaces. The first procedure concerning *in situ* synthesis was published in 1994 by Jacobs and colleagues[5] from Affymetrix. They used masks – as commonly applied in photolithography – for the preparation of arrays. Molecules positioned on defined islands on the solid support were deprotected and coupled with the next activated monomer. By this means it was possible to construct thousands of all possible variations of polymers *in situ*. At that time, all possible variations of an oligonucleotide octamer on one and the same support, i.e. 65,536 different molecular species[6] were synthesized. Also, peptide libraries were constructed using the above approach[7]. In the same year an alternative approach was described by Southern and Maskos[8]. It was successfully applied for the synthesis of large numbers of different oligonucleotides on solid supports. They applied the conventional solid-phase synthesis chemistry for the preparation of the oligonucleotides. In 1996 an approach combining standard phosphoramidite chemistry and standard photolithographic resist technology was developed by McGall and colleagues[9]. In this approach, in analogy to techniques common to semiconductor manufacturing, the support is covered with a photosensitive resist. This resist is patterned by light exposure through a suitable photolithographic mask and serves as a physical barrier to prevent selected areas from exposure to chemical reagents in the course of the synthesis cycle. Due to their nonlinear response to illumination intensity, photoresists allow the generation of patterned surfaces with submicron features[4]. However, all the *in situ* strategies have some drawbacks. The strategies concerning light-sensitive deprotection involve an expensive chemistry. In addition, the light-sensitive protection procedure is still problematic, i.e. only relatively short molecules can be synthesized. Furthermore, special cost-intense clean rooms are required. For the above reasons the standard amidite chemistry is a desirable feature. Therefore, the major advantage of the procedure of Southern and Maskos[1] is the use of conventional synthesis reagents. However, due to its capillary system, where all capillaries have to be filled with the appropriate reagent for the synthesis, this technology is not very economic. In addition, the tiny capillaries are sensitive and therefore, often blocked by particles. For these reasons more efficient and economic procedures had to be developed. One procedure[10] which can be applied in all laboratories and which can be used to synthesize all variations of a given polymer by conventional means, involves chemically inert stamps[11]. In Figure 2, a PDMS stamp with 64,000 islands per cm^2 is shown. The active surface of these stamps is microstructured according to defined synthesis algorithms (Figure 3).

Producing chips involving light-directed, spatially addressable, parallel chemical synthesis, it was possible to build up arrays with as many as 400,000 distinct oligonucleotides, each on a surface area of 20 μm^2 (ref. 12).

Gene expression studies

Most current array-based investigations concern the monitoring of RNA expression levels. The tools are powerful for both prokaryotic and eukaryotic genomes. Regarding all worldwide efforts, it is therefore not surprising that gene-expression analysis technologies – especially DNA arrays – are already producing real, valuable results in ground-breaking research studies, particularly in cancer and other diseases. However, concerning one major bottleneck of the above technology, namely the software for the interpretation of the data, much work has still to be done.

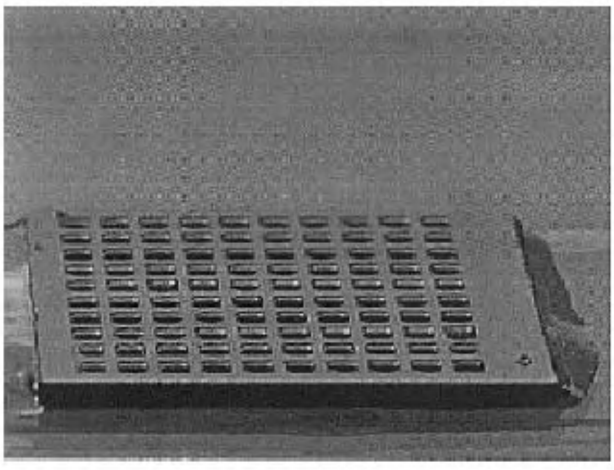

Figure 1. 10 × 10 µm UNIPAD immobilization matrix with hydrophilic and lipophilic areas. Drops of water remain on the hydrophilic spots; chip 1 cm × 1 cm.

