Preface

Dramatic technological advances have marked every quantum leap in our understanding of biological systems. Advances in manipulating and sequencing DNA triggered the last such leap. More than 20 years ago this technology began finding its way into biology laboratories, and multiple landmark discoveries have followed, including the sequencing of the genomes of several prokaryotic and eukaryotic species. This, in turn, spawned a new interest in bioinformatics, structural biology, and high-throughput methods that would allow scientists to look at the response of the entire genome to physiological, pharmacological, and pathological changes. Whether this transition was driven by exhaustion of the existing hypothesis pool, a subconscious adoption of the old "new paradigm" of biological complexity, or an instinctual urge among biologists to look for new and interesting phenomena is not really important. What is important is that high-throughput methods are becoming more and more routine and available, and experimentalists and theoreticians must be prepared to take advantage of them. Spotted DNA microarrays fall into this category. They power functional genomics, a nascent research field dealing with the structure and activity of genomes and global relationships between genotype and phenotype. The birth of the field can be traced back to three seminal papers ***(2,3,12)***. These three works grew from the realization that by placing individual sequence elements on a solid surface one can probe by hybridization a nearly unlimited number of targets simultaneously. Since then, similar *ad infinitum* approaches have been used to monitor relative protein levels ***(6)***, protein functions ***(15)***, cellular activities ***(16)***, and molecular interactions ***(10)***. These breakthrough studies will set the stage for the development of the fields of functional proteomics, metabolomics, etc. However, at the moment only functional genomic techniques on solid surfaces enjoy relatively wide acceptance because they are based on sounder physical-chemical principles.

The long-term impact of functional genomics will depend on multiple factors— standardization and simplification of the protocols used, robust error assessment, streamlining of the analytical techniques, the sustainability of cost. The goal of the volume is to familiarize its readers with available, reproducible protocols in the field, and to attempt to introduce an audience of biologists to data processing techniques that will become critically important as we start dealing with increasing quantities of information. The "Methods" are divided in two sections: (i) Methods in Data Generation and (ii) Methods in Data Analysis. The first section focuses on bench techniques that have been

developed and are being routinely used in several hard-core genomics laboratories. Besides general applicability of the techniques, the articles represented in the first section were selected on the basis of one major criterion: that they give sufficiently robust protocols to be adopted without modification by workers who have just begun their journeys in the field of genomics. The section opens with articles describing ways to manufacture and use spotted microarrays on three different solid surfaces: glass, plastic, and nylon membranes. Arrays manufactured on glass surfaces are usually interrogated with fluorescent nucleic acid probes, and Chapter 4 describes an optimized RNA labeling procedure that is applicable to the known spectrum of RNA sources. This chapter is followed by articles dealing with issues and protocols that are common to the field of bacterial functional genomics. The last two years' work in the field of functional genomics was marked by development of specialized applications that have added to its depth. In this period there were techniques introduced that allow one to monitor subcellular RNA localization *in masse* ***(4,9,14)***, to map chromosomes at the resolution of a single gene ***(8,13)***, and to survey the steady-state genome-wide distribution of DNA binding proteins in vivo ***(5,7,11)***. Chapters 6–8 deal primarily with the methodologies behind these advances; Chapter 7 also provides a link between expression profiling and determination of gene copy number using whole-genome DNA microarrays.

The issues of inference, experimental design, and reproducibility are of the paramount importance to researchers who deal with massive data sets. The second section of the "Methods" volume focuses on experimental design, data analysis, data display techniques, and bioinformatics. The section opens with a comprehensive overview of the inferential issues in microarray data analysis. The next four articles (Chapters 10–13) address in sequential manner some of the issues outlined in the overview article ***(9)***: design of microarray experiments ***(10)***; choice of the test statistic and assessment of the significance of observations ***(11)***; data reduction ***(12)***, and clustering, the most popular technique for microarray data classification ***(13)***. Accelerating and making the most of functional genomics studies are impossible without visualization and data storage, both of which—like the remaining methods in the field—are still in their infancy. These problems cannot be overlooked, however, because they allow genomics researchers and their colleagues in the scientific community to examine and mine experimental results. Two articles (***14*** and ***15***) describe some available approaches to the data visualization and database related issues. There are several things that were not reflected in any of the featured articles but are nevertheless worth noting. Spotted DNA microarrays are currently being used mainly for two purposes: 1. screening and 2. modeling. Truly successful application of microarrays in these two areas depends, albeit to different degrees, on technology standardization as well as the free, unimpeded flow

of experimental data. Although we believe that methods will become fairly standard in the near future, a good deal of useful and valuable information will be lost in the short run owing to the lack of enforceable standards and controlled vocabularies for experimental annotation. The Minimum Information About a Microarray Experiment (MIAME) specifically addresses this issue ***(1)***, and should be read by anyone who wants to engage in expression profiling.

