Biological Sciences Publications   Department of Biological Sciences

2002

Developmental biology: an array of new possibilities

Adnan Ali
University of Windsor

Michael J. Crawford
University of Windsor

2002

Developmental biology: an array of new possibilities

Adnan Ali

Michael J. Crawford

BIOTECHNOLOGY
ADVANCES

ELSEVIER

Developmental Biology: An Array of New Possibilities

Adnan Ali* and Michael J. Crawford*

Dept. Biological Sciences, University of Windsor, 401 Sunset, Windsor, Ontario, N9B 3P4,

CANADA

Abstract

Microarrays offer biologists comprehensive and powerful tools to analyze the involvement of genes in developmental processes at an unprecedented scale. Microarrays which employ defined sequences will permit us to elucidate genetic relationships and responses, while those which employ undefined DNA sequences (ESTs, cDNA or genomic libraries) will help us to discover new genes, relate them to documented gene networks, and to examine the way in which genes (and the process that they themselves control) are regulated. With access to broad new avenues of research come strategic and logistical headaches, most of which are embodied in the reams of data that are created over the course of an experiment. The solutions to these problems have provided interesting computational tools which will allow us to compile huge data sets and to construct a genome-wide view of development. We are on the threshold of a new vista of possibilities where we might consider in comprehensive and yet specific detail, for example, the degree to which diverse organisms utilize similar genetic networks to achieve similar ends.

* Contributed equally

email addresses: mcrawfo@uwindsor.ca, aali@uwindsor.ca

Development of Arrays

The development of DNA arrays heralded a fabulous opportunity to rapidly assess the transcriptional status of genes *thousands* at a time, but few could have predicted the versatility and value of these tools (for review see Schena et al., 1998; Schulze and Downward, 2001; Carulli et al., 1998). While microarrays are useful for obtaining a comprehensive snapshot of transcriptional activity in cells and tissues, they have also proved to be potent instruments for isolating new genes, establishing likely down-stream targets of characterized loci, and even for identifying partners and collaborators in signal transduction pathways. With one microarray and two differentially labeled cDNA samples, a single hybridization experiment has the potential to reveal differences in transcriptional activity at a level of complexity and detail formerly unimaginable. In a standard microarray experiment two pools of RNA are isolated from biological samples: these pools might represent two different treatments or two individual specimens under investigation. Each RNA pool is reverse transcribed using a nucleotide mixture including fluorescently labeled nucleotides (e.g. Cyanine 3-dCTP or Cyanine 5-dCTP). Following a reverse transcription reaction, the labeled cDNAs are mixed and hybridized with a microarray that is then washed under optimally stringent conditions. The microarray is imaged in a laser scanner to generate two microarray images that correspond to the emission of each of the two fluors used during the labeling reaction. Investigators in the field of Developmental Biology have been quick to exploit array technology to its fullest potential for the identification, isolation, and characterization of genes and gene networks. Indeed, it is perhaps in the developmental genetics of the plant *Arabidopsis thaliana* and of the lowly nematode worm *Caenorhabdites elegans* that the technology has been most thoroughly and imaginatively utilized.

Array Substrate, Design, and Utility

Much of the experimental design in an array-based experiment is determined by whether an investigator is searching for a specific gene the function of which has been inferred, or, for differences in general patterns of gene expression. If the latter, there may be no *a priori* assumptions with regard to function, activity, or relationships that differentially expressed genes bear to previously characterized loci. Ideally, the object of an experiment would determine the design of microarray employed. However, given the relative expense and scarcity of available microarrays, investigators frequently make the best of what they can get their hands on. The DNA arrays which are used in Developmental Biology come in several flavors: three commonly employed arrays include oligonucleotide-, cDNA-, and EST-spotted arrays. The use of protein-based arrays which are just coming into use in yeast studies cannot be far away for development studies and are under extensive study (Stoll et al., 2002; Kumar and Snyder, 2001; Zhu et al., 2001). A recent effort has been made to acquire peptide chips which overcome the limitations associated with regular protein chips (Houseman et al., 2002). Another interesting example is the development of protein nanoarrays which will provide a unique opportunity to analyse cell adhesion by providing details regarding cell surface receptor interactions (Lee et al., 2002). One might well imagine how investigators who combine DNA-based arrays with protein arrays will enjoy unparalleled access to developmental mechanisms: the benefits of such an interdisciplinary approach are tantalizing to say the least. While micro-arrays are frequently associated with glass substrates, they have also been printed at high-density on nylon membranes for *Arabidopsis*, sea urchin, nematode worm, and frog (Mochii et al., 1999; Rast et al., 2000; Sasaki et al., 2000; EnvirogenX; and Zorn, unpublished respectively).

The arrays currently in use have densities that vary from 400 to tens of thousands of spots per array. Sensitivity is a major asset of glass-based arrays, and studies using *Arabidopsis* materials have indicated that the acuity of the technique is such that investigators can resolve as

few as two transcripts per cell (Girke et al., 2000). The starting material for an array can be quite broad, for example comprising clones intended to be representative of nearly the entire *C. elegans* genome (Jiang et al., 2001), or highly stage- and tissue-specific. Specialized arrays can be representative of say, mouse genital ridge just prior to gender differentiation (Grimmond et al., 2000), hippocampus (Mody et al., 2001), or neural retina (Livesey et al., 2000). Similarly, the arrayed spots can be of defined sequences such as oligonucleotides (Leemans et al., 2001) or PCR-amplified cDNAs (Jiang et al., 2001), or of uncharacterized clones such as ESTs (Xu et al., 2001). Although best employed in a species-specific manner, a reasonable conservation of mRNA sequences will permit some measure of inter-species utility. For example, *Arabidopsis* arrays were useful for profiling the expression patterns revealed using rapeseed samples (Girke et al., 2000).

