现代生物技术前沿

MICROARRAY ANALYSIS

〔美〕M. 谢纳 著

生物芯片分析

（影印版）

科学出版社
www.sciencep.com

现代生物技术前沿

MICROARRAY ANALYSIS

〔美〕M. 谢纳 著

生物芯片分析

（影印版）

To my wife, family, friends, mentors, and humankind.

科学出版社
北京

图字 01－2003－5896

内 容 简 介

在生物学和医学前沿领域中，基因组测序、高通量药物筛选、DNA芯片（微阵）等技术应用越来越广泛。使用给定细胞中的mRNA，在给定的时间及条件下，DNA芯片能够快速检测出该细胞中的基因表达水平，可以研究基础的生物学现象（如进化、发育）、判断新基因的功能、检测药物反应及其他基因表达组成等。本书是对生物芯片（微阵）技术进行全面论述的专业性著作，对概念及理论基础进行了较为透彻的阐释，涵盖范围从生物化学原理一直到方法学的讨论以及研究中所用的基因数据的分析，内容涉及生物芯片技术相关信息、基因表达、遗传聚类以及新创建的芯片技术，芯片的商品化应用和芯片技术在临床诊断中的应用也有涉及。

本书内容较为全面，使相关领域的研究者对相关知识容易领会并应用，兼具理论性和实用性。适于从事生物芯片研发应用以及分子生物学、生物化学、细胞生物学、免疫学、细胞生物学、生物信息学、生物技术、生物工程、蛋白质组学、基因组学等生命科学相关研究领域的教学科研人员参考使用，也可用作教材和教辅读物。

图书在版编目(CIP)数据

生物芯片分析/(美)谢纳(Schena, M.)著. —影印版. —北京：科学出版社，2003.10

(现代生物技术前沿)

ISBN 7-03-012247-X

Ⅰ.生… Ⅱ.谢… Ⅲ.生物-芯片-分析-英文 Ⅳ.Q78

中国版本图书馆CIP数据核字(2003)第085869号

责任编辑：莫结胜

责任印制：刘士平／封面设计：王 浩

科学出版社 出版

北京东黄城根北街16号

邮政编码：100717

http://www.sciencep.com

新蕾印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2003年10月第 一 版 开本：890×1240 1/16

2003年10月第一次印刷 印张：40 插页：8

印数：1—3 000 字数：1 333 000

定价：80.00元

（如有印装质量问题，我社负责调换〈环伟〉）

Preface

Some painters transform the sun into a yellow spot; others transform a yellow spot into the sun. —*Pablo Picasso*

What does not destroy me, makes me strong. —*Friedrich Nietzsche*

Why are we here? What is the meaning of life? What happens when we die? Does God exist? This textbook is unlikely to provide immediate answers to these fundamental questions, but it does promise a broad introduction to microarray analysis, which may actually seem more like a punishment than a primer once you study its contents. But seriously...please understand that a scientist is no more equipped to spend 18 months (1.9 millions) writing a textbook than a writer would be if he or she were required to spend 18 months in a laboratory. This preface endeavors to direct, though admittedly it also conveys some of the delirium generated at the hands of a most challenging and demanding project.

I was warned amply about Ron Davis before departing the University of California at San Francisco for Stanford, and I heeded those warnings for several years before falling. But what started as biology, ended up technology, and it is the latter that is the main focus of this textbook. In a brief conversation in a Stanford café in 1990, Ron said in no uncertain terms (and with all the prescience of the French astronomer Nostradamus) that he thought I should develop a "major new technology," and I suppose that is exactly what happened, albeit several years after his suggestion and not necessarily the technology that either of us (certainly not I) had envisioned. Ron and I began our odyssey together exploring the function of transcription factors in the flowering plant *Arabidopsis thaliana* and, through this endeavor, came to realize that a new gene-expression technology based on solid-surface assays would revolutionize the medical and agricultural sciences. I'm not sure either of us understood the magnitude of what we were embarking on, and perhaps neither of us (again, certainly not I) can anticipate the full scope of discoveries that will be made with this technological advance.

There is always a "eureka" moment in research, a term that is even more appropriate in this context because *Eureka* is the motto of California. During a discussion in Ron's office, he told me about "a company in Santa Clara" that was developing high-density arrays of oligonucleotides for gene sequencing

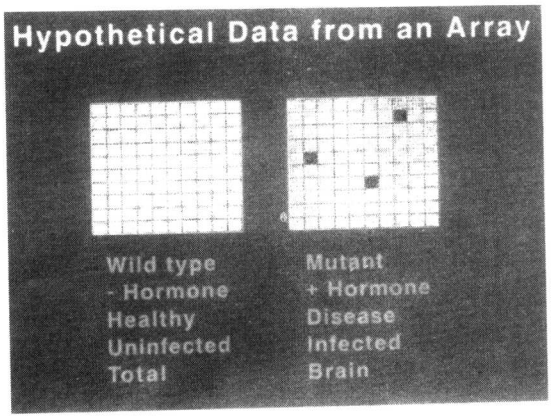


FIGURE P.1. *Microarray analysis concept. Microarrays can be used to profile gene expression patterns in labeled mixtures prepared from reference* (left) *and test* (right) *samples, yielding information on genotype, hormone action, disease state, infection, and tissue-specific gene expression. (Reprinted with permission from M. Schena, 1994. Data courtesy of M. Schena and R. Davis, Stanford University.)*

applications. After seeing some unpublished Affymetrix data, the idea of using microarrays for gene expression studies hit me with the force of a thunderbolt. Instead of using short oligonucleotide arrays for sequencing by hybridization, it seemed much more powerful to use microarrays of long oligonucleotides for gene expression analysis. Eureka!

We began our work on microarrays at a time when there was no evidence that biological experiments could be performed on glass chips. The microarray concept was presented in Holland in the summer of 1994 amid (I kid you not) howling laughter from the audience; and it was on this occasion that I outlined the basic experimental approach that would come to be known as microarray analysis (Fig. P.1). During that presentation, I also introduced the first microarray enzymatic labeling procedure, demonstrating the feasibility of preparing fluorescent probes from yeast and plant messenger RNA (Fig. P.2).

