An Overview Of DNA Microarrays: From Technology To Biology And Beyond

by
Lance David Miller

Introduction.

The so-called central dogma of molecular biology describes the flow of biological information: from DNA, to RNA, to protein. In molecular terms, when a biological pathway is operative, genomic DNA (eg, a gene) is transcribed into messenger RNA (mRNA; also known as the transcript), and this RNA is translated into protein (the gene product). The protein is the primary molecular machinery that carries out cellular functions, and for this reason, the activity of proteins – as defined by their relative abundance, activation states and molecular modifications – are generally regarded as the crucial determinants of a cell’s behavior. Genes, on the other hand, serve as the blueprints from which proteins are constructed, and when accessed by the cell’s transcriptional machinery, are transcribed into the scaffolding on which proteins are built: the mRNA. As such, the relative abundance and complexity of mRNA transcripts (collectively known as the transcriptome) reflects the potential of all cellular pathways. Knowledge of the transcriptome therefore provides clues as to the molecular processes that are operative, or not, in a cell.

For decades, scientists have used methods of gene expression analysis to better understand the molecular logic underlying biological processes and cellular states. This has, in many cases, led to the discovery of genes that play important roles in cell behavior and human disease. Historically, the techniques employed for determining gene expression levels, such as Northern Blot analysis or Differential Display, have been limited to the gene-by-gene approach, that is, looking for transcripts one or a few at a time. Recently, however, the invention of genome-wide transcript screening methods such as SAGE (serial analysis of gene expression) and DNA microarray technology have overcome this limitation, allowing for the determination of the relative abundance and complexity of tens of thousands of transcripts simultaneously.

Expression Microarray Technology.

DNA microarrays, also called Gene Chips, are tools for studying gene dynamics in a highly parallel fashion. In simplest terms, they are ordered collections of DNA sequences (probes) deposited on – and chemically bound to – solid surfaces (like microscope slides) or three-dimensional matrices. Though their composition and purpose have evolved considerably since their introduction in 1995, the original and still most popular format is the expression microarray designed to measure the relative abundance of mRNA transcripts. Here, the DNA probes, which can number in the hundreds of thousands on a single chip, are derived from the transcribed regions of genes. In form,
the probes may be long sequences several hundred to several thousand bases in length amplified from cloned mRNA transcripts, called cDNA probes, or short DNA sequences called oligonucleotide probes that are 20 to 80 bases in length and can be synthesized in situ (on the microarray surface) by phosphoramidite chemistry and light-sensitive enzymes.

The fundamental principle by which microarrays work is often referred to as DNA hybridization, whereby a single strand of DNA (the probe, for example) is capable of annealing to, and thus forming a highly stable duplex structure with, a complementary strand of DNA (also known as the target in a microarray experiment; see illustration). In a typical two-color microarray hybridization experiment, the first step is the labelling of the cDNA targets. Here, cellular mRNA is extracted from two cell populations (or tissues) for which relative gene expression levels are to be compared. These populations are usually referred to as the “test” and “reference” RNAs. The RNAs are then reverse transcribed into cDNA by a simple enzymatic reaction, and fluorescently labelled with different color fluorophores (eg, Cy3 and Cy5 dyes) whereby one cDNA target population will fluoresce green (Cy3) and the other will fluoresce red (Cy5). These targets are then purified, mixed together and simultaneously hybridized to the same microarray. Following the hybridization reaction, which is often accomplished in an overnight incubation, the microarrays are washed, dried, and scanned for detection of fluorescence on the DNA probes. Presumably, if a gene is expressed in one or both of the RNA samples, then its mRNA will be converted into fluorescently labeled cDNA, which will subsequently bind to its corresponding probe during the hybridization reaction. This hybridization event is detected by a microarray scanner which focuses specific wavelengths of laser light on the probes in order to excite fluorescence of the Cy3 and Cy5 fluorophores. This signal information is then captured in a two-channel 16-bit TIFF image which encodes the emitted fluorescence of each fluorophore as relative units of pixel color saturation (also called signal intensity) in each channel of the TIFF image. If the magnitude of the average signal intensity for a given gene probe is equivalent in the Cy3 and Cy5 channels, then the actual transcript levels of that gene in the two RNA samples are said to be equivalently expressed. If, however, they are not equivalent (ie, there being more red signal or more green signal), that gene is said to be differentially expressed, where the channel with the higher intensity corresponds to the sample in which the gene is more highly expressed. (Note that though this is widely accepted as the simplest explanation of the microarray readout, it also generally regarded as somewhat of an overstatement: differences in transcript levels can also occur as the result of variation in mRNA half-life, and not necessarily because of differences in the amount of gene expression, per se.)

