Microarrays and clinical dentistry

WINSTON PATRICK KUO, D.D.S., M.S.; MARK E. WHIPPLE, M.D., M.S.; TOR-KRISTIAN JENSSEN, Ph.D.; RANDY TODD, D.M.D., M.D., D.M.Sc.; JOEL B. EPSTEIN, D.M.D., M.S.D.; LUCILA OHNO-MACHADO, M.D., Ph.D., M.H.A.; STEPHEN T. SONIS, D.M.D., D.M.Sc.; PETER J. PARK, Ph.D.

Prevention, diagnosis and treatment in dental practice are based on an understanding of the biology underlying oral health and disease. Few aspects of patient care will remain untouched by today's rapid advances in biological research. In the future, dentists may use inexpensive but remarkably sophisticated diagnostic tests to diagnose infection, oral lesions and symptoms of temporomandibular dysfunction, or TMD. The translation of "new biology" into clinical practice is likely to emerge as a result of large-scale, worldwide research collaborations.

Microarrays hold much promise for the analysis of diseases in the oral cavity.

The Human Genome Project, or HGP, is a multinational scientific effort to determine the 3.2 billion DNA bases that make up the human genome. These sequences are the instructional material for the mechanisms governing the behavior of each cell, including why some cells become part of a disease process and others do not. The small variations in the DNA sequence that lead to different characteristics (such as skin color, facial features or height) are known as polymorphisms, which also can cause or contribute to the development of many syndromes and diseases.[1] Therefore, the completion of the draft version of the human genome sequence in 2001 was an important first step for a more comprehensive understanding of the fundamental determinants of health and disease.[2] In addition, the HGP provides the basis for many other scientific questions. For example, the genomes of many other organisms are being sequenced; this will allow the examination of cross-species similarity and differences as represented in their genomic sequences and provide us with insights on human evolution.

The great challenge now is to understand how the

ABSTRACT

Background. The Human Genome Project, or HGP, has inspired a great deal of exciting biology recently by enabling the development of new technologies that will be essential for understanding the different types of abnormalities in diseases related to the oral cavity.

Literature Reviewed. The authors review current literature pertaining to the advanced microarray technologies arising from the HGP and how they can contribute to dentistry. This technology has become a standard tool for monitoring activities of genes at both academic and pharmaceutical research institutions.

Results. With the availability of the DNA sequences for the entire human genome, attention now is focused on understanding various diseases at the genome level. Deciphering the molecular behavior of genetically encoded proteins is crucial to obtaining a more comprehensive picture of disease processes. Important progress has been made using microarrays, which have been shown to be effective in identifying gene expression patterns and variations that correlate with cellular development, physiology and function. Arrays can be used to classify tissue samples accurately based on molecular profiles and to select candidate genes related to a number of cancers, including oral cancer. This type of oral genetic approach will aid in the understanding of disease progression, thus improving diagnosis and treatment for patients.

Clinical Implications. Microarrays hold much promise for the analysis of diseases in the oral cavity. As the technology evolves, dentists may see these tools as screening tests for better managing patients' dental care.

BOX

KEY TERMS.

COMPLEMENTARY DNA, OR cDNA: DNA synthesized from a messenger RNA template that corresponds to expressed sequences of genomic DNA. The term also may refer to DNA that is complementary to a particular DNA sequence. This single-stranded form often is used as a probe.

DNA: the molecule that encodes genetic information. DNA is a double-stranded molecule held together by weak bonds between base pairs of nucleotides. The four nucleotides in DNA contain the bases: adenine, or A; guanine, or G; cytosine, or C; and thymine, or T. A generally pairs with T and C pairs with G.

DNA SEQUENCE: the relative order of base pairs, whether in a fragment of DNA, a gene, a chromosome or an entire genome.

GENE: the functional and physical unit of heredity passed from parent to offspring. Genes are pieces of DNA, and most genes contain the information for making a specific protein.

GENE EXPRESSION: the process by the instructions encoded in DNA is converted into a unit of biological function in a cell.

GENOME: the genetic material in the chromosomes of a particular organism; its size generally is provided by the total number of base pairs.

HUMAN GENOME PROJECT, OR HGP: an international research project to map each human gene and to completely sequence human DNA.

HYBRIDIZATION: the process of joining two complementary strands of DNA or one strand each of DNA and RNA to form a double-stranded molecule. One strand often is labeled and used as a probe to detect the presence of the second strand.

MESSENGER RNA, OR mRNA: the template for protein synthesis; each set of three bases, called codons, specifies a certain amino acid in the sequence of amino acids that comprise the protein. The sequence of a strand of mRNA is based on the sequence of a complementary strand of DNA.

MICROARRAY TECHNOLOGY: tools that allow for the massive parallel analysis of genes and proteins.