As a proof of principle study, we manufactured the first microarrays in collaboration with Affymetrix and used fluorescent oligonucleotide probes to demonstrate the specificity of the microarray assay (Fig. P.3). Hybridization experiments with complex gene expression mixtures yielded promising data, although the hybridization specificity was somewhat compromised owing to a suboptimal probe preparation procedure (Fig. P.4). Improved hybridization

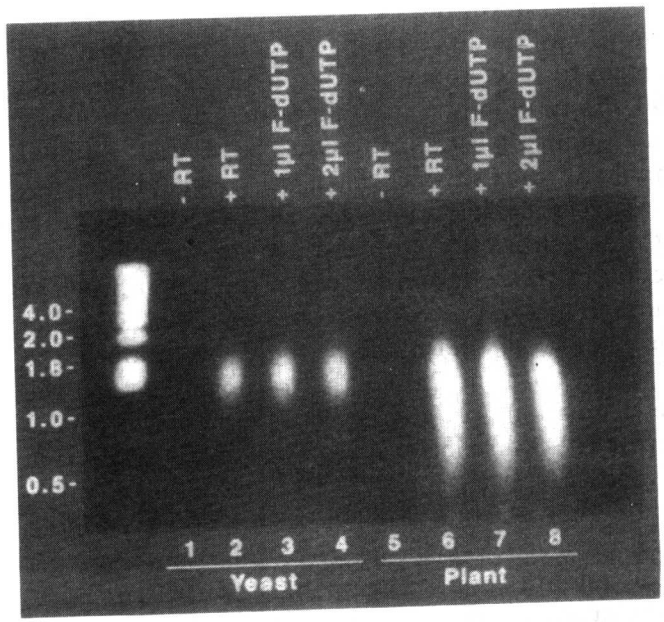


FIGURE P.2. *Data from the first microarray labeling procedure. Total messenger RNA samples from yeast and* Arabidopsis *were primed with oligo-dT, and labeled with reverse transcriptase in the presence of fluorescein-dUTP. The labeled products were separated by agarose gel electrophoresis and stained with ethidium bromide. (Reprinted with permission from M. Schena, 1994. Data courtesy of M. Schena and R. Davis, Stanford University.)*

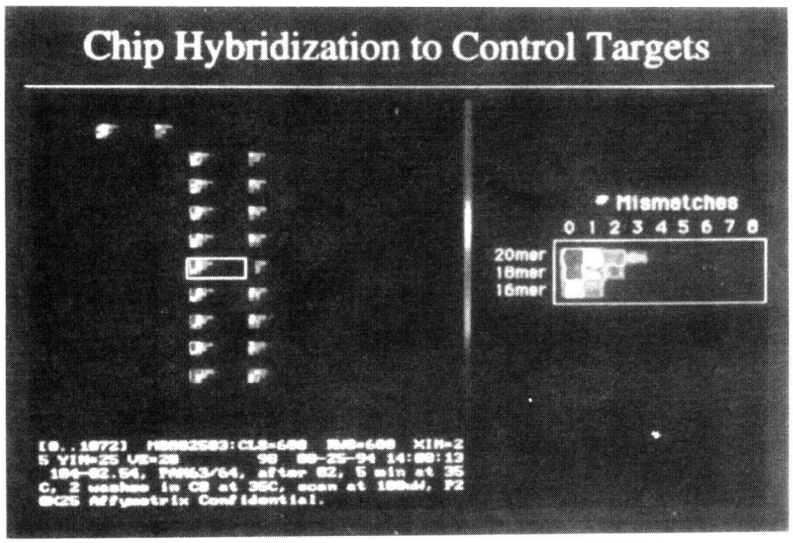


FIGURE P.3. *Data from the first manufactured microarray, which was by Affymetrix and used very large scale immobilized polymer synthesis (VLSIPS) technology. It was hybridized with two fluorescent oligonucleotides (PAN63 and PAN64) to assess assay specificity and detectivity. A small microarray region (white box) is enlarged for ease of viewing (right). (Reprinted with permission from M. Schena, 1994. Data courtesy of M. Schena and R. Davis, Stanford University.)*

specificity was obtained using complementary DNA microarrays in a second collaboration, this time with Dari Shalon and Patrick Brown at Stanford (Fig. P.5). A microarray robot was used for complementary DNA microarray manufacture and, as expected, these microarrays yielded unambiguous gene expression information for several different lines of *Arabidopsis*. The data from these experiments were presented at the University of Wisconsin at Madison in the summer of 1995 (Fig. P.6). Printed microarrays were also employed for the first human microarray analysis experiments, in a collaboration with Synteni (Palo Alto, CA). This work was undertaken in spite of warnings from some luminaries in the field who noted that repetitive sequences in the human genome would prevent the use of microarray assays for human studies (huh?). The first human microarray data were presented at the Stanford Sierra Retreat in October 1995 (Fig. P.7).

The field has blossomed immensely since the first paper on microarrays and microarray analysis appeared in *Science* magazine in the fall of 1995 (Schena

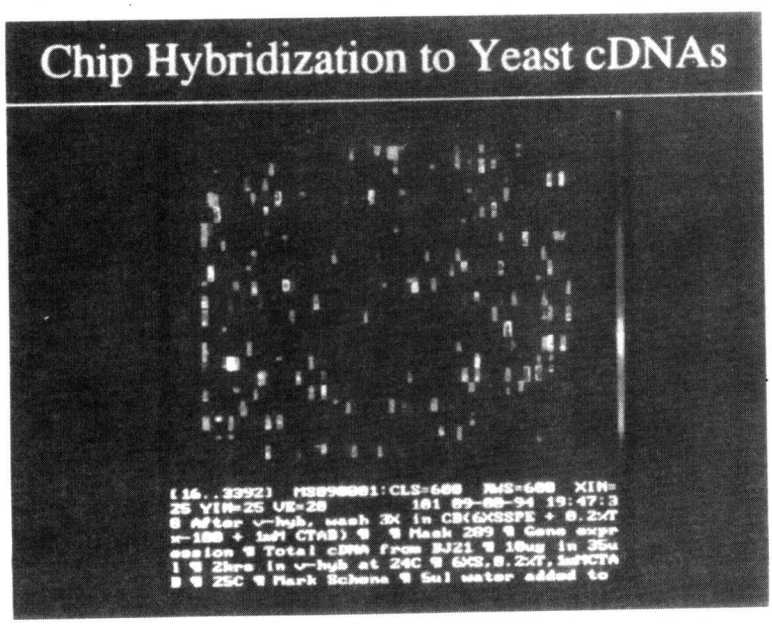


FIGURE P.4. *First microarray analysis experiment. Shown are data from the first microarray analysis experiment. The yeast oligonucleotide microarray was manufactured at Affymetrix (Santa Clara, CA) using VLSIPS technology, and hybridized under a cover slip using a fluorescent probe mixture prepared by reverse transcription of yeast total mRNA. Data were generated by Mark Schena on September 9, 1994, and presented on October 5, 1994 at the Stanford Sierra Camp. Data were provided courtesy of Mark Schena and Ron Davis (Stanford University).*