The "methods" we have compiled do not provide advice about or comparisons of the available spotting platforms, image extraction and analysis algorithms, data storage and retrieval devices, and data analysis compendia. Although the available options range from simple, relatively affordable solutions to high-end, sometimes extremely expensive, commercial ones, we believe that information accumulated in the field is not yet sufficient and/or systematic enough to provide comprehensive comparisons of individual solutions and/or specific recommendations.

In the course of preparing this volume we surveyed available microarray web resources. We encourage readers interested in developments in the field to keep an eye on the following sites:

http://www.microarrays.org
http://derisilab.ucsf.edu/
http://www.mged.org/Workgroups/MIAME/miame.html
http://www.bioconductor.org/
http://ihome.cuhk.edu.hk/~b400559

Michael J. Brownstein
Arkady B. Khodursky

References

1. Brazma, A., P. Hingamp, Quackenbush, J., et al. (2001) Minimum information about a microarray experiment (MIAME)–toward standards for microarray data. *Nat. Genet.* **29,** 365–371.
2. Chuang, S. E., Daniels, D. L., and Blattner, F. R. (1993) Global regulation of gene expression in *Escherichia coli. J. Bacteriol.* **175,** 2026–2036.
3. DeRisi, J. L., Iyer, V. R., and Brown, P. O. (1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278,** 680–686.
4. Diehn, M., Eisen, M. B., Botstein, D., and Brown, P. O. (2000) Large-scale identification of secreted and membrane-associated gene products using DNA microarrays. *Nat. Genet.* **25,** 58–62.
5. Gerton, J. L., DeRisi, J., Shroff, R., Lichten, M., Brown, P. O., and Petes, T. D. (2000) Inaugural article: global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **97,** 11383–11390.

6. Haab, B. B., Dunham, M. J., and Brown, P. O. (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.* **2,** RESEARCH0004.
7. 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,** 533–538.
8. Khodursky, A. B., Peter, B. J., Schmid, M. B., et al. (2000) Analysis of topoisomerase function in bacterial replication fork movement: use of DNA microarrays. *Proc. Natl. Acad. Sci. USA* **97,** 9419–9424.
9. Kuhn, K. M., DeRisi, J. L., Brown, P. O., and Sarnow, P. (2001) Global and specific translational regulation in the genomic response of *Saccharomyces cerevisiae* to a rapid transfer from a fermentable to a nonfermentable carbon source. *Mol. Cell Biol.* **21,** 916–927.
10. Kuruvilla, F. G., Shamji, A. F., Sternson, S. M., Hergenrother, P. J., and Schreiber, S. L. (2002) Dissecting glucose signalling with diversity-oriented synthesis and small-molecule microarrays. *Nature* **416,** 653–657.
11. Lieb, J. D., Liu, X., Botstein, D., and Brown, P. O. (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. *Nat. Genet.* **28,** 327–334.
12. Nelson, S. F., McCusker, J. H., Sander, M. A., Kee, Y., Modrich, P., and Brown, P. O. (1993) Genomic mismatch scanning: a new approach to genetic linkage mapping. *Nat. Genet.* **4,** 11–18.
13. Pollack, J. R., Perou, C. M., Alizadeh, A. A., et al. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23,** 41–46.
14. Takizawa, P. A., DeRisi, J. L., Wilhelm, J. E., and Vale, R. D. (2000) Plasma membrane compartmentalization in yeast by messenger RNA transport and a septin diffusion barrier. *Science* **290,** 341–344.
15. Zhu, H., Klemic, J. F., Chang, S., et al. (2000) Analysis of yeast protein kinases using protein chips. *Nat. Genet.* **26,** 283–289.
16. Ziauddin, J. and Sabatini, D. M. (2001) Microarrays of cells expressing defined cDNAs. *Nature* **411,** 107–110.