Opinion
# Have microarrays failed to deliver for developmental biology?
Rick Livesey

Address: Wellcome Trust/Cancer Research UK Institute of Cancer and Developmental Biology and Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QR, UK. E-mail: rick@welc.cam.ac.uk

## Abstract

Comprehensive microarrays covering large numbers of the predicted expressed transcripts for some invertebrates and vertebrates have been available for some time. Despite predictions that this technology will transform biology, to date there have been few published studies using microarrays to generate novel insights in developmental biology.

The cDNA microarray is a conceptually simple object, whether made on a glass slide within an academic lab or printed using complex technology in a commercial production setting. The best characterization of them is as glorified dot-blots (VG Cheung, personal communication), and this explains much of their appeal. Most life scientists have carried out a northern, Southern or *in situ* hybridization, and are therefore familiar with the main technology - hybridization - needed for microarray use. Yet it has proven difficult in some fields to translate the obvious promise of microarrays into tangible results. A pressing issue is why it appears that we are stuck in a 'proof of principle' stage, rather than a routine exploitation phase, especially in the field of developmental biology. The problem is that there have been few developmental biology microarray studies, indicating a very slow adoption of the technology in this field. Currently there appear to be two overriding concerns: access to the technology in a reliable form, and how best to apply the technology in a way that generates data that are useful in the short to medium term.

## Honourable exceptions

Interesting and insightful developmental studies using microarrays have been published, but they are all the more striking for their infrequency. There have been several studies of *Drosophila* development [1-3], including one identifying targets of the mesoderm-specific transcription factor *twist* [4]. Similarly, there have been a number of genome-wide studies of worm development, almost all from the Kim lab at Stanford [5-7]. The situation appears more bleak when looking for published microarray studies of vertebrate development. Aside from some small-scale studies of retinal development in the mouse (including our own) [8,9] and mesoderm induction in *Xenopus* [10], there are very few. The predicted deluge of data has failed to materialize, and it is not immediately obvious why this is the case. Commercial arrays of worm, fly, human and mouse genes have been available for several years, and several extensive cDNA clone sets are available for little or no cost. A cursory web search finds many academic microarray facilities offering mouse arrays, for example, yet little or no published data.

This is not the case when looking at other fields of research. The most striking successes are in classifying different cancers, including prognostic predictions (for examples, see [11-13]), and arrays are in general use in labs studying both the clinical and basic aspects of cancer biology. Other impressive studies have included comparative studies of bacterial genomes [14], and a number of groundbreaking studies of transcription and genome organization in yeast [15-17]. Aside from these high-profile studies, there also have been many other studies completed and published in these fields, all adding to the evidence of the widespread adoption of microarray technology by communities of researchers. The power and usefulness of microarrays are not in question, so why do there appear to be problems applying them to developmental biology?

## The perfect partnership?

Dynamic or temporally and spatially restricted patterns of gene expression are recurring mechanisms in the regulation of developmental processes. Where and when certain key sets of genes are expressed regulates processes such as responses to growth factors, cell-fate determination and differentiation. Many transcription factors and signaling pathways that are required for developmental processes have been identified, their functions carefully studied and organisms carrying targeted mutations in the corresponding genes generated. Technologies for generating targeted mutations in genes of interest, at particular times and in particular tissues, have been commonly used for many years in developmental biology. Such precise mutants are typically used to confirm predictions of gene function within a biological or cellular process, and are rarely used as tools for understanding the gene expression network within which a particular gene operates. It would appear therefore, with the addition of microarrays for gene-expression profiling, that all of the tools for complex dissections of both cellular and genetic pathways are available to developmental biologists.

One can picture a category of investigation that is likely to lead to a cycle of using arrays to analyze animals carrying targeted mutations to identify novel components of the pathway within which the gene product operates; the novel components are, in their turn, mutated and analyzed using arrays. The resulting complex datasets can then to be used to reconstruct the gene-expression networks operating in the relevant cells to regulate the processes under investigation. On a larger scale, and in the longer term, one goal will be to integrate data on gene expression, protein expression and cellular metabolism, so as to generate models of cellular behavior. Such models will allow us to make predictions about the changes resulting from interfering in a gene-expression network, for example by conditional gene knock-out, RNA interference (RNAi) [18,19], or overexpression, the results of which we can study using RNA and protein expression arrays. As similar studies are carried out on different genes and mutants in a given tissue in a high-throughput manner, it will be possible to model the development of that tissue at the genetic network level, reaching one of the long-term goals of developmental biology.