MUTATION: any permanent and heritable change in the DNA sequence; includes point mutations, deletions, insertions and changes in number and structure of chromosomes.

POLYMORPHISM: the existence, in a population, of two or more alleles of a gene, in which the frequency of rarer alleles is greater than can be explained by recurrent mutation alone (greater than 1 percent of the total population).

PROBE: single-stranded DNA or RNA or a small molecule of specific base sequence used to detect the function of a gene.

PROTEIN: a large complex molecule made up of one or more chains of amino acids in a specific order. Proteins are required for the structure, function and regulation of the body's cells, tissues and organs, and each protein has unique functions.

enormous amount of information contained in the genome is translated into molecular action. By examining the sequences, for example, researchers hope to implicate regions of DNA (genes) that control normal and disease processes. An early application of this effort has involved the development of microarrays: small devices, about the size of a thumbnail, that contain pieces of DNA sequences that can measure the gene expression levels of thousands of genes simultaneously.[3-5] In traditional biological research, genes were studied one at a time through a series of laborious and time-consuming experimental techniques. This is one reason that only a small fraction of all the genes have been characterized so far. By allowing simultaneous genome-wide measurements, microarrays can greatly expedite the research process, as well as enable new avenues of research. Currently, microarrays are used primarily in research, but their enormous clinical potential is beginning to be explored.

Several recent reports have discussed how the HGP can play an integrative role in the oral health care profession in terms of future directions and challenges.[6-8] In this review, we will discuss microarrays, emphasizing the biology behind the technology, the types of arrays available and some future applications to clinical dentistry. (Author's note: the box provides a glossary of some of the key terms used in this article.)

THE BIOLOGY BEHIND MICROARRAYS

The structure and function of cells are determined by proteins. Many diseases are caused by defects in the natural functions of proteins; hence, drugs often target certain proteins to alter their behavior. Depending on the environment imposed on a cell, the same protein may be modified to perform different, and sometimes opposing, functions. To treat various diseases effectively, the genetic makeup underlying the disease process ultimately needs to be characterized and understood at the protein level. However, obtaining a detailed view of the cellular process by measuring the levels of thousands of proteins simultaneously is an extremely daunting prospect.

While proteins carry out the activities in a cell, they ultimately are controlled by the instructions

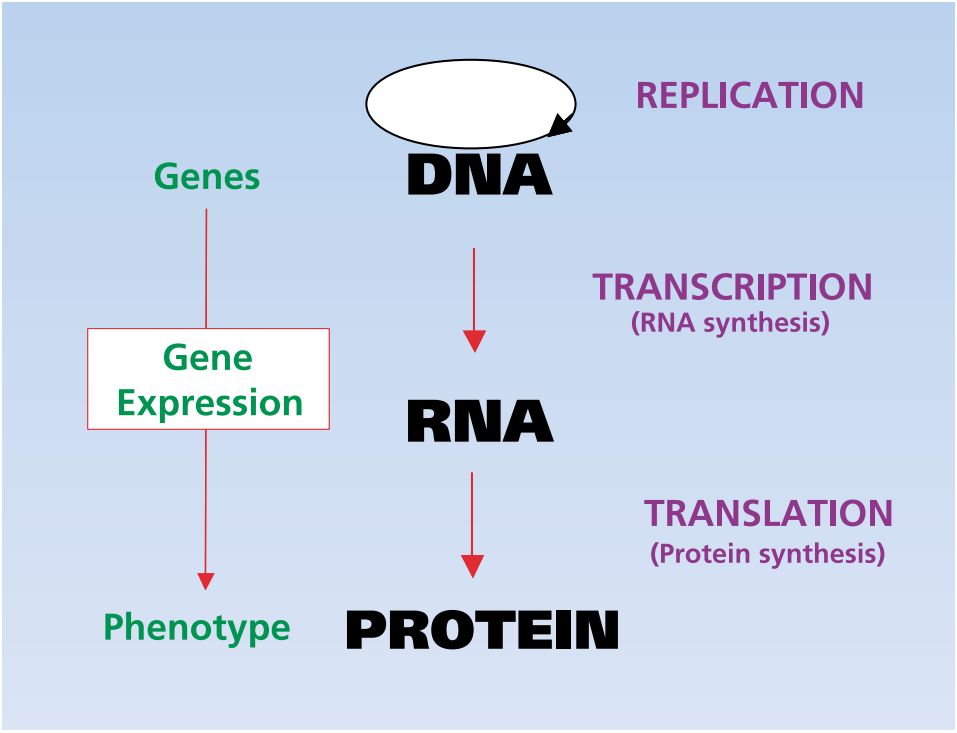


Figure 1. Transcription of DNA to RNA to protein.

contained in the DNA sequences, through an intermediate step involving messenger RNA, or mRNA. mRNA is an intermediary molecule that carries the genetic information from the cell nucleus to the cytoplasm, where proteins are made. When certain genes become active, or "expressed," many copies of mRNAs corresponding to those genes are produced (a process called "transcription"), and they in turn generate the necessary proteins (a process called "translation") (Figure 1). Therefore, by knowing which mRNAs are present in a cell, we can identify which genes are being expressed, and we can begin to understand some of the basic elements controlling the structure and function of the cell.