Quantitation, Normalization and Analysis of Data

In many of the routinely employed molecular methodologies, the focus of techniques has been to provide qualitative results. However, microarray experiments provide researchers with very large sets of quantitative information. The correct processing of that information is critical for the discovery of novel genes, and for the analyses of a multitude of transcriptional responses. A major challenge for researchers is to evaluate not only the data arising from a biological experiment, but also to critically and strategically establish the parameters which will constrain analysis – there are different types of data transformations that can be applied to make the data comprehensible. In short, a careful evaluation of the data analysis methodology is crucial to the derivation of meaningful results.
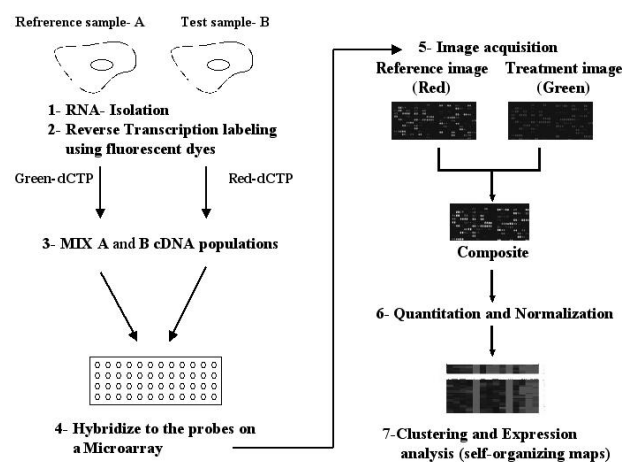
Two concerns that are paramount in microarray studies are common to all array-based approaches: namely, the reliability, and repeatability of the obtained data. Firstly, the array requires duplicated control DNA spots which are distributed in such a way as to provide a measure of reassurance with regard to the evenness of hybridization. If one expects to be capable of detecting minute changes in transcriptional activity, it is critical that DNA spots are distributed and hybridized evenly. Ideally, the microarrays will be tested under identical conditions more than once in order to ensure slide-to-slide consistency, and control spots will hybridize in a similar manner each time. Also, fluor-reversal is commonly used to normalize the differences due to variations in the labeling efficiencies of the two fluors being used. Moreover, the patterns of differential probe hybridization should be consistent among the other genes that are arrayed.

The imaging process takes advantage of the different excitation and emission wavelengths of the two fluors to obtain a picture (see figure 1a). Following the acquisition of images, a composite of the two images reveals the comparative hybridization pattern of the samples used. Images are processed using software to analyze the fluorescence intensities by defining the spots and the background regions, and then the spot light emission intensities are quantified. Standard image processing methods include fixed mask, percentile defined or adaptive mask methods. These methods are applied depending upon quality of chips or according to the needs of a particular experiment. Variable algorithms allow one to define the region of each spot on the microarray in a semi-automated fashion to facilitate spot recognition.

A

Deleted: .

Deleted: These methods are applied depending upon quality of chips or according to the needs of a particular experiment. These vVariable algorithms allow one to define the region of each spot on the microarray in a semi-automated fashion to facilitate spot recognition. These Suchmethods provide the flexibility to handle variations in spot size, shape, and alignment with

359

Formatted

Refrerence sample- A   Test sample- B

1- RNA- Isolation
2- Reverse Transcription labeling
   using fluorescent dyes

Green-dCTP   Red-dCTP

3- MIX A and B cDNA populations

4- Hybridize to the probes on
   a Microarray

5- Image acquisition
Reference image   Treatment image
   (Red)   (Green)

Composite

6- Quantitation and Normalization

7-Clustering and Expression
  analysis (self-organizing maps)

B.

Figure 1. Flow chart of microarray probing, quantitation and analysis. A) The routine mode of data acquisition and analysis is complemented by B) compiled data set transformations to permit re-interpretation and contextualization of individual data points by means of a topomap. Self-organizing cluster maps are normally presented in red and green, and we represent only a portion of a real scan analysis here. The scatter plot and topomap are simplified and representational, and intended for illustration only. Readers should refer to Kim et al., (2001) for a good example of practical interpretation methods.

Such methods provide the flexibility to handle variations in spot size, shape, and alignment with reference to other spots on a microarray. The average intensities for each spot and the surrounding background are compared and a ratio of the intensity of the two fluors is calculated for every spot on the microarray. This ratio has to be corrected by a calculated normalization factor that scales the spot intensities to provide an approximation of the ratio of gene expression between the two samples. Common methods of normalization include using housekeeping gene

expression (Heller et al., 1997), global intensity/ratio averaging (Johannes et al., 1999; Hardwick et al., 1999), sub-array intensity/ratio averaging as well as normalization with known control genes from distantly related species (Loftus et al., 1999; Heller et al., 1997; Lashkari et al., 1997a; Lashkari et al., 1997b). Individually or in combination with the control genes, an accurate and normalized data set reflecting the intensities of gene expression is achieved - this data set can then be subjected to computational analysis to infer comprehensive expression profiles.

Technical Impediments and Considerations Regarding Data Analysis

Like any other experiment, variation can play a large role in compromising the integrity of microarray results. Firstly, variability is introduced during DNA spotting with the result that the array is comprised of differently sized or shaped spots which will hybridize probe in an irregular manner. Secondly, irregularities in array structure i.e., alignment and/or size of spots can make subsequent data collection (laser scanning) and processing problematic. Finally, uneven distribution, hybridization, or washing of sample can have similarly deleterious effects upon data quality. Computational methods for spot finding, delimiting, and assessment are not yet perfect, therefore identification, quantitation, and normalization methods can have a significant effect on the quality of data and hence upon the interpretation of the biology and genetics of the system in question (reviewed by Quackenbush, 2001; Goodman, 2002; Kerr, 2001). In addition to array quality, false signals (noise) either from the hardware itself or from contaminating signals will affect quantitation. In other words, great care must be exercised in selecting appropriate methods for quantitation, normalization, discrimination, and analysis.

Given the ever-increasing density of microarray DNA spots possible, the variable intensity of hybridized probe, and the numbers and types of treatment/developmental state comparisons possible, it is easy to see that the volume of data generated and the complexity of analysis can present a substantial computational and conceptualization challenge. How best to compare the varying probe intensities elicited by different samples hybridizing with thousands of spots of DNA? Indeed "number crunching" has proven to present one of the greatest challenges for microarray analysis.

Another limitation in the final assessment of microarray results is the lack of standards for the presentation of huge data sets. Recently, efforts are being exerted to set standards data presentation and exchange through the establishment of a universally acceptable format, namely, the Minimum Information About a Microarray Experiment (MIAME) format (Brazma et al., 2001; Becker, 2001). A standardized recording of microarray-based expression data in the array databases will allow an independent evaluation, verification, logical interpretation, comparison of high throughput genetic profiles, and sharing of the data. A validated protocol is also necessary to ensure that microarray data from any given experiment provides a legitimate assessment of gene expression profiles.

With several teams working upon these problems, it is not surprising that several approaches to the analysis of large and complex data sets have arisen, nor that each of these methods are evolving rapidly (reviewed by Zweiger, 1999). A fortuitous but perhaps unforeseen consequence of this evolution is that researchers can go back and re-examine their previous microarray scans using modified and intelligent data-mining protocols. This adds further to the immense amount of data generated, but it offers the possibility of consolidating experiments in larger sets than originally envisaged. This complex combination of factors and potential uses creates a profound need to have microarray databases structured to permit researchers access

Deleted: ,

Deleted: Not incidentally, the imaginative approaches taken in response to this wealth of data have provided some of the most powerful means for inferring relationships between genes, their products, their partners, and their targets.