## Scale and resolution

Important technical issues about using microarrays for developmental biology are the small size of available tissues and the correspondingly small amounts of total RNA, and also the resolution at which expression data can be generated. The majority of array studies use microgram amounts of total RNA for each hybridization [20], quantities that can be difficult to achieve from developing tissues. The simplest solution to the RNA availability problem is to collect very large numbers of samples, so as to generate enough RNA for an array probe, for example using *Drosophila* embryo

sorters [21]. This may not always be practically possible or scientifically desirable. Alternatively, there are several different technologies available for RNA/cDNA amplification from limiting amounts of RNA (for review see [22]). All of these methods are, however, subject to some concerns about the fidelity of representation of the original complex RNA population in the resulting amplified material.

The second issue is the level of resolution at which expression studies can be carried out. Ideally, developmental biologists wish to study gene expression in populations of identical cells, rather than in a whole tissue or a mixed population of cells (Figure 1). If sufficient starting material and antibodies to cell-specific surface antigens are available, one approach is to purify homogenous population of cells by fluorescence-activated cell sorting (FACS). But FACS sorting of cell populations is unlikely to produce large enough numbers of cells to generate sufficient RNA for a traditional microarray probe, particularly from developing tissues. RNA extracted from purified cells is therefore likely to require amplification. Alternatively, methods have been available for some time for cDNA synthesis and amplification from single cells [23-25], with several novel methods reported to be under development. This is an attractive alternative to cell sorting, given that populations generated from FACS sorting are unlikely to be comprised of 100% pure, identical cells. Single-cell cDNA preparations are beginning to be successfully analyzed using microarrays, raising the possibility that it will soon be possible to reliably study gene expression at the single-cell level during development.

Finally, there is the question of whether there are particular technical problems that could confound the use of microarrays in developmental biology. These would include the ability to detect functionally important transcripts expressed at low levels, or a lack of availability of arrays that distinguish between functionally significant splice variants. There is no general principle that interesting genes are expressed at low levels, but in any case detection of low-abundance transcripts is not generally considered an issue, particularly with amplification of small amounts of starting material. Similarly, the use of splice variants in developmental settings, while important, is not the only, or even the primary, mechanism used for generating functional differences between cells. In both of these cases, the current situation is that we cannot know if any speculations are valid, given the small numbers of microarray studies that have been carried out in developmental biology to date. It seems unlikely *a priori* that there are any issues specific to developmental biology that will prevent expression profiling from being applicable to this field.