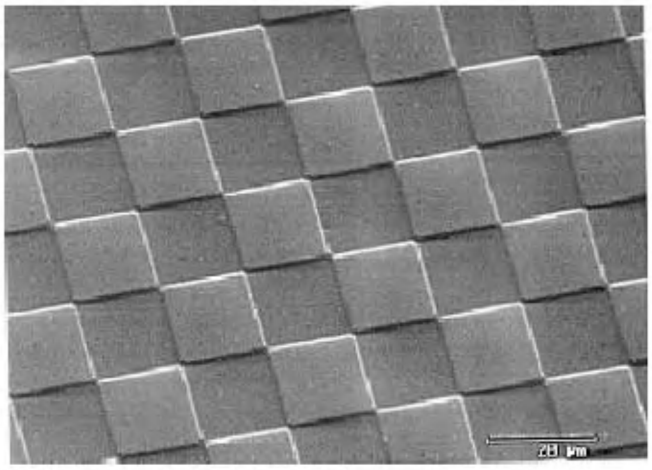

Figure 2. Section of a structured stamp-surface area. One cm^2 of the surface of this PDMS stamp contains 64,000 islands.

The impact of the DNA chips on cancer research has been already influenced by a number of investigations published in the last couple of years[13–19]. The role of DNA chips in expression monitoring is to specifically point out the targets of the causative agents in host response and the specific role of genes of the pathogen itself, e.g. role of *HBV* and *HCV* in liver cancer, *EBV* in nasopharyngeal cancer and specific lymphomas, *HPV* in cervical cancers and helicobacter pylori in gastric cancers constituting a category of cancer-risk factors, besides the role of certain other microbial agents, such as *Chlamydia* or *Plasmodia*. Such events can be specifically studied by serial analysis of infections at different time points and knowing the influence of the various genes on the pathogen involved. Such studies have been performed to identify novel transcripts involved in tumour suppression in the human melanoma cell line UACC-903, and the suppression by introduction of normal human chromosome 6 (ref. 20). The same application has been utilized to identify the preferentially expressed mRNA molecules in pancreatic cancer cells[21].

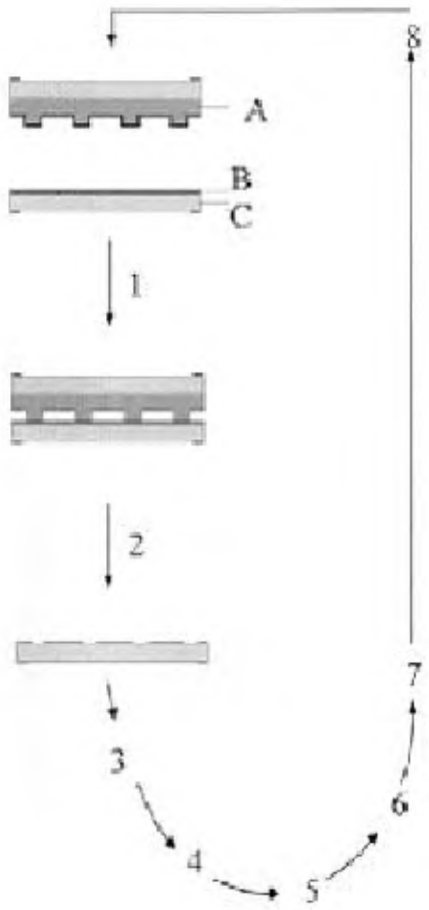

Figure 3. Schematic representation of a synthesis cycle. A, Elastomeric stamp which is covered with the deprotecting agent; B, Protecting group, and C, Matrix. Numbers 1–8 indicate the individual steps of a synthesis cycle: 1, Stamping; 2, Stamped areas deprotected; 3, Coupling; 4, Washing; 5, Oxidation; 6, Washing; 7, Capping and 8, Washing.

Moreover, the use of other pioneering bimolecular techniques, such as suppressive subtractive hybridization or representational difference analysis (RDA)[22], in conjunction with DNA chips has been utilized to identify the rarely transcribed differentially expressed genes in human