Deleted: findings

through simply defined and user-friendly retrieval and manipulation of protocols. For instance, it would be useful for a microarray database encompassing the developmental gene expression profiles from one organism to be viewed and compared with the profiles of another.

Finally, genes which are identified as likely candidates for subsequent study need to be validated by alternative means. This means that care has to be taken to confirm transcriptional changes using comparative Northern blot analyses, RNase protection, or more frequently, RT-PCR, or *in situ* hybridization of the same embryos or tissues from which samples were derived. When expression differences are confirmed and where the technology permits, a gene's involvement in a developmental process can be further defined with respect to a developmental event through RNAi, gene knockout, and knock-in or rescue experiments. The most comprehensive of these approaches rests within the communities of researchers engaged in *C. elegans,* and the *Arabidopsis* projects, however, we will also outline other approaches in *Xenopus, Drosophila*, Sea Urchin, and mammals.

Deleted: briefly

Examples of Applications for Developmental Microarrays

The huge advantage that is conferred by microarray analysis is that there is usually a high level of delineation between various samples under investigation: each treatment or developmental stage impinges upon and reveals a separate subset of genes. The most straightforward use of microarrays is to employ them as tools to identify potential downstream targets of a specific gene of interest. For example, mutants of a recently characterized pan-neural zinc finger protein, sequoia, were employed to generate probe and to identify genes potentially regulated by this transcription factor in *Drosophila* (Brenman et al., 2001). Similarly, potential murine T-cell differentiation regulators were isolated following a screen in which thymic cells from *Whn* (thymus transcription factor) mutant and wildtype embryos were employed as the sample source for microarrays (Bleul and Boehm, 2001). In addition, microarrays have been employed to discriminate between the effects of different mutations which lead to cardiac hypertrophy (Aronow et al., 2001). In this investigation, the same phenotype can be achieved in mice mutant for four different genes. Contrary to expectation, microarray analysis revealed that the phenotypes are reached through the elicitation of distinct pathways involving very different sets of genes. Other investigators have elected to use microarrays to reveal broad sets of genes which are temporally regulated during early development by discriminating between maternal and early zygotic embryos (Altmann et al., 2001). Temporal discrimination need not be restricted to transitions as marked as different developmental stages: it can be so acute that microarrays have been employed to reveal regulators of circadian rhythms in *Drosophila* (Claridge-Chang et al., 2001; McDonald and Rosbash, 2001). Alternatively, microarrays have been employed to discover new genes with specific attributes, but without prior knowledge of interacting partners or activators: parthenogenetic and normal wildtype mouse embryos were the source of sample to search for paternally imprinted loci (Kobayashi et al., 2000).

Examination of mechanisms of Developmental Gene Regulation

The field of Developmental Biology has been intensely investigated to identify potent regulatory molecules. Through a cascade of protein-protein, cell-cell, and environmental interactions, embryonic cells utilize particular signaling pathways that lead to specific changes in the expression of particular genes. Indeed, transcriptional regulation in itself is a very complex operation and involves a combination of proteins and regulatory elements. In its simplest definition, signal transduction requires that *trans*-acting factors identify specific *cis*-acting DNA

binding sites and thereby initiate gene transcription. Several regulatory DNA elements may exist in a promoter and work in concert to regulate the transcriptional status of a specific gene. Proceeding beyond a developmental stage in an organism may depend upon the activation or inactivation of numerous genes, each of which is regulated by specific transcription factors. Although altered patterns of expression for a given gene during development could be due to the activity of a subset of regulatory factors, microarrays offer the opportunity to establish whether these regulatory factors are binding with other promoters *in vivo* to activate, in effect, a battery of genes. The complexion of the situation is further complicated by the feature that changes occur in transcription factor target specificity, DNA-binding activity, and with access to promoter regions. In recent studies, microarrays have been exploited to identify DNA binding sites in promoter regions at a genomic scale (Iyer et al., 2001). In this *in vivo* analysis of protein-DNA interactions, the binding sites for the yeast transcription factors SBF and MBF were determined using microarrays that contained over 6K of amplified intergenic regions from yeast genomic DNA. Probes were made first by cross-linking proteins to DNA (*in vivo*), and then by affinity purifying cross-linked DNA using antibodies against the two transcription factors. After amplifying and labeling the co-immunoprecipitated DNA, it was hybridized to microarrays. The results were astonishing since, in addition to the previously known target sites for these factors, 200 new binding sites were identified for these transcription factors. Although these binding sites may not transpire to be directly responsible for in vivo transcriptional responses, the study provided ample fodder for intelligently directed future studies that may shed light on, for example, the role of these sites and binding factors in chromatin assembly and remodeling. There is great potential that the same strategy can be employed at various stages of embryonic development to investigate where regulatory networks impinge by selecting for instances where interactions between known *trans*-acting factors and their corresponding *cis*-acting elements occur in unknown gene promoters. Such studies will help us to understand the multitude of links that comprise signaling networks within the developing embryo. In addition to identifying putative target sites for transcription factors, such an approach may also clarify how these factors are temporally restricted *in vivo*.

Deleted: ing

Sea Urchin

In an attempt to reveal the regulatory networks at play during embryogenesis, a group of researchers has recently employed a combination of molecular and cellular approaches and differential macroarray screens during sea urchin development (Davidson et al., 2002). In this study, macroarrays were sequentially hybridized with pre- and post-subtraction endomesodermal cDNA pools. The non-redundant gene products which were prevalent in post-subtraction samples were characterized as transcription factors, signaling pathway proteins, and a number of other modulators of embryogenesis. Furthermore, the results from these macroarray experiments were correlated with the phenotyptic consequences of functional knockouts and genetic perturbations. This permitted investigators to build a model for the genetic networks that regulate cellular specification during development.