## Genetics versus genomics

Aside from technical concerns, there may be a lack of attraction for developmental biologists to arrays, and to genomic
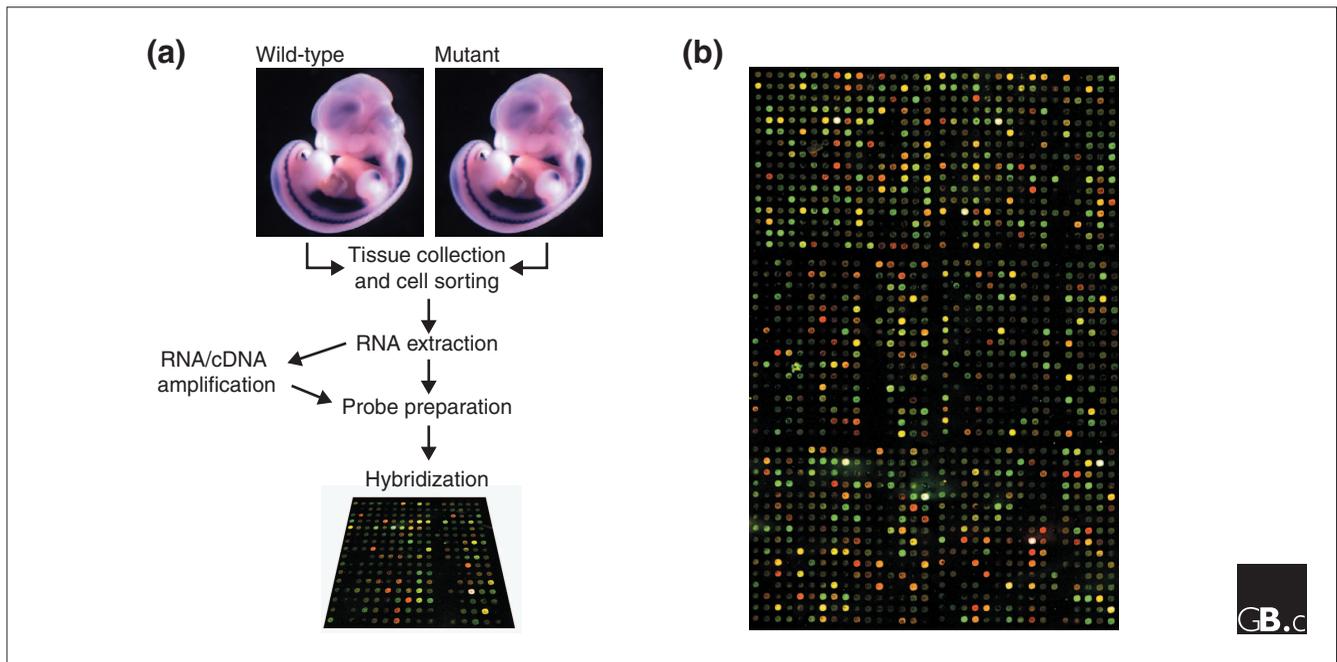
**Figure 1**
Microarrays of gene expression in mouse embryos. **(a)** A typical single microarray comparison, using mouse embryos as an example, illustrating some of the issues that arise concerning the use of microarrays to study development. The whole-mount *in situ* hybridization shown in the top panels demonstrates the complexity of gene expression during development, with tissue-specific expression in a small number of areas. Studying the consequences of altering normal gene expression requires dissection of the tissues that gene is expressed in, or sorting of the specific cell populations of interest. A consequence of this is that the extracted RNA will probably need to be amplified to generate sufficient cDNA for a microarray probe. **(b)** Detail of a microarray hybridization comparing gene expression in a single mouse neural progenitor cell (red) with that in total brain (green). One issue for developmental biologists is whether arrays can be used to profile gene expression in small numbers of cells. We and others have found that it is possible to reproducibly profile gene expression from single cells, using current cDNA amplification technologies.

technologies in general. In this regard, there is a clear distinction between newer genome-based technologies and classical genetics, which has proven a very fruitful approach for understanding development. Developmental biology, historically an experimental science, has traditionally been function-led. A common view is that it is better to carry out a well-designed genetic screen to identify a small group of genes that when mutated give clear phenotypes related to the process being studied, rather than to identify hundreds of transcripts whose expressions correlate with key aspects of the same process, followed by functional studies of those deemed most promising. But this view is based on a misunderstanding of how complementary these approaches are, as discussed above, in extracting even more information from available mutants. Furthermore, it is a view that characterizes microarrays simply as gene-discovery tools. Although arrays are very useful in that role, one of the benefits of generating expression data from large numbers of genes simultaneously is that the entire dataset, taken together, contains useful information on transcription within cells or tissues that can be used for modeling gene-expression networks.

More importantly, there are organisms, developmental stages and processes for which genetic screens cannot be performed,

or are extremely expensive and difficult to carry out. This is particularly true of the later stages of vertebrate development, most notably of the mammalian central nervous system. In such cases, there has to come a point where investigating the conservation of the fundamental principles of development identified in powerful invertebrate systems must be left behind, and directed efforts must be made to identify the developmental mechanisms underlying vertebrate- and mammal-specific structures. This is where arrays can complement genetic approaches, to identify candidate genes or networks involved in development of those structures. In addition, arrays are becoming useful tools for identifying candidate genes underlying complex traits in both vertebrates and invertebrates [26], further complementing genetic approaches.