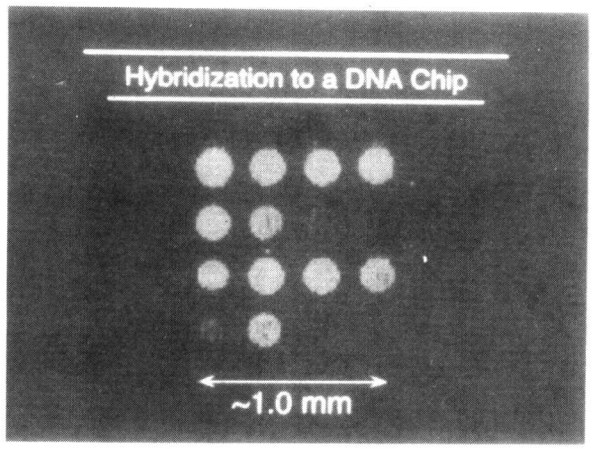


FIGURE P.5. *Data from the first printed complementary DNA (cDNA) microarray, which was manufactured in the laboratory of P. Brown using a tweezer-based printing robot that deposited small droplets of amplified* Arabidopsis *cDNA onto a treated glass microscope slide. The microarray was hybridized with fluorescent cDNA to assess the specificity and detectivity of the assay. (Reprinted with permission from M. Schena and D. Shalon, 1995. Data courtesy of M. Schena and R. Davis, Stanford University.)*

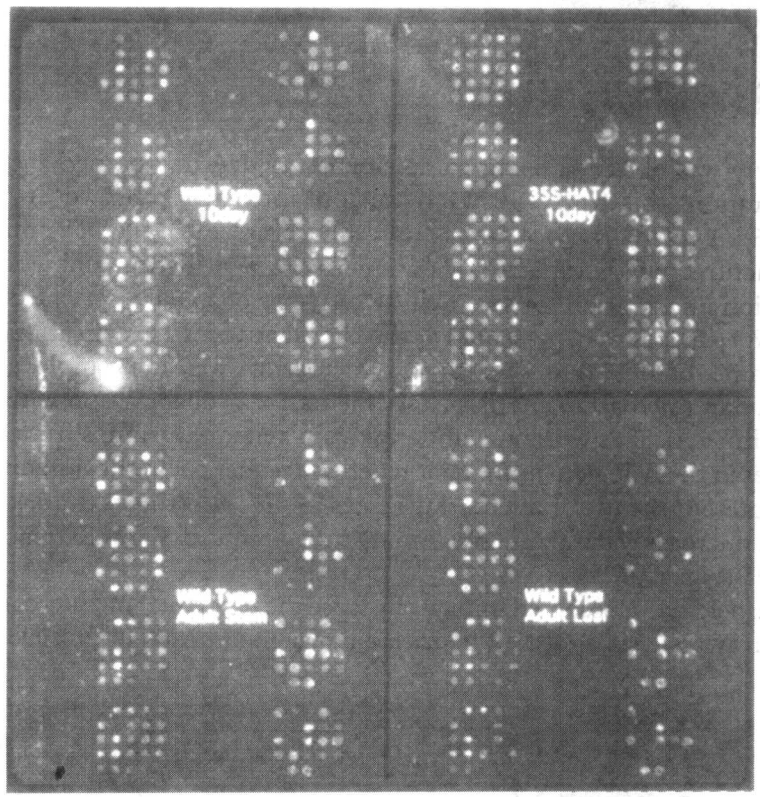


FIGURE P.6. *Data from the first complementary DNA (cDNA) microarray analysis experiment. An* Arabidopsis *cDNA microarray was hybridized with fluorescent cDNA mixtures prepared from four different sources of* Arabidopsis *messenger RNA. The experiments were performed to profile gene expression patterns in a wild-type line, a transgenic line overexpressing the* HAT4 *cDNA, wild-type stem tissue, and wild-type leaf tissue. (Reprinted with permission from M. Schena and D. Shalon, 1995. Data courtesy of M. Schena and R. Davis, Stanford University.)*

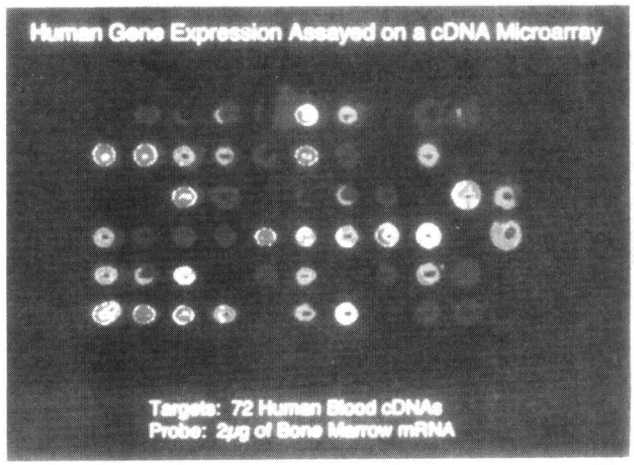


FIGURE P.7. *Data from the first human microarray manufactured by printing 72 amplified complementary DNA (cDNA) inserts from a human blood cDNA library onto a glass substrate. Microarray analysis was performed by hybridizing a fluorescent cDNA mixture labeled from bone-marrow messenger RNA and reading the fluorescent signals, which were coded to a rainbow palette. (Reprinted with permission from M. Schena 1995. Data courtesy of M. Schena and R. Davis, Stanford University.)*

et al., 1995). Assays now encompass a remarkable breadth of organisms, including human, yeast, mouse, rat, chimp, gorilla, fruit fly, worm, corn, rice, and bacteria, and enjoys an equally impressive breadth of applications, including genotyping, tissue analysis, and protein studies. More than 3000 scientific publications showcase a broadening repertoire of microarray assays (see References), and this wealth of published work emphasizes the richness of the technology, the collegiality of the field, and the abundant availability of commercial resources.

The creation of *Microarray Analysis* was motivated by the explosive proliferation of the technology, the highly technical nature of the field, and the flood of requests from young scientists for a foundational compendium. The textbook was also motivated by a clear need to assemble a resource aimed at complementing the specialty books and primary scientific literature in the field, and by my good friend Tom Tisone. I have been compensated enormously over the years for my technical expertise, and the notion of giving something back to the scientific community at this juncture seemed entirely reasonable, logistically possible, and probably a bit overdue.

And so when Luna Han, life and medical sciences editor at John Wiley & Sons, contacted my publicist Paul "Hollywood" Haje to initiate *Microarray Analysis,* I jumped at the opportunity and an agreement was hammered out quickly. I slated six months to write the textbook, and the project took almost exactly three times that long, a remarkably poor estimate for someone with a Ph.D. in biochemistry. In my defense, let me assure you that no amount of grooming by the best mentors at the top biochemistry departments in the country could prepare a biochemist for writing a 1500-page science manuscript. I joked several times with Luna that I came to consider myself a POW (Prisoner of Wiley), but in retrospect that title turned out to be less than accurate because real POWs actually enjoy somewhat better treatment than textbook writers! With all kidding aside, the project would not have succeeded without Luna's constant support and guidance, and dearest thanks are due to Luna and the entire staff at Wiley for making *Microarray Analysis* a success.