Deleted: s

Drosophila

A genome-wide approach has been used to access the genetics which underlies development in *Drosophila* (White et al., 1999). Genes that were unique to certain stages of insect metamorphosis were identified using 19 arrays representing 6 time points with reference to pupa formation. Pupation is a particularly useful event in which to employ microarrays since it allows investigators to examine the effect of a potent hormone stimulus, namely an ecdysone pulse, upon gene expression. In essence, the hormone acts as a trigger to set in motion a complex

series of genetic and developmental programs. Since the arrays contain several thousand genes, many of which have been previously characterized, and since some of the inter-relationships for these genes are known, imaginative use can be made of prior knowledge to identify new elements in a specific metabolic pathway: the existing genes act as a sort of validating tool. For example, if it is known that genes x, y, and z are among the candidate genes involved in a specific metabolic pathway or signal transduction network, then inspections of the probed arrays can be evaluated in light of the known relationships. Newly identified gene candidates which cluster with a group of previously linked network of genes can then be assessed with regard to their behavior in another experiment or at another sampling time-point documented to affect the known genes x, y, and z. Essentially, analysis program sub-routines can be created that perform an informed or "supervised" discrimination of gene expression patterns by comparing them to previously defined gene networks. These algorithms have a great power and potential to reveal relationships between gene expression and the functional classification of highlighted genes, thereby allowing one to stream a vast amount of raw data into a logical and comprehensible output.

Clearly this sort of approach will be most effective when a substantial body of information has already been established - most experiments presently do not have the luxury of access to this degree of analysis, and inspections of gene activity are therefore "unsupervised". However, in the *Drosophila* experiments, sufficient details were already known to permit limited supervised inspections. More than 10% of genes exhibited in excess of a 3 fold change in relative mRNA abundance. Using pairwise correlations, also confirmed by self organizing maps, it was possible to group genes into those clusters that exhibited similar expression changes and temporal profiles with reference to pupa formation. The analyses showed that 44% of genes are repressed by the metamorphosis-inducing hormone ecdysone since their expression is higher prior to the ecdysone surge. Conversely, 31% of genes are inducible by ecdysone and are expressed at high levels during late larval stages. Clustering new genes by means of supervised analysis methodologies helped investigators both to predict and to integrate these genes into specific and discrete modules of genetic activity. Appropriate grouping and assignment of these genes narrowed the number of candidate genes which required subsequent investigation regarding a specific metabolic pathway. In addition to the identification of ecdysone-inducible genes, nine genes involved in glycolytic pathways, as well as others important to the citric acid cycle, oxidative phosphorylation, amino acid metabolism, glycogen synthesis, and fatty acid oxidation were identified as showing changes in their expression. Clearly, changes in the expression of genes regulating cellular metabolism during metamorphosis are important to the normal execution of development. A direct genome-wide look at these changes allows us to view hierarchically ordered and developmentally programmed molecular processes. For example, the study revealed a clear and conserved correlation of expression changes among the genes responsible for larval muscle breakdown, adult myogenesis, and CNS differentiation and restructuring.

The abundant data that is obtained through such studies reveals a plethora of genes which might be regulated in a temporally or spatially restricted manner. However, the mere involvement of genes in modules of activity does not necessarily indicate their functional significance. Questions which have to remain at the forefront of an investigator's mind remain:

- is a change in gene expression specific for the process under study or is it a secondary (indirect) effect/artifact;
- how can the master regulatory switches within the "jungle" of genetic changers be revealed;

- how are the different developmental transcription networks interconnected and how can they regulate a series of metamorphic changes with such fidelity?

Mutations become critically important in this regard. Model organisms, where mutants are available, can provide a valuable entré where one can compare treatment groups and exclude extraneous and indirect changes on gene expression. Site directed mutagenesis of specified genes can give useful information that, in combination with wild-type expression, can be used to decipher genetic networks. Furthermore, the advent of more sophisticated technology and more innovative approaches are going to extend this kind of analysis. For instance, bead-based array (multiplex microsphere arrays) methodology can benefit post-microarray analysis of developmentally regulated genes. In such experiments, a handful of genes that show big changes during a specific developmental event can be used to screen hundreds of mutants simultaneously. This would provide further confirmatory microarray data as well as aid in specifying the significance of each gene in the mechanics of regulated development.

Caenorhabdites elegans (nematode worm)

In one of the most comprehensive series of microarray-based experiments, a consortium of investigators employed a *C. elegans* microarray to compare the gene expression profiles and to derive general associations or assemblies of genes which were co-regulated during development. Initially, various worms mutant for germ cell development were compared with wildtype using microarrays (Jiang et al., 2001). Microarray hybridization patterns were scanned and processed as previously described. Probe intensity differences at the gene spots were then compared through the use of self-organizing maps (Tamayo et al., 1999) or through the use of hierarchical clustering algorithms. In this way genes which were either up- or down- regulated in similar fashion following a specific treatment or under specific conditions were identified. This process demonstrated that 2,171 out of 17,871 genes were sex-regulated (involved in male vs hermaphrodite development), and 1416 were involved specifically in germ cell development (Jiang et al., 2001). Of these, approximately 650 were involved in sperm development (Reinke et al., 2000). A gene's annotation and functional relevance is typically based upon the genes' response to specific conditions. Moreover, a specific gene product may have some properties and exert influences that are still unknown. The uncovering of these diverse effects is one of the aims that drove these investigators to consolidate information from several different experiments (Kim et al., 2001). It was hoped that microarrays would provide an opportunity to demonstrate new functions of genes that exhibit big and perhaps unforeseen changes in transcriptional status in an experiment.

Data-mining from such a consolidation of results is best exemplified by the *C. elegans* project. The means to build matrices representing a compendia of results from many microarray experiments had been demonstrated in an earlier project on yeast (Hughes et al, 2000). The innovation of Kim and his collaborators was to provide a simpler method to visualize compiled pairwise comparisons of *C. elegans* gene expression profiles from many different experiments (different mutants, or experimental conditions)(Kim et al., 2001). When the profiles of gene activity for 553 experiments were installed in this vast data matrix, Pearson correlations were made in a pairwise fashion to construct a 2D scatter plot which highlighted the clusters of genes that exhibited the greatest changes in expression (see figure 1b). This scatter plot was converted into a three dimensional topographical representation (terrain map) where mound height signified the density of genes exhibiting similar activity profiles (Kim et al., 2001). Essentially, the derived "topomap" separates genes into 44 "mountains". Genes which separate into one mount might tend to be involved in oocyte development, while those that separate into another

might be linked to spermatogenesis. Genes which are very tightly linked in a pathway often tend to segregate into sub-regions of a mountain. In the case of germ cell differentiation genes, many co-localizing gene profiles were loci which had been previously identified as genes involved in specific pathways. There were a few additional genes of unknown function, however, for which oligonucleotides were made to perform RNA interference experiments: a proof of principle for this approach to data analysis would necessitate that interference with the activity of these new loci would result in developmental effects which were similar to those of the previously characterized loci. It transpired that the new genes did indeed represent members of the same differentiation pathway, and their mutant phenotypes were rescue-able by cosmid constructs containing the complete genomic sequence (Kim, 2000).