Another popular microarray platform, often referred to as the “single-color” or “single-channel” approach, or simply as the Affymetrix method (after the Affymetrix GeneChip® product), utilizes an alternative strategy: only one sample is assessed per chip. The mRNA from a single sample is reverse transcribed into cDNA which is then used to transcribe and amplify target cRNA. This biotinylated cRNA is then hybridized to the chip (much in the same way that cDNA is hybridized to a chip), and in subsequent labeling reactions, it is bound by the fluorescing molecule phycoerythrin. The array is
then scanned for fluorescent signals in much the same way as previously described, and
the resultant image can be analyzed for probe signal intensities that approximate the
absolute expression levels of the bound mRNA transcripts.

Signal Intensities and Gene Expression Measurements.

Many mathematical and statistical methods for calculating gene expression ratios and/or
absolute transcript levels, and for normalizing these expression measurements within and
across microarray experiments, have been described. Indeed, from the many publications
seeking to rationalize these approaches, one truth has emerged: that there is no one
correct method for calculating expression measurements or for normalizing expression
data. It is therefore not the goal of this section to dissect and explain the different
methodologies in detail (the following chapters will have this honor), but rather to
describe the basic concepts from the point of view of conventional methodologies.

One of the first hurdles in microarray analysis is the decision of how to quantify gene
expression. By convention, gene expression is quantified using fluorescence intensity
measurements taken from a high resolution image. Here, the boundary of a microarray
probe (ie, the spot) is usually well demarcated on its edge by the contrast between the
high foreground signal (ie, within the spot) and the lower background signal (around the
spot; see illustration). Once this boundary is defined, the pixels comprising the
foreground signal are used to calculate a representative signal value, such as the mean or
the median of the pixel intensities. Indeed, exactly which is more correct to use is a
heavily debated issue; however, certain spot-related artifacts commonly observed in two-
color microarrays such as spots with black holes in their centers (resulting from printing-
related damage to the glass surface or in some cases, insufficient probe concentration
during printing) suggest that the median value, or at least a trimmed mean value, would
be most practical. In two-color arrays, this value must be calculated in each of the two
channels from which an expression ratio will be derived. Typically, the ratio of the
means or the ratio of the medians is calculated for this purpose. However, other rational
calculations can be considered: for example, the mean of the ratios or median of the
ratios. In this approach, it is not the totality of pixels in a channel from which the mean
or median is calculated, but rather the signal ratios of all two-channel pairs of pixels.
Currently, there is no one universally agreed upon method, however, the ratio of the
means and ratio of the medians are decidedly the most popular.

In the single-color array platform, absolute expression levels can be calculated in
essentially the same way as described above: by calculating the mean or median signal
for each spot. However, for greater quantitative certainty, it is common on single-color
platforms to have multiple probes representing the same gene and thus comprising what
is called a probe set. This approach requires an additional analytical step: the averaging
of the signal measurements within a probe set. The averaging of signal measurements
across multiple probes representing the same gene is thought to provide a more robust
and accurate measure of gene expression. Furthermore, to gain confidence in the
specificity of the probe measurements, the single-channel arrays such as the Affymetrix
chips often include a mismatch (MM) probe for every perfect match (PM) probe. Here, the MM probe varies from its corresponding PM probe by only the central base position in the oligonucleotide sequence. Because of its central placement (such as the 13th position of a 25-mer probe), a hybridization event between probe and target that is specific for the intended target should result in minimal or no signal on the MM probe, but relatively high signal for the PM probe. In this case, the ratio of PM to MM signal can be applied to the overall quantification of gene expression as a function of the predicted probe-target specificity.