Microarray Analysis is written for a broad audience of undergraduates, graduate students, postdoctoral fellows, faculty members, and deans as well as researchers, clinicians, investors, lawyers, and businesspeople from universities, companies, hospitals, government agencies, and nonprofit organizations. The book is intended to provide a conceptual, experimental, and methodological foundation for the full spectrum of activities in modern microarray analysis. Chapter 1 introduces the field, and Chapters 2–4 cover basic concepts in chemistry, biochemistry, and genomics pertinent to microarray analysis. Chapters 5–9 explore microarray surfaces, targets and probes, manufacturing, detection, and data analysis and modeling. Chapters 10–13 explain microarray methodology, cleanroom technology, gene expression profiling, and genotyping and diagnostics. Separate chapters are devoted to novel technologies (Chapter 14), commercial opportunities (Chapter 15), and future trends (Chapter 16). Each chapter also contains a set of questions intended for use as a teaching aid, so that university courses, technical workshops, and adult-education programs can be designed around *Microarray Analysis.* The continued expansion and success of the field is predicated on having highly educated students and scientists, and it is my hope that *Microarray Analysis* will assist in teaching a new generation of microarray researchers, including our youngest and most precocious students (Fig. P.8).

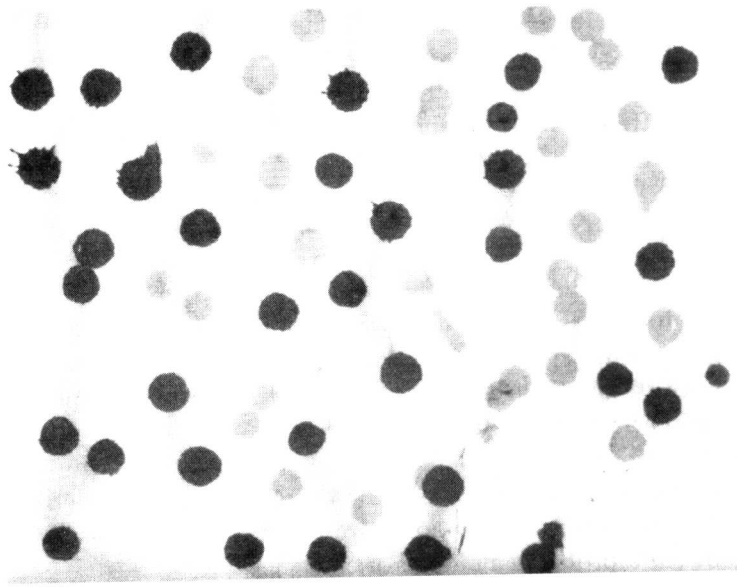


FIGURE P.8. *A first attempt at microarray manufacture from the youngest known microarray researcher. (Data courtesy of Julia Kaplan, age 5, San Carlos, CA.)*

Have we reached a turning point in our civilization? Since our earliest days as humans on the plains of Africa, genetic and infectious diseases have ravaged our communities. What is disease? Disease is little more than the disruption of normal cellular gene function, and microarrays can be used to rapidly understand disease states by enabling the analysis of gene expression patterns, sequence variation, and other biochemical processes. Fifty years from now, and long after human disease has been eradicated, we will look back incredulously at the start of this millennium and wonder how we ever endured cancer, heart disease, AIDS, and the thousands of other illnesses that compromise our well-being. But that will be then, and this is now. Let's get to work!

Mark Schena

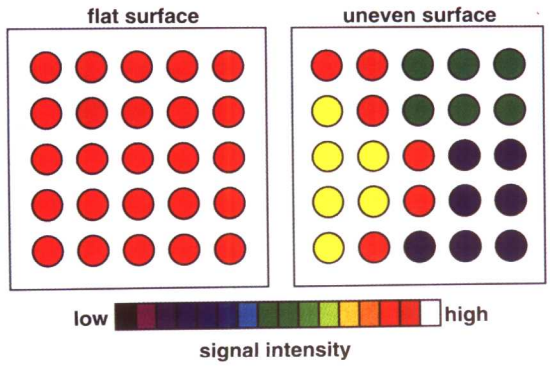


FIGURE 5.2. Because microarray detectors have a limited depth of focus, a hypothetical microarray of identical samples formed on a flat surface will produce a uniform image, whereas the same microarray formed on an uneven surface will produce an image that has variations in signal intensity.

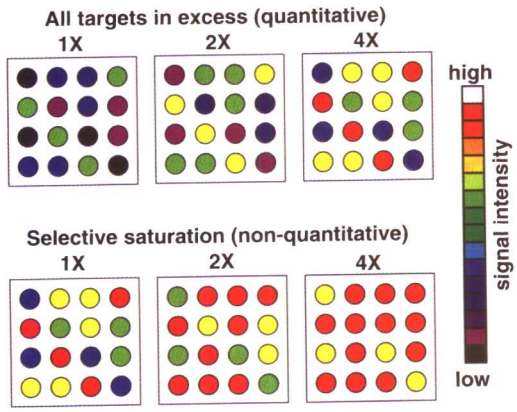


FIGURE 5.11. Top, *When all the targets are present in excess of probe, successive twofold increases in probe concentration produce successive twofold increases in signal intensity at each microarray location.* ***Bottom,*** *Under conditions of probe excess, there is a selective saturation of targets corresponding to abundant species, so that successive twofold increases in probe concentration produce twofold increases in signal intensity at only some microarray locations.*

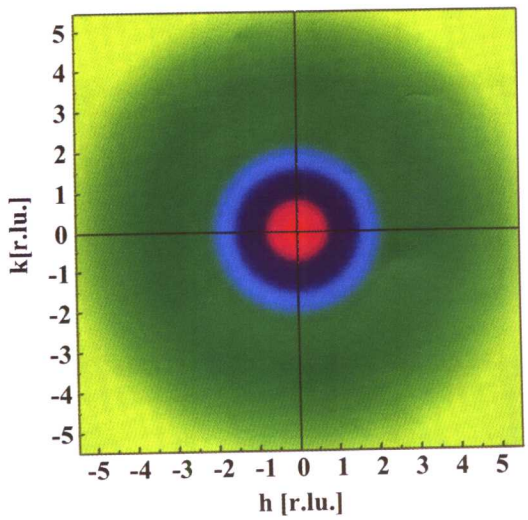


FIGURE 5.13. *An x-ray diffraction pattern of a silicon atom, one of the two main atomic components of glass. Electron densities are represented in a gray scale. (Data courtesy of T. Proffen and R. B. Neder, University of Wurzburg, Germany.)*