In a systematic manner, the Consortium is knocking out all identified *C elegans* genes using RNAi (Kim, 2001a; Kim, 2001b). *C. elegans* offers the additional advantages that there have been many mutant lines isolated, PCR-based gene knockout technologies are available to confirm the phenotypes of RNAi and mutagenized alleles, and there is an extensive array of cosmids which can be employed to test the "rescue" efficacy of genomic fragments. In other words, once potential players in a gene network have been identified by microarray analysis, they can be first "knocked out" and the phenotypes then "rescued" using genomic cosmids. The consortium then plans to construct a protein-interaction map using yeast two-hybrid approaches.

Arabidopsis

A comprehensive approach has also been undertaken by the Arabidopsis Functional Genomic Consortium (http://afgc.stanford.edu). This group has established a database where microarray results can be deposited and compared. They have also linked the microarray printing effort to a centralized gene knock-out facility. A large collection of T-DNA mutagenized lines have been generated and isolated (Krysan et al., 1999). For a nominal fee, investigators can have the library of mutants screened for their (disrupted) gene of interest.

Xenopus laevis (South African Clawed Frog) and Zebrafish

Microarray use in frog embryology studies is still at its infancy. Recently, microarrays comprising defined genes and ESTs have been employed to identify changes in gene activity which occur as zygotic transcription supplants maternally-encoded transcript as the template for translated protein (Altmann et al., 2001). In addition, as proof of principle, the technique was used to identify novel genes, genes regulated in response to experimental manipulation, and genes which are expressed in a spatially restricted manner.

Zebrafish microarrays are currently under development, and an interesting strategy is being employed to reduce redundancy of EST sequences represented on the microarrays. A PCR-based strategy called oligonucleotide fingerprinting (OFP) has been used to weed out EST clones which overlap and thereby reduce the number of clones which need to be spotted on an array (Clark et al., 2001). It is anticipated that high density arrays possessed of twenty-five thousand unique sequences will be available for use shortly.

Mammals

Human and mouse arrays have been among the earliest to undergo development, but they have tended to be expensive, in short supply, and available to laboratories which are well funded to do health-related research. Human embryological studies are just under way, and recently investigators have made use of an arrayed cDNA library to isolated genes which are differentially expressed in adult and fetal testes (Xiao et al., 2002). Good use has been made of arrays and the numerous gene knock-out and transgenic lines to identify lineage-specific or

Deleted: ¶

signal transduction pathway-specific genes of interest (Grimmond et al., 2000; Bleul and Boehm, 2001; Mody et al., 2001; Aronow et al., 2001). In the last few months, microarrays have even been employed to identify loci subject to genetic imprinting (Kobayashi et al., 2000; Mizuno et al., 2002). In one of the studies, a cDNA library was spotted and material from parthenogenetic and androgenetic embryos employed to discriminate. Eight formerly identified and three new loci were isolated and confirmed as imprinted loci by secondary means (Mizuno et al., 2002).

An important technical innovation which is likely to have great importance in developmental studies was designed to assess the functional significance of differentially expressed genes (Ziauddin and Sabatini,, 2001). Human cDNAs cloned into an expression system were spotted onto an array, and human embryonic kidney (HEK293) cells were then plated directly onto the slide. The experiments were designed to test the efficiency of cell transfection as well as the ability of different tryosine kinase signaling cascade members to elicit a response detectable with tagged antibodies. Cells expressed proteins encoded by marker constructs (green fluorescent protein), and they did so in a spatially restricted and highly regular punctate pattern that was coincident with the pattern of array spotting (Ziauddin and Sabatini,, 2001). A similar strategy is likely to prove valuable for discerning the activity of specific genes in the regulation of, for example, ectodermal placode formation, or mesenchymal cell patterning etc.

Dictyostelium (slime mold)

A recent addition to the research tool chest for the social myxamoeba is the development of *Dictyostelium* microarrays. The first, indeed the only study published to date describes a project in which the 24 hour developmental cycle of this organism has been divided into two hour segments and gene expression profiles obtained and compared (Sasik et al., 2002). The intent of the project is to identify networks of genetic activity which are invoked at different stages of development.

Future prospects

Microarrays are not limited to programs of gene discovery, but have the potential to offer access to additional and potent tools for the analysis of developmental functions. For example, microarrays have been employed to find translation-impeding oligonucleotides (aptamers) (Brody and Gold, 2000), and the efficacy of antisense oligonucleotides have been assessed by means of microarrays prior to *in vivo* use (Sohail et al., 2001).

Microarray-driven improvements to computational tools is also proving valuable for the construction of arrays specifically comprised of developmentally regulated genes. This later approach is being pioneered by plant scientists who employ "electronic Northern" analysis to discern candidates for spotting upon the basis of sequence representation in published cDNA and EST libraries (Rafalski et al., 1998). A further use with great potential for disclosing mechanisms of gene regulation involves designing microarrays that contain specific amplified sequences from various exons of selective genes. This approach would allow investigators to identify and to characterize genes that use alternative splicing as a mechanism for regulation during embryogenesis or organogenesis.

Finally, microarrays have recently emerged as important tools for gene expression profiling in other fields; for genomic DNA analyses; for the classification of cancer types; and for the identification of single nucleotide polymorphisms. The major role currently being played by microarrays in Developmental Biology is being mirrored in other endeavors such as gene discovery, disease diagnosis, drug discovery (Pharmacogenomics) (Debouck and Goodfellow, 1999; Zanders, 2000), and toxicological research (Toxicogenomics) (Nuwaysir et al., 1999).

Microarray technology is still in its infancy, but the power and the comprehensive nature of its capacity augers well for significant advancements in the near term. Perhaps most of all, the technology may serve as the seed to foster multidisciplinary approaches to developmental questions for years to come.

Acknowledgments

References Cited

Altmann, C.R., Bell, E., Sczyrba, A., Pun, J., Bekiranov, S., Gaasterland, T. and Brivanlou, A.H. (2001) Microarray-based analysis of early development in Xenopus laevis. Dev Biol 236, 64-75.

Aronow, B.J., Richardson, B.D. and Handwerger, S. (2001) Microarray analysis of trophoblast differentiation: gene expression reprogramming in key gene function categories. Physiol Genomics 6, 105-16.

Becker, K.G. (2001) The sharing of cDNA microarray data. Nat Rev Neurosci 2, 438-40.