The gene expression measurement of a microarray probe is still further shaped by a process known as background subtraction. Background signal is generally defined as the detectable fluorescence in the vicinity of a probe spot not attributable to the hybridization of target to probe. It is usually measured in a local fashion as the median intensity of the pixels in a defined area of non-probe array surface surrounding a DNA probe spot. It can derive from a variety of sources including slide coating artifacts, the deposition of unincorporated Cy dyes or other fluorescent contaminants during the hybridization, or electronic noise emanating from the array scanner. An underlying assumption regarding background signal is that it fluoresces from the probe spot at an intensity equivalent to that of the surrounding non-probe surface. For this reason, the general practice is to subtract the background signal (in each channel) from the spot’s mean or median signal intensity to generate a background-corrected spot signal intensity. However, many groups have observed that this practice sometimes leads to noisy signal measurements, particularly in the low signal intensity range. In a hybridization artifact commonly observed on spotted arrays and referred to as “antiprobe” or “spot inversion”, the fluorescence signal around a spot can be relatively high, while no signal is detectable on the spot itself, resulting in a black hole-like appearance. This phenomenon suggests that not all types of non-specific fluorescence have an equal impact on DNA spots and non-probe surfaces. Other experimental and analytical results support this hypothesis. As such, some groups have begun opting not to subtract background signal from their spot intensities. This matter remains unresolved in the microarray community.

Once the average signal intensities of the probes have been calculated, the data must be normalized for intra- and inter-array comparisons. For two-color arrays, a ratio calibration is often employed to correct for potential signal imbalances. A popular method for this is the global normalization procedure, whereby the mean or the median of the log expression ratios is set to zero and the difference is used as a correction factor to adjust all log ratios. This approach is based on the assumption that most genes on the array are not differentially expressed, and therefore, the mean or median ratio would best approximate an equivalently expressed gene. However, as this assumption may not hold true in all biological contexts, some camps feel it is more practical to adjust the ratios based on the average or median value derived from a set of housekeeping genes represented on the microarray. Housekeeping genes are so-called because their expression levels are thought to vary minimally from one cell type to another, that is, they are naturally occurring non-differentially expressed genes. However, this method, too, is based on an assumption that may not always hold true: that these genes are expressed at similar levels across all possible biological conditions. Still other
Normalization methods may be preferred that account for the possibility that the distribution of the log ratios may, to some extent, be dependent on the magnitude of the signal intensities. Indeed, it is commonly observed that after global normalization, or normalization using housekeeping genes, probes in the lower range of signal intensity tend to be noisy – showing greater variability in ratio measurements – which likely reflects a higher degree of false positive results. The locally weighted linear regression (LOWESS) curve normalization is perhaps the most popular method for addressing this problem. With the LOWESS transformation, every probe is not necessarily corrected by the same amount, but rather the ratios at all intensities are adjusted to a mean of zero. Finally, local normalization strategies have recently been suggested to address instances of spatial variation on arrays, such as that stemming from print-tip-specific variation or signal intensity gradients owing to pronounced slide curvature or hybridization anomalies. For single-color arrays, the normalization methods are more specialized, and will be discussed in greater detail in the following chapters.

Detecting Differential Expression.

Perhaps the most popular buzzword in the microarray field is “differential expression”. The reason why is because this is what expression microarrays do best: they reveal changes in genes expression – more so than report absolute transcript levels. Though the signal intensity of a probe is, in general, an estimate of the absolute abundance of a gene transcript (and as previously discussed, single-channel platforms are based largely on an approximation of this), this intensity is usually not uniformly linear with respect to actual transcript levels, and for this reason microarray data is widely regarded as semi-quantitative at best. Many studies have compared microarray-derived expression levels or ratios with expression levels/ratios determined by more quantitative mRNA assessments, such as quantitative or real-time PCR techniques. A common finding in these studies is that small changes in gene expression (e.g., 2-4 fold) detected by microarray tend to be more accurate, and higher fold changes (e.g., 5-50 fold) tend to be under-estimates of actual transcript levels. This phenomenon can be partly explained by a lack of sensitivity in detecting transcript levels in the low range of signal. Seldom are microarray probes observed that have a signal intensity of zero, and this is particularly true of the two-color microarrays. Presumably, this is the consequence of low, steady-state levels of nonspecific signal that preclude the accurate detection of very low or absent signal. Thus, when a gene is turned completely “off” in one sample, yet is “on” in the other, instead of resulting in an infinite ratio (e.g., with zero in the denominator in the case of absent signal), the nonspecific signal in the “off” sample results in the calculation of an expression ratio with under-estimated fold change. Consistent with this hypothesis are the repeated observations that higher probe signal intensities well above the level of nonspecific signal (in both channels of a two-color array, or on two single-channel arrays) tend to yield more quantitatively accurate expression ratios. Indeed, in the higher range of signal intensity, highly reproducible differences in gene expression as little as 1.2 fold have been observed. Perhaps surprisingly, this general lack of quantitative accuracy does not discourage most users. Simply put, the ability to detect differential expression, and the directionality of the change, is considered by most to be more
important than knowing the exact transcript abundances. Afterall, who’s to say if a 1.2-fold change in gene expression is any less biologically significant than a 100-fold change? Once differentially expressed genes are identified, questions such as this can be subsequently resolved in the context of more focused and quantitative assays.