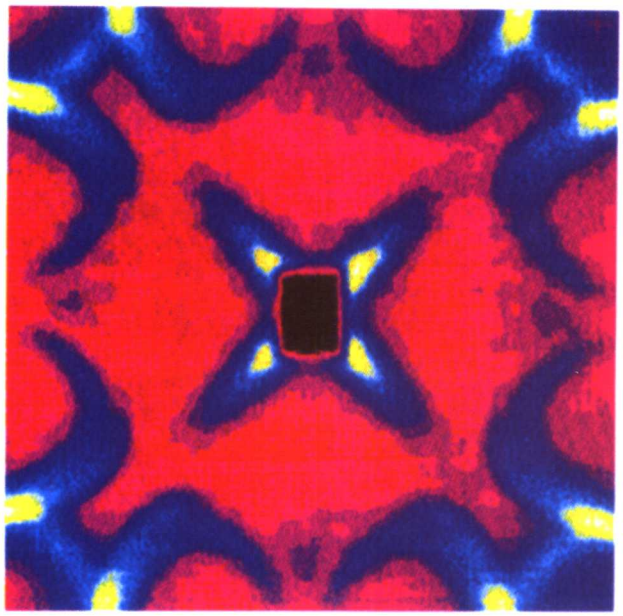


FIGURE 5.14. *SANS of silicon dioxide (SiO_2), the principle component of glass. A single crystal of silicon was heated for 500 h at 600°C, and trace oxygen (~ 30 ppm) diffused to form SiO_2 precipitates. (Data courtesy of ILL, France.)*

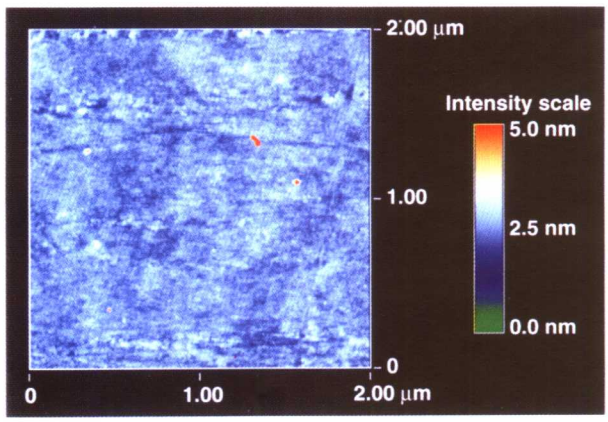


FIGURE 5.15. *An optically flat glass substrate was examined by AFM over an area of 4 μm^2. The roughness image is displayed in a rainbow intensity scale, and the mean flatness is ~2.1 nm. (Data courtesy of TeleChem/ArrayIt.com, Sunnyvale, CA.)*

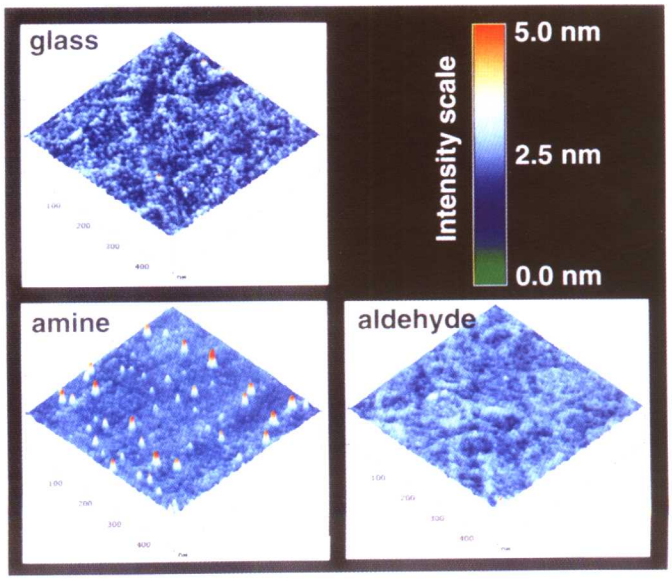


FIGURE 5.18. *AFM was used to examine three surfaces over an area of 0.25 μm^2. Data for optically flat glass, optically flat glass with reactive amine groups, and optically flat glass with reactive aldehyde groups are represented in a rainbow intensity scale. (Data courtesy of TeleChem/ArrayIt.com, Sunnyvale, CA.)*

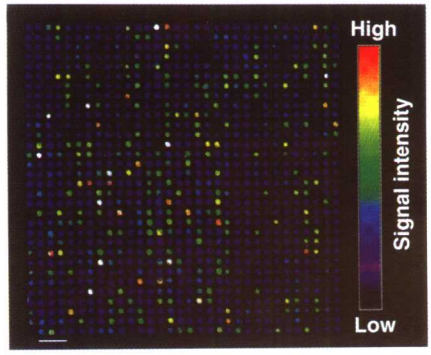


FIGURE 5.21. *Target cDNAs corresponding to 1024 human genes were printed on a reactive amine surface and hybridized to fluorescent probes prepared from human liver RNA. The scanned data are represented in a rainbow color palette. Space bar = 500μm. (Data courtesy of Y. Li, Fudan University, Shanghai, People's Republic of China.)*

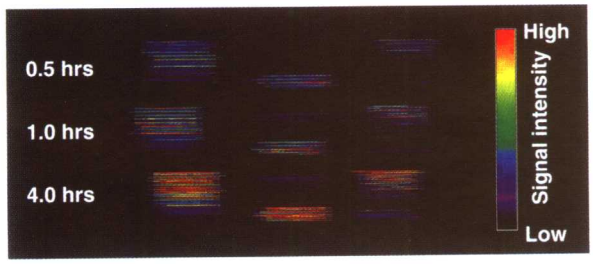


FIGURE 5.25. *Oligonucleotides (15mers) containing amino linkers were attached to an aldehyde surface and hybridized for different lengths of time (0.5–4 h) with a mixture of fluorescent oligonucleotides (15mers) at a concentration of 2 μM for each microarray. After hybridization, the chips were washed to remove unbound probe and scanned for Cy3 emission. Fluorescence intensities are represented in a rainbow scale. Because the hybridization was carried out under conditions of target excess, signal intensities increased as a function of time.*

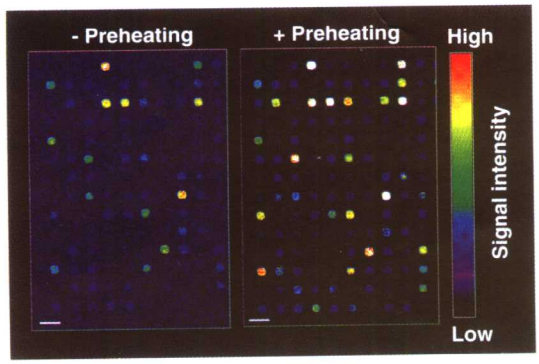


FIGURE 5.26. *Two microarrays were hybridized with fluorescent probes prepared from 3 μg of total* Arabidopsis *RNA and scanned for Cy3 emission. Image data are displayed in rainbow palette. Hybridizations were initiated by adding room-temperature probe to a room-temperature aldehyde microarray (−Preheating) or a 55°C probe to a 55°C aldehyde microarray (+Preheating). Space bar = 250 μm. The materials used include a Submicro Expression Array detection kit (Genisphere, Montvale, NJ), GlassHyb buffer (Clontech, Palo Alto, CA), and SuperAldehyde substrates (TeleChem/ArrayIt.com, Sunnyvale, CA). (Data courtesy of S. Ruuska, Michigan State University, East Lansing, MI.)*