One final possible reason for the slow adoption of array technologies in developmental biology may be the lack of availability of arrays for many model organisms used in developmental biology, including *Xenopus* and zebrafish. Organism- and tissue-specific arrays are straightforward to construct, as we and others have done for organisms such as chick and ferret, whose genomes have not been sequenced or are not supported by large EST projects. The simplest way to

do this is to make arrays of random, unsequenced clones from cDNA libraries from particular tissues or developmental stages, then use the arrays as normal and only sequence those clones that prove to be of interest after the data-analysis step. The alternative is to sequence sets of clones initially and then array a non-redundant subset of those clones. In either case, such arrays can be produced quickly and relatively cheaply within individual labs that have access to liquid-handling systems and arrayers, and arrays are now becoming generally available (see, for example [10]).

## The informatics challenge
Much has been written and said about the interpretation of array data, the analytical challenge presented by large datasets, and how this should intimidate unwary biological scientists [27]. This may have delayed the entry of many researchers into applying array technology to their own research, by raising concerns that, even if individual labs do generate good gene-expression datasets, they will be unable to extract useful information from them. Thanks to the efforts of many public-spirited labs around the world, however, all of the necessary computing tools are now freely available to academic users. These include database systems for storing and accessing raw data and desktop software for analyzing data extracted from the databases to identify interesting features. These analytical tools include the now-standard cluster analysis algorithms, along with other novel statistical techniques, and are remarkably straightforward to use. In addition, there are now a number of training courses offered by genome centres on the theoretical and applied aspects of expression data analysis, so labs wishing to use these technologies can become proficient very quickly (see, for example, [28]).

In summary, developmental biologists have been relatively slow to adopt what is, in many ways, an ideal technology for answering some of the major questions in development. There is nothing fundamentally new about this technology that means that normal rules will not apply. As for every other new technology, the proof-of-principle stage will be followed by the widespread-adoption stage, accompanied or followed by a sharp improvement in the standard and the complexity of studies using these technologies. But it is reasonable to ask why this process is taking so long. The first microarray paper from Patrick Brown's lab appeared in 1995 [29], seven years ago. There was a frenzy of interest in arrays by four or more years ago, when any seminar that mentioned array technologies was full to capacity. There are likely to be many reasons why this was not followed by a large number of studies in developmental biology using array technologies. Early on, limited access to reagents was a definite factor slowing widespread array use. Another factor has been the intellectual shift away from single-gene studies to thinking about gene-expression networks, allied to a dependence on statistics, mathematics and computing. A reluctance to embrace these newer approaches is likely to be a generational issue, as there is now a cohort of students and post-doctoral fellows that have been trained in an environment where such quantitative methods are the norm. Given the wide availability of arrays and the necessary computing infrastructure to interpret the data, developmental biology should be able to make increasing and impressive use of these approaches in the near future, despite a relatively slow start.