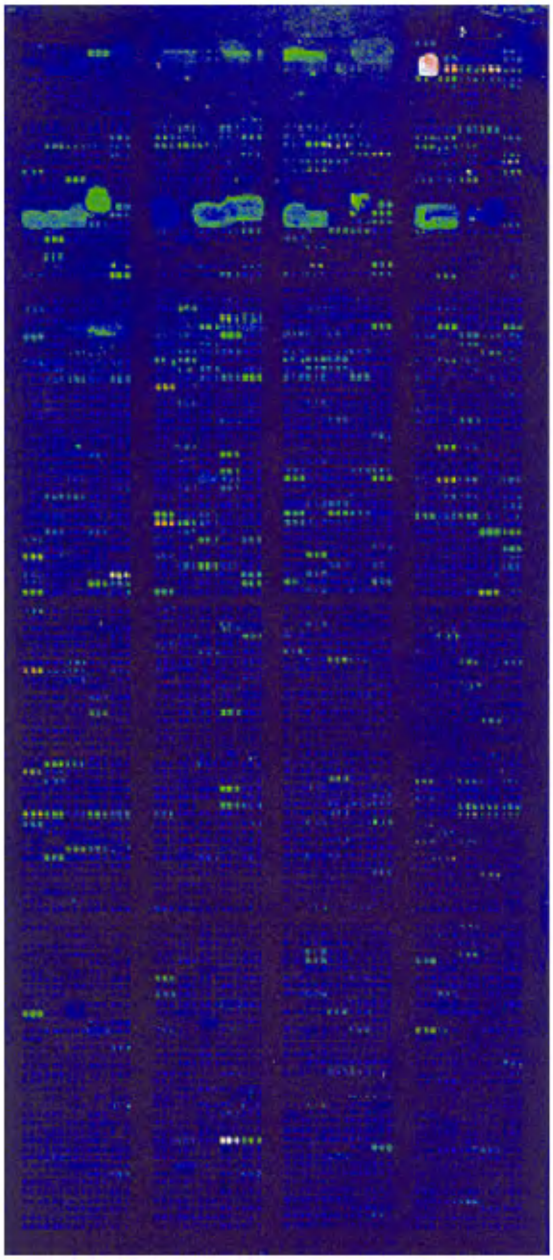

Figure 4. *In vivo* recognition-element library combined with array technology. Known proteins with a high affinity to DNA, such as transcription factors, are covalently immobilized in living cells by means of nanosecond-UV irradiation. The nucleo-protein complexes are specifically isolated. By a special procedure the footprints are determined (Wu *et al.*, in publication). Upon digestion of the proteins, the promoter areas containing gene sequences are positioned and immobilized on chip surfaces. The immobilized DNA is hybridized with labelled mRNA from cells or tissues, e.g. from patients or from differently treated biological samples. By this means, differential gene expression in the context of specific protein/DNA-interactions can be investigated, revealing novel targets.

cell lines[23–25]. In another study by Javed and colleagues[26] involving microarrays, the gene expression induced by the chimerical transcription factor *PAX3-FKHR* in NIH3T3 cells and the myogenic properties of these oncogenic transcription factors in *Alveolar rhabdomyosarcoma* were elucidated. Several recent and ongoing cancer studies using microarrays led to the discovery of promising cancer targets. In a pioneering study published last year, researchers led by Eric Lander and Todd Golub[27] generated cancer relevant gene-expression profiles. The consequence of their results was the power of discrimination between acute myeloid leukaemia and acute lymphoblastic leukaemia, the two types of leukaemia which were extremely difficult to be specifically analysed. Other important recent investigations based on gene-expression monitoring revealed new subtypes of lymphoma[28]. Patrick Brown *et al.*[29] identified by the above means, two classes of breast tumours, in only one of which the gene for the estrogen receptor was expressed. These two types of breast cancer were already known; however, the same studies indicated that these two tumour groups have further subgroups. In another study it was found that melanoma patients could be classified into two subgroups according to gene-expression profiles[20]. These subgroups were not distinguishable pathologically.

One of the main challenges that faces current molecular biology is the integration of new genetic information that is pooled in the diagnostic research. Microarrays provide the backbone for such procedures without sacrificing the sensitivity and miniaturization of samples which leads to reduction in material, cost and preparation time. Parallelization of this method has thus greater advantage with DNA diagnostics than conventional techniques.

The first emphasis will be on the differential sequencing to identify mutations and polymorphism of a known target sequence. Such an approach was utilized by Chee and colleagues[30] to analyse the entire human mitochondrial genome, using an oligonucleotide array of 135,000 features to investigate 179 of totally 180 polymorphisms with a precision of 99%.