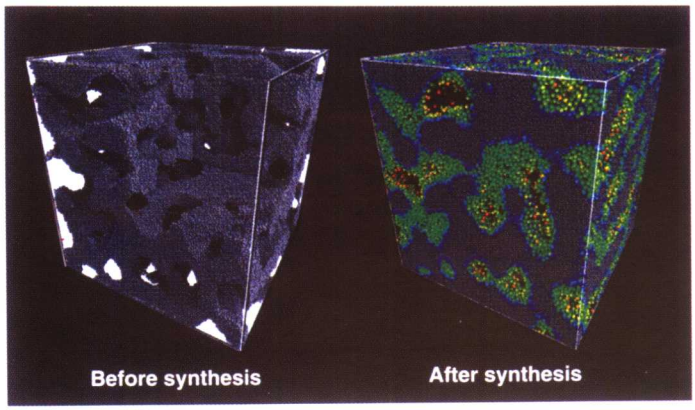
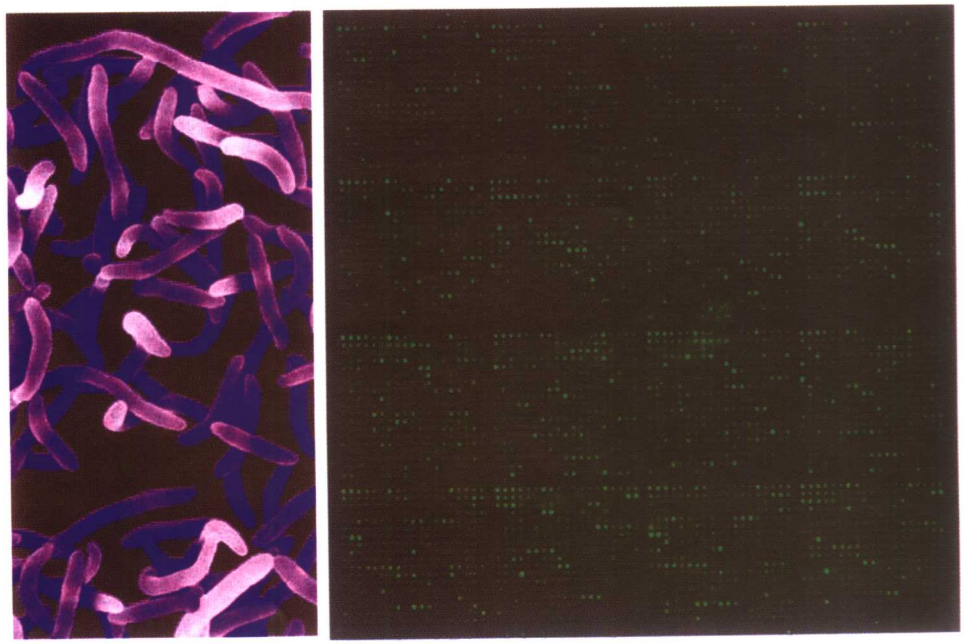


FIGURE 6.12. *Computer-generated models of CPG simulating the oligonucleotide synthesis matrix before and after oligonucleotide synthesis. The first DNA base is bound directly to silicon dioxide atoms* (gray matrix). *During the synthesis process, phosphoramidite reagents* (white spheres) *enter the porous CPG and become incorporated into the growing oligonucleotide chains.(Courtesy of Dr. L. Gelb, Department of Chemistry, Florida State University, Tallahassee.)*



(A) (B)

FIGURE 6.21. **A,** *Photomicrograph of the bacterium* V. cholera. *(Courtesy of Dennis Kunkel Microscopy, Kailua, HI.)* **B,** *Fluorescent image of a microarray containing the entire genome of* V. cholera. *(Courtesy of K. Chong, Schoolnik Laboratory at Stanford University, Stanford, CA.) Total RNA from* V. cholera *was labeled with reverse transcriptase in the presence of Alexa546-dUTP, and the direct-labeled probe was hybridized to the microarray and scanned with a ScanArray 5000 (Packard Biochip Technologies, Billerica, MA). The fluorescent TIF data are coded to a monochromatic green look-up table. (Photomicrograph was provided)*

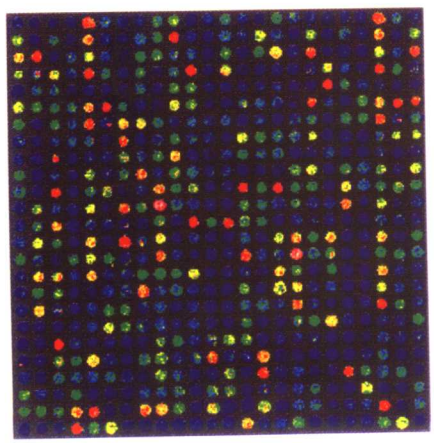


FIGURE 6.24. *Human total RNA (4 μg) was labeled and hybridized to a microarray containing full-length human cDNAs prepared using Full-Length Expressed Gene (FLEX) technology (AlphaGene, Woburn, MA). The hybridized microarray was then stained with TSA reagents from a MICROMAX system (NEN Life Science Products, Boston), and fluorescent gene expression signals were displayed in a rainbow palette.*

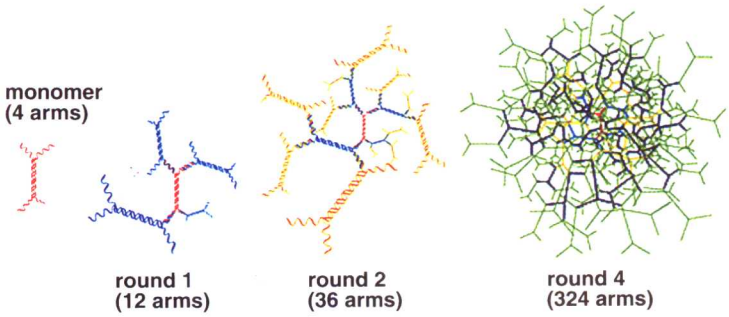


FIGURE 6.25. *For the dendrimer synthesis process, two oligonucleotides are annealed to produce a molecule (monomer) that has a double-stranded central portion and four single-stranded arms. Monomers are then annealed and cross-linked in a stepwise manner; four successive rounds of annealing and cross-linking produce dendrimers with 324 arms. If the synthesis process is performed with fluorescent oligonucleotides, each dendrimer contains as many as 324 fluorescent dye molecules. (Courtesy of Dr. R. C. Getts, Genisphere, Montvale, NJ.)*