So the primary readout of expression microarrays is a measure of differential expression, or put another way, a measure of how the transcript levels of a given gene differ among different biological contexts. Experimentally, there are two fundamental ways in which differential expression can be measured: either by direct comparison or indirect comparison. In direct comparisons (applicable to two-color microarrays), the two RNA samples to be compared are differentially labelled (one with Cy3, the other with Cy5) and hybridized simultaneously to the same array. In this way, the ratio of the (average) signal intensities from the two channels is an expression ratio that directly compares expression levels. An example of when this experimental strategy might be optimally used is a timecourse experiment where the objective is to identify genes that are differentially expressed as a function of time. Here, the samples to be compared might be an untreated sample versus a treated one, or two samples treated at different points in time. Thus, an expression ratio derived from a single experiment reflects the change in expression level between the two biological states. The alternative method for measuring differential expression is the indirect comparison. Here, the two RNA test samples to be compared are each hybridized to separate microarrays together with a common reference RNA sample. In this case, the expression ratios generated on any one microarray are, by themselves, biologically meaningless. However, a comparison of ratios between the two microarray experiments recovers biological meaning: differential expression between the two test RNAs can now be measured as the ratio of their expression ratios. Since the technical variation with in an array experiment is usually less than that observed between array experiments, the direct comparison has somewhat greater accuracy in its quantitation of differential expression. However, the indirect comparison approach has the advantage of allowing any combination of cross-sample comparisons. An example of an experimental situation in which this methodology is advantageous is the study of tumor expression profiles. Here, each tumor sample is hybridized together with a common reference RNA. Thus, gene expression levels can be compared across a continuum of tumor samples, and differential expression can be queried in a variety of biological contexts, such as which genes are differentially expressed between certain clinical tumor classes, or which tumor groups show variable expression of certain select genes. In effect, the experimental method used to assess differential gene expression depends largely on the biological and analytical objectives.

**Evolving Array Technologies.**

Though the DNA microarray was originally designed for gene expression analysis, the flexibility of the technology has allowed its migration into new territories where microarray methods have evolved for interrogating other genome dynamics. For example, one of the earliest spin-off technology was the array-CGH (comparative genomic hybridization). Here, a conventional cDNA microarray was used for an alternate
purpose: to identify genomic copy number changes. Often in cancer and other genetic diseases, the genome becomes unstable and certain chromosomal regions are amplified or deleted. The amplification of oncogenes (ie, genes that promote tumorigenesis) and the deletion of tumor suppressor genes can result in populations of cells with a growth advantage leading to eventual tumor outgrowth. The concept of array CGH is simple: instead of using RNA to generate fluorescent targets, genomic DNA is used, and when hybridized to a cDNA microarray together with a normal reference, the relative copy number ratios for the genomic regions represented by the probes can be quantified. To validate this approach, Pollack et. al., first purified DNA from different cell lines containing various numbers of copies of the X chromosome (from 1 to 5 copies). These DNA samples were then enzymatically fragmented and labelled with fluorophores. Each sample was labelled with Cy5, with the exception of the normal female sample with 2X karyotype which was labeled with Cy3 and used as the reference RNA in subsequent experiments. The 2X sample was then compared to each of the others by conventional two-color cDNA microarray hybridization. By analyzing the fluorescence intensity ratios for all cDNA probes representing genes on chromosome X (n=160), the authors demonstrated the ability of this method to accurately approximate the “dose” of X chromosome in each sample. Next, they examined various cancer cell lines with known copy number changes and demonstrated their ability to detect these changes by microarray. Today, array-CGH is widely used in the study of human cancers for identifying recurring genomic alterations that might pinpoint the location of new cancer susceptibility genes or correlate with other clinically relevant tumor properties[]. In addition to using cDNA probes designed for expression analysis, array-CGH microarrays are often comprised of PCR products or long oligonucleotides representing intergenic sequences, or alternatively, very large contiguous fragments of chromosomal DNA contained within BACs (bacterial artificial chromosomes)[].