Another feature in the genome era will be the elucidation of the functions of tens of thousands of EST (> 400,000) sequences of the genes. The chips-based approach is a sensitive method to accomplish such tasks. Schena and his group[31] used the cDNA-microarray technology to identify the role of novel heat-shock and phorbol-ester, regulated genes in human T-cells. This was the first work to monitor differential expression of the human genome. They were able to discover four yet unknown genes, hence implementing that novel genes can be found without the prior knowledge of the corresponding sequence information. In a similar study the same group showed the involvement of the selected molecules in rheumatoid arthritis and inflammatory bowel diseases[32]. The study revealed the novel participation of cytokine interleukine 3, chemokine Groα and the metalloproteinase matrix mettaelastase in both diseases.

The basic aspect in using diagnostic chips will be to implement the fundamental strategies in homogenous linear amplification of specific labelled probes. Although physical (absolute) quantification of signals arising from the fluorescence labelled hybrids on a chip is still not possible and will be one of the main challenges to be exposed for knowing the precise viral or pathogenic loads in patient samples. These data will indicate, e.g. the state of progression of a given disease at a certain time point. Although this is one of the major challenging phases of the diagnostic chips and the present chip infrastructure which is still in infancy, but using accessible tools in clinics for patient management and diagnostics will certainly be a breakthrough. Accessible chip diagnostics should be cheap, simple and very easy to interpret for large scale screening.

One of our major objectives has been the identification of the differentially expressed genes in various cell lines and/or tissues upon bacterial or viral infections. These experiments have been performed by cDNA microarrays. Using this method several hundred novel potential drug targets out of approximately 40,000 cDNA clones could be found. The precise functions of these molecules are currently under investigation (Iqbal *et al.*, unpublished data). Another miniaturized approach concerns *in vivo* recognition element libraries[33]. This technology has been modified and drastically improved to study both genome-wide interaction *in vivo* of protein factors with their promoters, and the simultaneous differential expression of the corresponding genes (Wu *et al.*, unpublished data). An example of such a recognition-element library combined with array technology is shown in Figure 4.

In search for drug targets in microbial genomes, it is feasible to use microarrays for the comparative analysis of the complete sequences of the bacterial or viral genomes. The genes that are conserved in the different genomes are often said to be essential for the pathogenicity. These well-defined pathogenicity islands can be attractive targets for a new, broad-spectrum antibiotics. Species-specific genes on the other hand can be used to design drugs against particularly narrow groups of pathogens[18]. The metabolic pathways that have been elucidated by microarrays in yeast under diauxic shifts can be implemented in various pathogenic organisms to design drugs that block the key enzymes of the pathogen[34]. A comparison of the expression profiles of the virulent pathogen and the non-virulent strains will lead to key fundamental genes that are responsible for pathogenicity.

Similarly, Vishwanath Iyer and colleagues[35] were able to demonstrate a temporal programme in gene expression of the response of fibroblasts to serum. The study suggested the relation of the transcriptional programme with a given physiological state of the cells.

Conclusion

Comparative expression profiling is one of the remarkable aspects about DNA chips, and now everyone is finding new ways to make them. From the results so far, it seems that DNA-chip technology will become as powerful as PCR within the next few years. A clear trend of transition has started in expression-profiling studies. One single publication encompasses relevant aspects of thousands of genes and here lies the hallmark transition of structural genomics to functional genomics.

The real challenge in expression arrays is in developing the experimental design to exploit the full power of global-perspective experimental manipulation, like responses to the microenvironment and state of the arrays, hybridization time, scanning procedures and other related aspects which need to be rigorously controlled. Another challenge that will be of concern is the study of the expression levels of very small quantities of target tissues. One should keep in mind while interpreting the data from DNA chips that transcription levels cannot be equalled to protein abundance or the rate of transcription is dependent on the half-life and decay direction of the mRNA. Also the signal strength does not reflect the level of potential translating mRNA molecules, as the protein can be regulated post-translationally. Although the problem is a pleasant one, it can be improved and it will provide us insight into the cellular function and allow us to come closer to the art of the network functions in the cell.