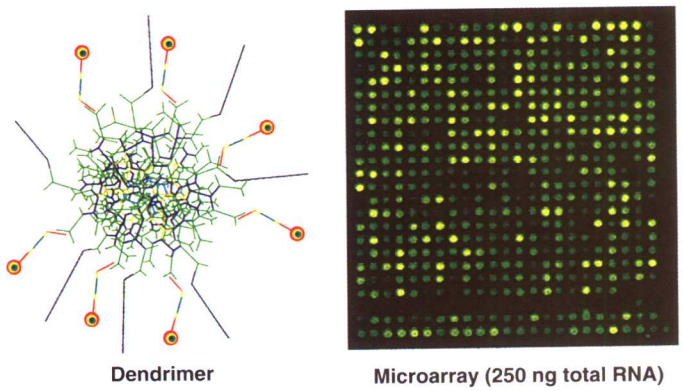


FIGURE 6.26. *A dendrimer reagent was used to label a microarray indirectly by hybridizing the single-stranded capture sequences* (black lines) *on the dendrimers to a microarray hybridized with a probe mixture prepared from 250 ng mouse total RNA. The intense fluorescent signals are due to passive signal enhancement by the dye molecules* (circles) *attached to the dendrimers. (Dendrimer courtesy of Dr. R. C. Getts, Genisphere, Montvale, NJ; microarray data courtesy of University of California at San Francisco Cancer Center.)*

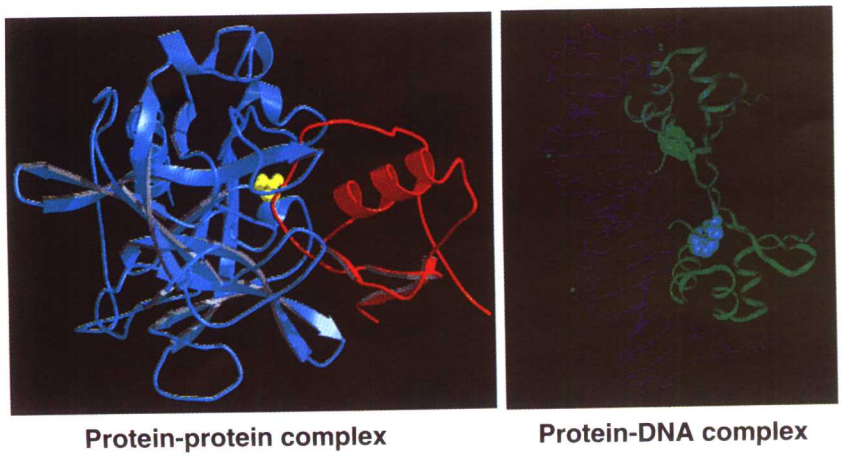


FIGURE 6.31. *Crystal structures of a protein–protein binding reaction between the elastase enzyme* (light blue) *and the ovomucoid substrate* (red), *and a protein–DNA binding reaction between the Cro-repressor protein* (green) *and double-stranded DNA* (blue and magenta). *(Protein-protein complex courtesy of Dr. K. P. Murphy, Department of Biochemistry, University of Iowa at Iowa City; protein–DNA complex courtesy of Dr. M. C. Mossing, University of Notre Dame, Notre Dame, IN.)*

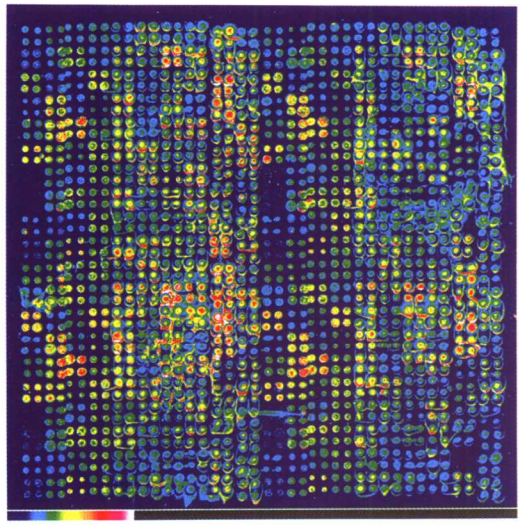


FIGURE 8.18. *Fluorescent image of a cDNA microarray printed with a prototype ink-jet dispensor and stained with the OliGreen dye. Detection was made with a confocal scanning device using an Argon laser for excitation; data are represented in a rainbow palette. (Courtesy of M. Schena and R. W. Davis, Stanford University, Stanford, CA, and Combion, Pasadena, CA.)*

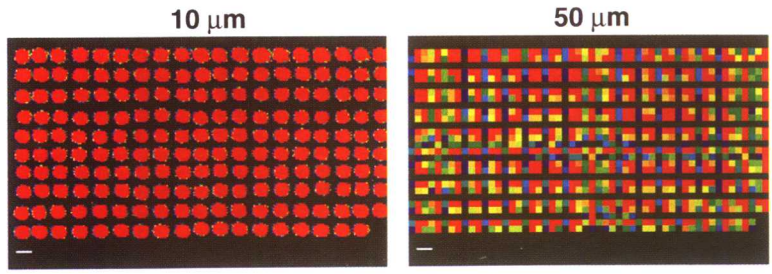


FIGURE 8.36. *Two microarray images obtained from the same fluorescent microarray, scanned at a pixel size of 10 μm and 50 μm. The TIFF data were obtained using a ScanArray 3000 (Packard Biochip, Billerica, MA) and are presented in a rainbow palette from* NIH Image 1.62 *software developed by Rosband (National Institutes of Health, Bethesda, MD). Space bars = 150 μm.*

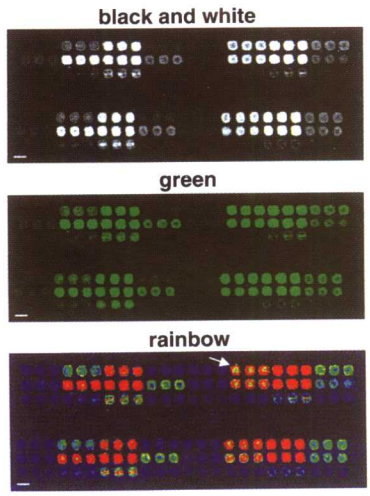


FIGURE 8.37. *A single microarray image with the values from the TIFF file used to index a LUT composed of a black and white, green, or rainbow scale. The brightest fluorescent signals are represented as white, bright green, and red in the three respective palettes. Space bars = 150 μm.* Arrow, *the portion of the image shown in Figure 8.38. (Data generated with a ScanArray 3000 instrument set to the Cy3 channel.)*

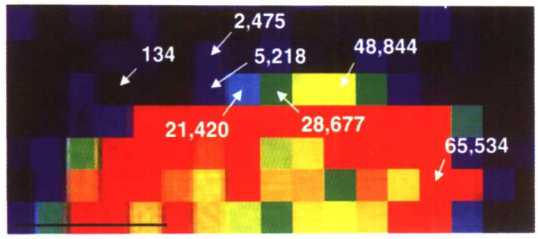


FIGURE 8.38. *An enlarged image of a microarray (from Fig. 8.37) with intensity values given for seven different pixels. The custom rainbow palette was indexed with a 16-bit TIFF file. Space bar = 50 μm.*