Though gene expression microarrays can tell us which genes are turned on and off in different biological contexts, they do not provide information on how genes are mechanistically regulated. With just a few procedural and structural modifications to the existing technology platform, an alternate microarray approach was created to address this problem. Known as chIP-CHIP technology (for chromatin Immuno-Precipitation – chip (ie, DNA chip)), this microarray method was developed for pinpointing precise genomic regions that contain the binding sites for DNA-binding proteins. Transcription factors are DNA-binding proteins that regulate gene expression by binding to specific DNA response elements in gene promoters (ie, gene regulatory regions generally found just upstream of the gene start site). The chIP-CHIP method was designed to gain a genome-wide view of all target genes of a particular transcription factor and thus elucidate the effector molecules associated with its biological function. Procedurally, the technique involves capturing the transcription factor in action – that is, while bound to its target DNA – then isolating the transcription factor/DNA complex using antibodies that bind specifically to the transcription factor (ie, immunoprecipitation). The DNA in this complex can then be fragmented, labelled with a fluorophore, and hybridized to a chIP-CHIP array. On chIP-CHIP arrays, the probes are designed to detect gene promoter regions rather than mRNA transcripts. In this way, the probes that light up on the array indicate which gene promoters the transcription factor of interest interacts with. Similar
to the expression and CGH arrays, the chIP-CHIP arrays can be designed with either cDNA or oligonucleotide probes. Recently, microarrays have been manufactured that contain overlapping oligonucleotide probes tiled across large genomic tracts, or even whole genomes, in some cases. Within the next year, it is estimated that microarray sets (consisting of 20-40 arrays per set) representing the whole human genome will be commercially available for array-CGH or chIP-CHIP applications.

Though the technology for manufacturing whole genome arrays was only recently achieved, the concept of using overlapping oligonucleotide probes to interrogate genomic sequence is actually not new. Beginning in the late 1980’s, methodologies were proposed for using large libraries of short oligonucleotide probes to sequence DNA in a highly rapid and large-scale manner. Today, this technique of sequencing by hybridization has reached its height of optimization in the form of a DNA resequencing array. Resequencing arrays are primarily used to identify single-base mutations or small nucleotide polymorphisms in a highly parallel fashion. These arrays contain short in situ synthesized oligonucleotide probes that overlap each other at 1-base resolution and are derived from a reference sequence such as a gene involved in human disease or the genome of a small pathogen. Array-based resequencing depends on the differential hybridization of genomic fragments (prepared much in the same way as fluorescent target is prepared for array-CGH) to short perfect-match (PM) and corresponding mismatch (MM) oligonucleotide probes. For example, on a platform of 29-mer oligonucleotides, each base to be sequenced (a “T” in this example) is located at the 15th position of a PM 29-mer probe surrounded by its flanking sequence. For each PM probe, 3 MM probes are also included on the array, each possessing one of the three possible alternative bases at the 15th position (“G”, “C”, or “A”). To make a base call at this position, a sequence prediction algorithm evaluates the hybridization signal intensities of the PM probe and its corresponding MM probes, and the one with significantly higher signal intensity determines the correct base at that position. Recent studies suggest that the accuracy of DNA sequence generated in this way can exceed 99.9% across tens of thousands of bases, and that the process can be parallelized on a level adequate for population-based investigations.

Using microarrays to analyze samples on a population scale, however, remains a daunting proposition, largely due to cost constraints. Further advances in microarray manufacture, however, should shorten this hurdle considerable in the coming years. When that time comes, certain microarray applications may leap ahead into the mainstream. One such application in development is a pathogen detection array for use in clinical diagnostics. Recent studies have shown that a microarray composed of probes designed against the most highly conserved genomic regions of viruses, can be used on patient tissue samples (where the nucleic acid has been isolated and PCR amplified to make target) to generate a viral fingerprint that can predict the cause of infection. Indeed, it is already possible to design probes that, in combination, specifically recognize viral families, genera, species, and even highly related strains. This logic can also be extended to bacteria and fungi, and probes for all pathogen types could theoretically be included on the same microarray platform and therefore assayed simultaneously in a single test. Such arrays would take the guesswork out of the current strategy of pathogen diagnostics which generally relies
on a prediction of the likely pathogen involved followed by a subsequent test for the predicted pathogen. Not only would this array be useful in determining which pathogen, or assortment thereof, is the cause of a patient’s illness, but it should also aid in the molecular characterization of newly arising recombinants, as was the case for the SARS coronavirus.