Bleul, C.C. and Boehm, T. (2001) Laser capture microdissection-based expression profiling identifies PD1- ligand as a target of the nude locus gene product. Eur J Immunol 31, 2497-503.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C. P., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., and Vingron, M. (2001) Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nature Genetics 29, 365-71.

Brenman, J.E., Gao, F.B., Jan, L.Y. and Jan, Y.N. (2001) Sequoia, a tramtrack-related zinc finger protein, functions as a pan- neural regulator for dendrite and axon morphogenesis in Drosophila. Dev Cell 1, 667-77.

Brody, E.N. and Gold, L. (2000) Aptamers as therapeutic and diagnostic agents. J Biotechnol 74, 5-13.

Carulli, J.P., Artinger, M., Swain, P.M., Root, C.D., Chee, L., Tulig, C., Guerin, J., Osborne, M., Stein, G., Lian, J. and Lomedico, P.T. (1998) High throughput analysis of differential gene expression. J Cell Biochem Suppl 30-31, 286-96.

Claridge-Chang, A., Wijnen, H., Naef, F., Boothroyd, C., Rajewsky, N. and Young, M.W. (2001) Circadian regulation of gene expression systems in the Drosophila head. Neuron 32, 657-71.

Clark, M.D., Hennig, S., Herwig, R., Clifton, S.W., Marra, M.A., Lehrach, H., Johnson, S.L. and Group t, W. (2001) An oligonucleotide fingerprint normalized and expressed sequence tag characterized zebrafish cDNA library. Genome Res 11, 1594-602.

Davidson, E. H., Rast, J. P., Oliveri, P., Rasnick, A., Calestani, C., Yuh, C-H., Minokawa, T., Amore, G., Hinman, V., Arenas-Mena, C., Otim, O., Brown, C. T., Livi, C. B., Lee, P. Y., Revilla, R., Rust, A. G., Pan, Z. J., Schilstra, M. J., Clarke, P. J. C., Arnone, M. I., Rowen, L., Cameron, R. A., McClay, D. R., Hood, L., and Bolouri, H. (2002) A genomic regulatory network for development. Science 295, 1669-78.

Debouck, C. and Goodfellow, P.N. (1999) DNA microarrays in drug discovery and development. Nat Genet 21, 48-50.

Girke, T., Todd, J., Ruuska, S., White, J., Benning, C. and Ohlrogge, J. (2000) Microarray analysis of developing Arabidopsis seeds. Plant Physiol 124, 1570-81.

Goodman, N. (2002) Biological data becomes computer literate: new advances in bioinformatics. Curr Opin Biotechnol 13, 68-71.

Grimmond, S., Van Hateren, N., Siggers, P., Arkell, R., Larder, R., Soares, M.B., de Fatima Bonaldo, M., Smith, L., Tymowska-Lalanne, Z., Wells, C. and Greenfield, A. (2000) Sexually dimorphic expression of protease nexin-1 and vanin-1 in the developing mouse gonad prior to overt differentiation suggests a role in mammalian sexual development. Hum Mol Genet 9, 1553-60.

Hardwick, J.S., Kuruvilla, F.G., Tong, J.K., Shamji, A.F. and Schreiber, S.L. (1999) Rapamycin-modulated transcription defines the subset of nutrient- sensitive signaling pathways directly controlled by the Tor proteins. Proc Natl Acad Sci U S A 96, 14866-70.

Heller, R.A., Schena, M., Chai, A., Shalon, D., Bedilion, T., Gilmore, J., Woolley, D.E. and Davis, R.W. (1997) Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. Proc Natl Acad Sci U S A 94, 2150-5.

Deleted: ⸺Page Break⸺

Houseman, B. T., Huh, J. H., Kron, S. J., and Mrksich, M. (2002) Peptide chips for the quantitative evaluation of protein kinase activity. Nature Biotechnology 20, 270-4.

Hughes, T.R., Marton, M.J., Jones, A.R., Roberts, C.J., Stoughton, R., Armour, C.D., Bennett, H.A., Coffey, E., Dai, H., He, Y.D., Kidd, M.J., King, A.M., Meyer, M.R., Slade, D., Lum, P.Y., Stepaniants, S.B., Shoemaker, D.D., Gachotte, D., Chakraburtty, K., Simon, J., Bard, M. and Friend, S.H. (2000) Functional discovery via a compendium of expression profiles. Cell 102, 109-26.

Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M. and Brown, P.O. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature 409, 533-8.

Jiang, M., Ryu, J., Kiraly, M., Duke, K., Reinke, V. and Kim, S.K. (2001) Genome-wide analysis of developmental and sex-regulated gene expression profiles in Caenorhabditis elegans. Proc Natl Acad Sci U S A 98, 218-23.

Johannes G, Carter MS, Eisen MB, Brown PO, and Sarnow P. (1999) Identification of eukaryotic mRNAs that are translated at reduced cap binding complex eIF4F concentrations using a cDNA microarray. Proc Natl Acad Sci U S A 96, 13118-23.

Kerr, M.K. and Churchill, G.A. (2001) Statistical design and the analysis of gene expression microarray data. Genet Res 77, 123-8.

Kim, S.K. (2000) C. elegans functional genomics: profiling gene expression patterns with DNA microarrays. In, Society for Developmental Biology 59th Annual Meeting, Developmental Biology, Boulder, Colorado, Vol. 222, pp. 230.

Kim, S.K. (2001a) C. elegans: mining the functional genomic landscape. Nat Rev Genet 2, 681-9.

Kim, S.K. (2001b) Functional genomics: the worm scores a knockout. Curr Biol 11, R85-7.

Kim, S.K., Lund, J., Kiraly, M., Duke, K., Jiang, M., Stuart, J.M., Eizinger, A., Wylie, B.N. and Davidson, G.S. (2001) A gene expression map for Caenorhabditis elegans. Science 293, 2087-92.

Kobayashi, S., Wagatsuma, H., Ono, R., Ichikawa, H., Yamazaki, M., Tashiro, H., Aisaka, K., Miyoshi, N., Kohda, T., Ogura, A., Ohki, M., Kaneko-Ishino, T. and Ishino, F. (2000) Mouse Peg9/Dlk1 and human PEG9/DLK1 are paternally expressed imprinted genes closely located to the maternally expressed imprinted genes: mouse Meg3/Gtl2 and human MEG3. Genes Cells 5, 1029-37.

Krysan, P.J., Young, J.C. and Sussman, M.R. (1999) T-DNA as an insertional mutagen in Arabidopsis. Plant Cell 11, 2283-90.