Because measuring the levels of mRNAs is easier than measuring those of proteins, many researchers use mRNA as a surrogate marker. This is the idea behind the DNA microarrays. However, because mRNA is only a substitute for proteins, there may be some problems in relying on expression levels generated from microarrays. For example, although mRNA accurately reflects portions of expressed DNA, the extent of mRNA expression does not necessarily correlate with the amount and function of proteins they encode. Ideally, one would like to measure the final products of every gene, the proteins or—even better—the biochemical activity of these products, which are related more directly to biological functionality. Furthermore, mRNA is degraded easily and therefore has to be converted to the more stable form of complementary DNA, or cDNA, before an experiment can be conducted.

An estimated 30,000 to 40,000 genes are present in the human genome. While every cell contains every gene within that cell's DNA, only 10 to 20 percent of the genes are expressed in a given cell at a particular time. By measuring how these genes are differentially expressed differently among various conditions, it is possible to implicate certain genes as being relevant in the underlying process. For example, by comparing the different patterns of gene expression levels between a group of cancer patients and a group of normal patients, we can find the genes that may be associated with cancer. Without microarrays, one would have to study thousands of genes one at a time; with them, one can examine all of those thousands of genes simultaneously.

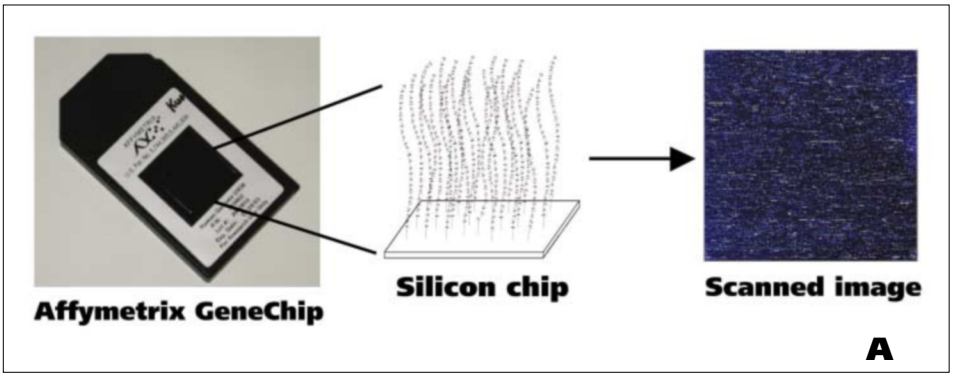
MICROARRAY TECHNOLOGY

A microarray is composed of a series of pieces of DNA, usually corresponding to segments of genes, that are placed in a prespecified arrangement, typically on a glass slide or a silicon chip.[9] In addition to the array designed for the human genome, different arrays are available for the genomes of a variety of model organisms such as yeast and mouse.[10-13]

The principle behind microarrays is the fact that complementary sequences will bind to each other under the right conditions. That is, if the DNA is 10 base pairs long and the sequence is ATGGCTTGAC, for example, the mRNA containing segment TACCGAACTG will "hybridize" to the probe. (The base pair A is complementary to T and C is complementary to G; the actual DNAs on the array are at least 25 base pairs long to maximize the chance of there being a unique

sequence for each gene.) On a microarray, many thousands of spots are placed on a rectangular grid, each spot containing a large number of pieces of DNA from a particular gene. When the sample of interest contains many copies of mRNA, many bindings will occur, indicating that the gene from which the mRNA is transcribed is highly expressed. The quantity of hybridization can be determined because each copy of mRNA is labeled in the experiment with a fluorescent or radioactive tag and a brighter signal is detected when more copies bind.

There are presently two types of microarray technology in common usage: cDNA microarrays and oligonucleotide microarrays. They differ in their method, design and image. cDNA arrays contain long fragments of DNA (from 100 to thousands of base pairs) and oligonucleotide arrays contain short fragments of DNA (25 base pairs) (Figure 2). cDNA microarrays are created by robotically spotting individual samples of purified cDNA clones onto a glass slide, whereas oligonucleotide microarrays are made by constructing exact probe sequences using a photolithography masking technique similar to the one used for fabricating computer chips. cDNA arrays often are made by individual laboratories based on their particular resources and needs; oligonucleotide arrays are built base by base on the array surface by repeated cycles of photodeprotection and nucleotide addition until the desired length is reached. These arrays generally are more costly to make and are purchased through the manufacturer.