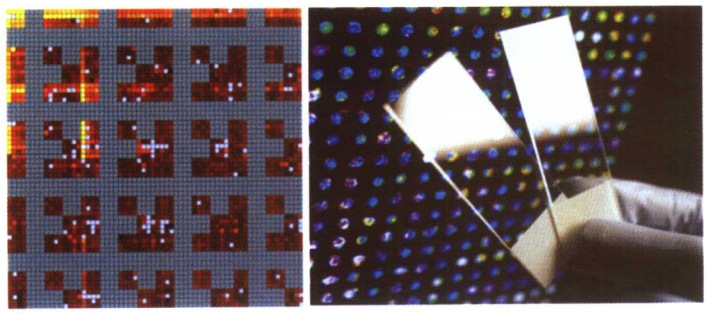


FIGURE 8.39. A, *Oligonucleotide hybridization data displayed via a custom color palette. (Reprinted with permission from* Nucleic Acids Research.*)* ***B,*** *Corning Microarray Technology slides and gene expression data displayed in a custom color palette. (Data courtesy of Corning, Corning NY.)*

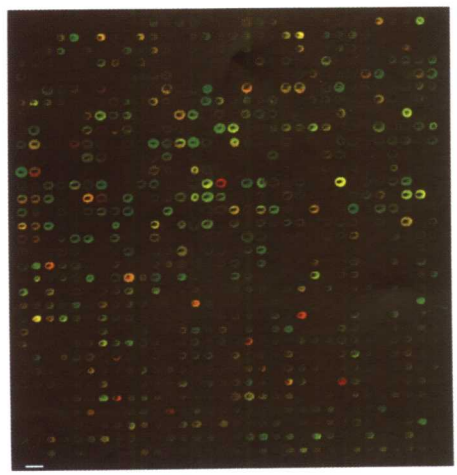


FIGURE 8.40. *A green–red overlay is created by subtracting image values obtained from separate channels of a microarray scanner and superimposing the images to create a composite. The samples were prepared by labeling mRNA from human cells grown at 37 and 42°C, respectively. (Courtesy of Dr. M. Schena and Dr. R. W. Davis, Stanford University, Stanford, CA, and by Dr. D. Shalon, Synteni, Palo Alto, CA.)*

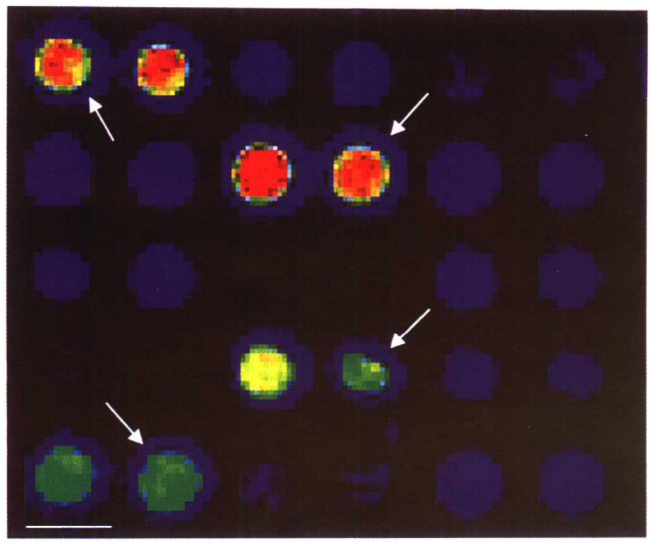


FIGURE 8.48. *Fluorescent microarray spots captured with a CCD-based camera, with high signal intensities manifesting a blooming phenomenon* (arrows). *Fully charged pixels transfer excess charge to adjacent pixels that are empty or filled partially. Data are represented in a rainbow palette. Space bar = 150 μm.*

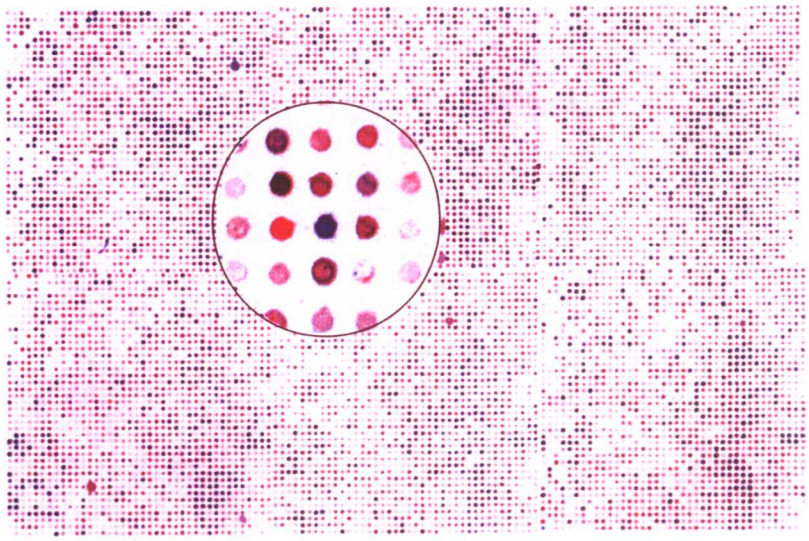


FIGURE 8.50. *Human oral carcinoma cells were treated with the chemotherapy drug camptothecin and compared to untreated cells. Two-color data are represented in a composite image so that dark and light circles correspond to genes that are activated and repressed, respectively, by the drug. A portion of the array* (circle) *is enlarged for ease of viewing and the data were obtained by Dr. Konan Peck. (Reprinted with permission from Academic Press, San Diego, CA.)*

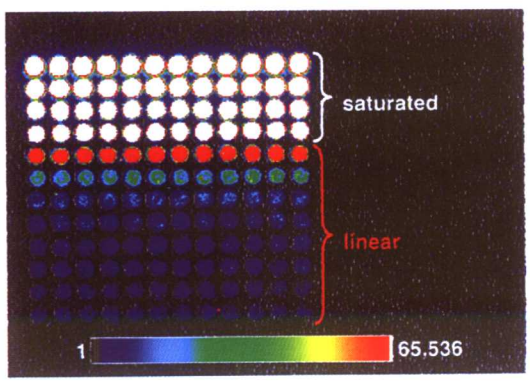


FIGURE 9.51. *Saturated signals. Microarray data from a fluorescent scan are represented in a rainbow palette coded to signal intensities from 1 to 65,536 counts as shown in the color bar (bottom). Rows correspond to 10-fold dilutions of a fluorescent oligonucleotide, ranging from 100 pmol/μL to 0.1 attomol/μL (rows 1–10) with 12 replicates per dilution. Rows 11 and 12 (bottom two rows) contain buffer only. (Data courtesy of Dr. R. Stears, TeleChem/arrayit.com Sunnyvale, CA.)*