1. Southern, E. M., *J. Mol. Biol.*, 1975, **98**, 503–517.
2. Lockhart, D. J. *et al.*, *Nature Biotechnol.*, 1996, **14**, 1675–1680.
3. Shalon, D., Smith, S. J. and Brown, P. O., *Genome Res.*, 1996, **6**, 639–645.
4. Ermantraut, J., Wölfl, S. and Saluz, H. P., in *Microsystem Technology: A Powerful Tool for Biomolecular Studies* (eds Koehler, J. M., Mejevaia, T. and Saluz, H. P.), Birkhäuser/Springer-Verlag, Basel, 1999, pp. 255–265.
5. Jacobs, J. W. and Fodor, S. P. A., *Trends Biotechnol.*, 1994, **12**, 19–26.
6. Pease, A. C., Solas, D., Sullivan, E. J., Cronin, M. T., Holm, C. T. and Fodor, S. P. A., *Proc. Natl. Acad. Sci. USA*, 1994, **91**, 5022–5026.
7. Gruber, S. M., Yu-Pang, P. and Fodor, S. P. A., Proc. 12th Am. Pep. Symp., 1992.
8. Southern, E. M. and Maskos, U., *J. Biotechnol.*, 1994, **35**, 217–227.
9. McGall, G., Labadie, J., Brock, P., Wallraff, G., Nguen, T. and Hinsberg, W., *Proc. Natl. Acad. Sci. USA*, 1996, **93**, 13555–13560.
10. Ermantraut, E., Wölfl, S. and Saluz, H. P., Patent No. DE 195.43. 232.A1.
11. Kumar, A., Biebuyck, H. A. and Whitesides, G. M., *Langmuir*, 1994, **10**, 1498–1511.
12. Fodor, S. P. A. *et al.*, *Science*, 1991, **251**, 767–773.
13. Pastinen, T., Kurg, A., Metspalu, A., Peltonen, L. and Syvänen, A-C., *Genome Res.*, 1997, **7**, 606–614.
14. Guo, Z., Guilfoyle, R. A., Thiel, A. J., Wang, R. and Smith, L. M., *Nucleic Acids Res.*, 1994, **22**, 5456–5465.
15. Boguski, M. S., *Trends Biochem. Sci.*, 1995, **20**, 295–296.
16. Boguski, M. S. and Schuler, G. D., *Nature Genet.*, 1995, **10**, 369–371.
17. Carulli, J. P. *et al.*, *J. Cell. Biochem. Suppl.*, 1998, **30/31**, 286–296.
18. DeRisi, J., Iyer, V. R. and Brown, P. O., *Science*, 1997, **278**, 680–686.
19. Bowtell, D. D. L., *Nature Genet. Suppl.*, 1999, **21**, 25–32.
20. DeRisi, J., Penland, L. and Brown, P. O., *Nature Genet.*, 1996, **14**, 457–460.
21. Gress, T. M. *et al.*, *Oncogene*, 1996, **13**, 1819–1830.
22. Lisitsyn, N. and Wigler, M., *Science*, 1993, **259**, 946–951.
23. Yang George, P., Ross Douglas, T., Kuang Wayne, W., Brown Patrick, O. and Weigel Ronald, J., *Nucleic Acids Res.*, 1999, **27**.
24. Welford, S. M., Gregg, J., Chen, E., Garrison, D., Sorensen, P. H., Denny, Ch. T. and Nelson, S. F., *ibid*, 1998, **26**, 3059–3065.
25. Loftus, S. K. *et al.*, *Proc. Natl. Acad. Sci. USA*, 1999, **96**, 9277–9280.
26. Khan, J. *et al.*, *ibid*, 1999, **96**, 13264–13269.
27. Lander, E. S. *et al.*, *Science*, 1999, **286**, 531–537.
28. Ross, D. T. *et al.*, *Nature Genet.*, 2000, **24**, 227–235.
29. Brown, P. O. *et al.*, *Nature*, 2000, **406**, 747–752.
30. Chee, M. *et al.*, *Science*, 1996, **274**, 610–614.
31. Schena, M. *et al.*, *Proc. Natl. Acad. Sci. USA*, 1997, **94**, 2150–2155.
32. Schena, M., Shalon, D., Heller, R., Chai, A., Brown, P. O. and Davis, R. W., *ibid*, 1996, **93**, 10614–10619.
33. Santoro, Wölfl, S. and Saluz, H. P., *BBRC*, 1999, **256**, 68–74.
34. Debouck Christine and Goodfellow Peter, N., *Nature Genet.*, 1999, **21**, 48–50.
35. Iyer, V. R. *et al.*, *Science*, 1999, **283**, 83–87.