Kumar, A. and Snyder, M. (2001) Emerging technologies in yeast genomics. Nat Rev Genet 2, 302-12.

Lashkari, D.A., DeRisi, J.L., McCusker, J.H., Namath, A.F., Gentile, C., Hwang, S.Y., Brown, P.O. and Davis, R.W. (1997a) Yeast microarrays for genome wide parallel genetic and gene expression analysis. Proc Natl Acad Sci U S A 94, 13057-62.

Lashkari, D.A., McCusker, J.H. and Davis, R.W. (1997b) Whole genome analysis: experimental access to all genome sequenced segments through larger-scale efficient oligonucleotide synthesis and PCR. Proc Natl Acad Sci U S A 94, 8945-7.

Lee, K-B, Park, S-J., Mirkin, C. A., Smith, J. C., and Mrksich, M. (2002) Protein nanoarrays generated by Dip-Pen nanolithography. Science 295, 1702-5.

Leemans, R., Loop, T., Egger, B., He, H., Kammermeier, L., Hartmann, B., Certa, U., Reichert, H. and Hirth, F. (2001) Identification of candidate downstream genes for the homeodomain transcription factor Labial in Drosophila through oligonucleotide-array transcript imaging. Genome Biol 2.

Livesey, F.J., Furukawa, T., Steffen, M.A., Church, G.M. and Cepko, C.L. (2000) Microarray analysis of the transcriptional network controlled by the photoreceptor homeobox gene Crx. Curr Biol 10, 301-10.

Loftus, S.K., Chen, Y., Gooden, G., Ryan, J.F., Birznieks, G., Hilliard, M., Baxevanis, A.D., Bittner, M., Meltzer, P., Trent, J. and Pavan, W. (1999) Informatic selection of a neural crest-melanocyte cDNA set for microarray analysis. Proc Natl Acad Sci U S A 96, 9277-80.

McDonald, M.J. and Rosbash, M. (2001) Microarray analysis and organization of circadian gene expression in Drosophila. Cell 107, 567-78.

Mizuno, Y., Sotomaru, Y., Katsuzawa, Y., Kono, T., Meguro, M., Oshimura, M., Kawai, J., Tomaru, Y., Kiyosawa, H., Nikaido, I., Amanuma, H., Hayashizaki, Y. and Okazaki, Y. (2002) Asb4, Ata3, and Dcn are novel imprinted genes identified by high-throughput screening using RIKEN cDNA microarray. Biochem Biophys Res Commun 290, 1499-505.

Mochii, M., Yoshida, S., Morita, K., Kohara, Y. and Ueno, N. (1999) Identification of transforming growth factor-beta-regulated genes in caenorhabditis elegans by differential hybridization of arrayed cDNAs. Proc Natl Acad Sci U S A 96, 15020-5.

Mody, M., Cao, Y., Cui, Z., Tay, K.Y., Shyong, A., Shimizu, E., Pham, K., Schultz, P., Welsh, D. and Tsien, J.Z. (2001) Genome-wide gene expression profiles of the developing mouse hippocampus. Proc Natl Acad Sci U S A 98, 8862-7.

Nuwaysir, E.F., Bittner, M., Trent, J., Barrett, J.C. and Afshari, C.A. (1999) Microarrays and toxicology: the advent of toxicogenomics. Mol Carcinog 24, 153-9.

Quackenbush, J. (2001) Computational analysis of microarray data. Nat Rev Genet 2, 418-27.

Rafalski, J.A., Hanafey, M., Miao, G.H., Ching, A., Lee, J.M., Dolan, M. and Tingey, S. (1998) New experimental and computational approaches to the analysis of gene expression. Acta Biochim Pol 45, 929-34.

Rast, J.P., Amore, G., Calestani, C., Livi, C.B., Ransick, A. and Davidson, E.H. (2000) Recovery of developmentally defined gene sets from high-density cDNA macroarrays. Dev Biol 228, 270-86.

Reinke, V., Smith, H.E., Nance, J., Wang, J., Van Doren, C., Begley, R., Jones, S.J., Davis, E.B., Scherer, S., Ward, S. and Kim, S.K. (2000) A global profile of germline gene expression in C. elegans. Mol Cell 6, 605-16.

Sasaki, Y., Asamizu, E., Shibata, D., Nakamura, Y., Kaneko, T., Awai, K., Masuda, T., Shimada, H., Takamiya, K., Tabata, S. and Ohta, H. (2000) Genome-wide expression-monitoring of jasmonate-responsive genes of Arabidopsis using cDNA arrays. Biochem Soc Trans 28, 863-4.

Sasik, R., Iranfar, N., Hwa, T. and Loomis, W.F. (2002) Extracting transcriptional events from temporal gene expression patterns during Dictyostelium development. Bioinformatics 18, 61-6.

Schena, M., Heller, R.A., Theriault, T.P., Konrad, K., Lachenmeier, E. and Davis, R.W. (1998) Microarrays: biotechnology's discovery platform for functional genomics. Trends Biotechnol 16, 301-6.

Schulze, A. and Downward, J. (2001) Navigating gene expression using microarrays--a technology review. Nat Cell Biol 3, E190-5.

Sohail, M., Hochegger, H., Klotzbucher, A., Guellec, R.L., Hunt, T. and Southern, E.M. (2001) Antisense oligonucleotides selected by hybridisation to scanning arrays are effective reagents in vivo. Nucleic Acids Res 29, 2041-51.

Stoll, D., Templin, M.F., Schrenk, M., Traub, P.C., Vohringer, C.F. and Joos, T.O. (2002) Protein microarray technology. Front Biosci 7, C13-32.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E.S. and Golub, T.R. (1999) Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci U S A 96, 2907-12.

White, K.P., Rifkin, S.A., Hurban, P. and Hogness, D.S. (1999) Microarray analysis of Drosophila development during metamorphosis. Science 286, 2179-84.

Xiao, J., Xu, M., Li, J., Chang Chan, H., Lin, M., Zhu, H., Zhang, W., Zhou, Z., Zhao, B. and Sha, J. (2002) NYD-SP6, a novel gene potentially involved in regulating testicular development/spermatogenesis. Biochem Biophys Res Commun 291, 101-10.

Xu, X.R., Huang, J., Xu, Z.G., Qian, B.Z., Zhu, Z.D., Yan, Q., Cai, T., Zhang, X., Xiao, H.S., Qu, J., Liu, F., Huang, Q.H., Cheng, Z.H., Li, N.G., Du, J.J., Hu, W., Shen, K.T., Lu, G., Fu, G., Zhong, M., Xu, S.H., Gu, W.Y., Huang, W., Zhao, X.T., Hu, G.X., Gu, J.R., Chen, Z. and Han, Z.G. (2001) Insight into hepatocellular carcinogenesis at transcriptome level by comparing gene expression profiles of hepatocellular carcinoma with those of corresponding noncancerous liver. Proc Natl Acad Sci U S A 98, 15089-94.