## References
1. Brenman JE, Gao FB, Jan LY, Jan YN: **Sequoia, a tramtrack-related zinc finger protein, functions as a pan-neural regulator for dendrite and axon morphogenesis in *Drosophila*.** *Dev Cell* 2001, **1:**667-677.
2. White KP, Rifkin SA, Hurban P, Hogness DS: **Microarray analysis of *Drosophila* development during metamorphosis.** *Science* 1999, **286:**2179-8214.
3. Bryant Z, Subrahmanyan L, Tworoger M, LaTray L, Liu CR, Li MJ, van den Engh G, Ruohola-Baker H: **Characterization of differentially expressed genes in purified *Drosophila* follicle cells: toward a general strategy for cell type-specific developmental analysis.** *Proc Natl Acad Sci USA* 1999, **96:**5559-5564.
4. Furlong EE, Andersen EC, Null B, White KP, Scott MP: **Patterns of gene expression during *Drosophila* mesoderm development.** *Science* 2001, **293:**1629-1633.
5. Jiang M, Ryu J, Kiraly M, Duke K, Reinke V, Kim SK: **Genome-wide analysis of developmental and sex-regulated gene expression profiles in *Caenorhabditis elegans*.** *Proc Natl Acad Sci USA* 2001, **98:**218-223.
6. Reinke V, Smith HE, Nance J, Wang J, Van Doren C, Begley R, Jones SJ, Davis EB, Scherer S, Ward S, *et al.*: **A global profile of germline gene expression in C. elegans.** *Mol Cell* 2000, **6:**605-616.
7. Kim SK, Lund J, Kiraly M, Duke K, Jiang M, Stuart JM, Eizinger A, Wylie BN, Davidson, GS: **A gene expression map for *Caenorhabditis elegans*.** *Science* 2001, **293:**2087-2092.
8. Livesey FJ, Furukawa T, Steffen MA, Church GM, Cepko CL: **Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene Crx.** *Curr Biol* 2000, **10:**301-310.
9. Mu X, Zhao S, Pershad R, Hsieh TF, Scarpa A, Wang SW, White RA, Beremand PD, Thomas TL, Gan L, *et al.*: **Gene expression in the developing mouse retina by EST sequencing and microarray analysis.** *Nucleic Acids Res* 2001, **29:**4983-4993.
10. Altmann CR, Bell E, Sczyrba A, Pun J, Bekiranov S, Gaasterland T, Brivanlou, AH: **Microarray-based analysis of early development in *Xenopus laevis*.** *Dev Biol* 2001, **236:**64-75.
11. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA *et al.*: **Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.** *Science* 1999, **286:**531-537.
12. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, *et al.*: **Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.** *Nature* 2000, **403:**503-511.
13. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al.*: **Molecular portraits of human breast tumours.** *Nature* 2000, **406:**747-752.
14. Behr MA, Wilson MA, Gill WP, Salamon H, Schoolnik GK, Rane S, Small PM: **Comparative genomics of BCG vaccines by whole-genome DNA microarray.** *Science* 1999, **284:**1520-1523.
15. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM: **Systematic determination of genetic network architecture.** *Nat Genet* 1999, **22:**281-285.
16. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B: **Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization.** *Mol Biol Cell* 1998, **9:**3273-3297.
17. Cohen BA, Mitra RD, Hughes JD, Church GM: **A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression.** *Nat Genet* 2000, **26:**183-186.

18. Fire A: **RNA-triggered gene silencing.** *Trends Genet* 1999, **15:**358-363.
19. Elbashir SM, Harborth J, Lendeckel W, Yalcin A, Weber K, Tuschl T: **Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells.** *Nature* 2001, **411:**494-498.
20. Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J: **A concise guide to cDNA microarray analysis.** *Biotechniques* 2000, **29:**548-550.
21. Furlong EE, Profitt D, Scott MP: **Automated sorting of live transgenic embryos.** *Nat Biotechnol* 2001, **19:**153-156.
22. Blackshaw S, Livesey R: **Applying genomics technologies to neural development.** *Curr Opin Neurobiol* 2002, **12:**110-114.
23. Dulac C, Axel R: **A novel family of genes encoding putative pheromone receptors in mammals.** *Cell* 1995, **83:**195-206.
24. Eberwine J, Kacharmina JE, Andrews C, Miyashiro K, McIntosh T, Becker K, Barrett T, Hinkle D, Dent G, Marciano P: **mRNA expression analysis of tissue sections and single cells**. *J Neurosci* 2001, **21:**8310-8314.
25. Brady G, Iscove NN: **Construction of cDNA libraries from single cells.** *Meth Enzymol* 1993, **225:**611-623.
26. White KP: **Functional genomics and the study of development, variation and evolution**. *Nat Rev Genet* 2001, **2:**528-537.
27. Altman RB, Raychaudhuri S: **Whole-genome expression analysis: challenges beyond clustering.** *Curr Opin Struct Biol* 2001, **11:**340-347.
28. **EMBL/EMBO Course and Meetings Listing** [http://www-db.embl-heidelberg.de/jss/CoursesConferences.html]
29. Schena M, Shalon D, Davis RW, Brown PO: **Quantitative monitoring of gene expression patterns with a complementary DNA microarray.** *Science* 1995, **270:**467-470.