**The Path Ahead.**

In 2004, expression microarrays are being manufactured in hundreds if not thousands of locations, and have been applied to almost every biological niche where some genomic architecture is known. Indeed, at the time of the writing of this book, over 7,000 titles and/or abstracts of articles referenced on the PubMed database at the National Library of Medicine contain the search term “microarray”. Using the same search term on the popular Google search engine results in more than 700,000 hits, suggesting that the existing hype and interest in microarrays is roughly ten times greater than the current level of published applications. So where then are microarrays leading us? Towards larger, more daunting volumes of data, for one. Advances in microarray fabrication technologies are resulting in smaller probe sizes and higher probe densities which translate into more features per array than ever before seen. Competition among commercial providers of microarrays is slowly but steadily driving consumer costs down, thus allowing researchers to increase their usage of microarrays both across and within projects. Moreover, the rapid pace at which whole genomes are being sequenced these days perpetuates an ever-expanding market of new microarray platforms. To cope with the rising tsunami of data and to preserve its biological meaning, public database repositories are being established online, and top journals are making data submission to these repositories mandatory for publication to ensure that the data is sufficiently annotated and standardized for optimal public access and future re-analysis.

The collective thinking in the microarray community is cohering, achieving unity under common goals and necessities. We are moving into a more probe-conscious era, designing and re-designing algorithms for selecting gene probes with the highest specificity and hybridization performance. We are realizing the need for probes capable of distinguishing the various alternative splice variants that exist for most genes, as well as for detecting new biologically-active RNA molecules such as noncoding and antisense RNAs. The need to better understand how these new molecules, as well as known and newly discovered genes, interact in biological space is making us more pathway-conscious. Databases have been developed that seek to place all genes and their encoded proteins into functional classes that define their involvement in biochemical pathways and cellular processes. Interfaces for these databases are now being built that allow for the rapid extraction of this biological information and its immediate application to microarray results. To fully apply the data gathered from animal models of human disease will require carefully annotated maps of genes and proteins that are structurally and functionally conserved across species, such as a database that indexes paired human and mouse orthologous genes. Gene expression signatures are currently being discovered that predict for the tissue of origin of human tumors, clinical outcomes such as cancer
recurrence and patient survival, or the operational configuration of important biochemical pathways, but these require rigorous validation in independent datasets before being put to practice. Tools for studying the behavior of these genes across different microarray datasets and different microarray platforms will be needed to facilitate these validation studies. And of course, there will be increasingly greater analytical challenges whose solutions cannot be achieved by mere biologists or strict mathematicians, but rather by professionals trained in both disciplines and thus capable of interrogating biology as a living system. Currently, biologists and mathematicians working synergistically, like the authors of this book, are leading that charge.

DNA microarrays are indeed a portal to a new frontier in molecular science— the *living* genome. They afford us the opportunity to observe the genome in motion, and this allows us to witness, for the first time, the underlying logic and order inherent in biological systems. Given that the current prediction for the number of genes in the human genome is lower than previously thought, in the range of 30,000, the entire variance in the human phenotype rests in how these 30,000 units are expressed. Our explorations into the genome, limited as they have been, are already changing the way we view the cell and the organism. We have begun to visualize life itself as a dynamic continuum of genomic and proteomic reprogramming, and microarrays help us to see, dissect, and predict this continuum. The technology itself is flexible, and therefore it can be used to address both broad and specific biological questions. It can be applied in a holistic manner to characterize the state of a cell or an organism, or in a focused manner to provide context for understanding the activity of a molecule or a pathway. Microarrays are not platform-restricted tools. They can be used in both prokaryotic and eukaryotic systems, and the only limitation appears to be the current paucity of available genomic sequence information for more complex organisms. The expression data generated by microarray analysis also is not platform-restricted. Studies that have demonstrated the evolutionary conservation of function between orthologous proteins support the notion that gene expression behavior in one animal system may be recapitulated in another. The path ahead is indeed an exciting one, and the chapters that follow will illuminate important current issues in data analysis that just perhaps will guide us well as we delve yet deeper into the genomic fabric of life in search of biological answers.