Zanders, E.D. (2000) Gene expression analysis as an aid to the identification of drug targets. Pharmacogenomics 1, 375-84.

Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A., Gerstein, M. and Snyder, M. (2001) Global analysis of protein activities using proteome chips. Science 293, 2101-5.

Zweiger, G. (1999) Knowledge discovery in gene-expression-microarray data: mining the information output of the genome. Trends Biotechnol 17, 429-36.

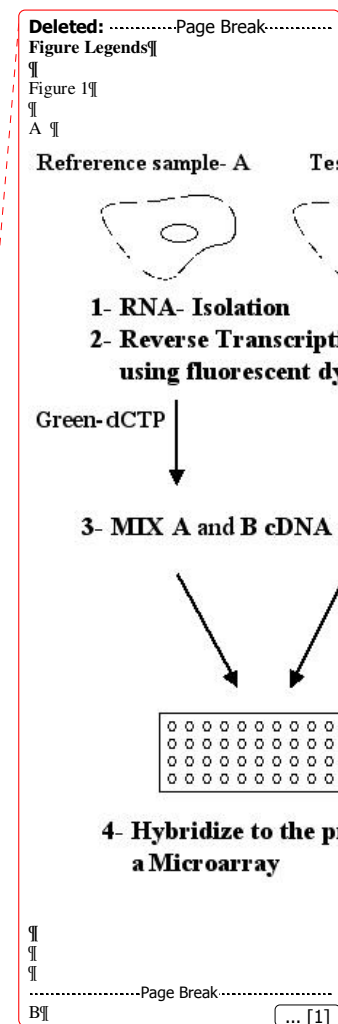Ziauddin, J. and Sabatini, D.M. (2001) Microarrays of cells expressing defined cDNAs. Nature 411, 107-10.

Deleted: ----------Page Break----------
Figure Legends¶
¶
Figure 1¶
¶
A ¶

Refference sample- A   Te

1- RNA- Isolation
2- Reverse Transcripti
   using fluorescent dy

Green-dCTP

3- MIX A and B cDNA

4- Hybridize to the p
   a Microarray

¶
¶
¶
----------Page Break----------
B¶   ... [1]

--Page Break--

Figure Legends

Figure 1

A

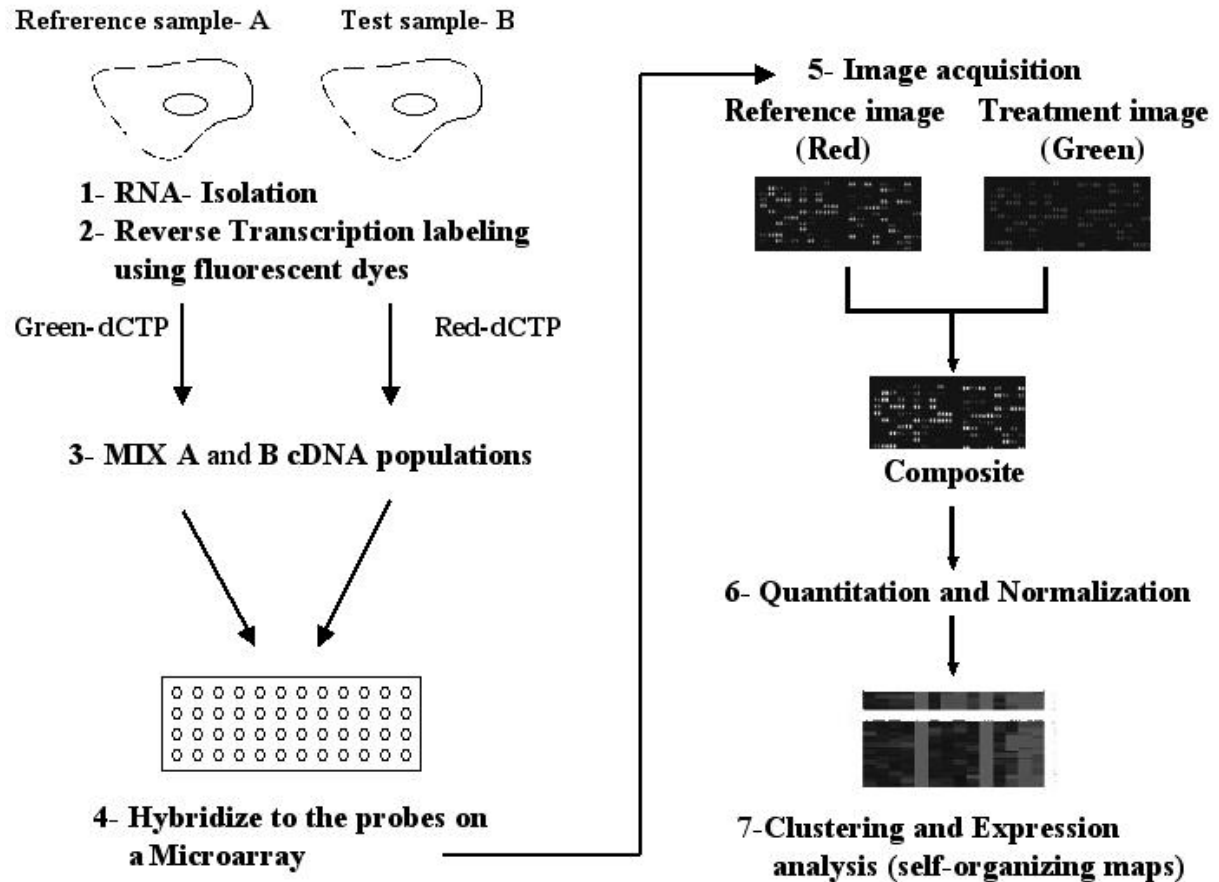--Page Break--

B

|

Figure 1. Flow chart of microarray probing, quantitation and analysis. A) The routine mode of data acquisition and analysis is complemented by B) compiled data set transformations to permit re-interpretation and contextualization of individual data points by means of a topomap. Self-organizing cluster maps are normally presented in red and green, and we represent only a portion of a real scan analysis here. The scatter plot and topomap are simplified and representational only, and intended for illustration only. Readers should refer to Kim et al., (2001) for a good example of practical interpretation methods.