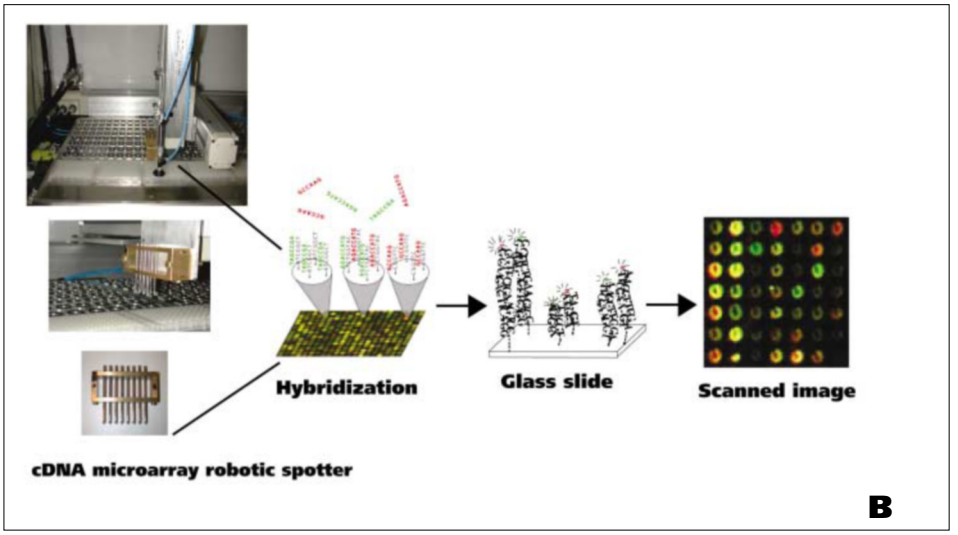


Figure 2. Oligonucleotide and complementary DNA, or cDNA, microarray. A. Affymetrix GeneChip (Santa Clara, Calif.) microarray cartridge, consisting of oligonucleotide probes (25 base pairs long). Within the microarray, up to 1 million unique oligonucleotide probes can be affixed to a silicon chip. The plastic cartridge protects the microarray and serves as a chamber for hybridization. After hybridization, the array is scanned. B. Spotted cDNA microarray robot. A light source causes the spots that contain cDNA probes (from 100 to thousands of base pairs long) hybridizing to the Cy3 labeled target strands to fluoresce red. Spots that hybridize to Cy5 labeled strands fluoresce green, and spots that hybridize both Cy3 and Cy5 fluoresce yellow. The intensity of the fluorescence correlates with the amount of hybridization. After hybridization, the array is scanned. Image of GeneChip probe array reproduced with permission of Affymetrix.

The idea behind this technology has been applied to other areas. There are genotyping arrays that contain oligonucleotides representing the various sequence possibilities at specific loci and can be used to detect single nucleotide polymorphisms, or SNPs.[14] This new approach provides a quick way to genotype thousands of different loci at a time, which is useful for association and linkage studies to isolate chromosomal regions related to a particular disease. This array also can be used to locate chromosomal aberrations related to cancer, such as segments of allelic imbalance, which can be identified by loss of heterozygosity.[15] By comparative genomic hybridization techniques on genomic DNA, amplified or deleted regions in the chromosomes can be identified, such as in the case of oral cancer.[16]

THE CHALLENGES OF USING MICROARRAYS

While these arrays are designed to give a genome-wide view of the cell on an unprecedented scale, there are some problems and obstacles. A major drawback is their high cost, although prices have dropped significantly in the past few years. Each oligonucleotide array typically costs hundreds of dollars; chemicals and labor needed for the experiment also can cost a substantial amount. This is in addition to the hundreds of thousands of dollars necessary for the initial setup.

Another challenge is the efficient management and accurate analysis of the large volume of data generated by microarrays. As with other high-throughput technologies, microarrays collect massive amounts of raw data. These data then need to be stored in a database in such a way that relevant information can be retrieved efficiently and be linked to various databases on the Internet containing biological annotations. Analysis of the data requires considerable expertise in statistics and data-mining techniques, especially given the low quality of data at present. In addition to the usual biological variations, microarray data can be very noisy owing to many technical shortcomings that are yet to be overcome. Recently, poor reproducibility between different technologies for the same samples has been observed.[17] Standard methods of analysis include cluster analysis for grouping similar patient samples or similar genes, as well as classification techniques for finding genes differentially expressed between conditions and predicting the class of a new sample. Many software packages are available for standard analysis, but the validity of the statistical methods in these packages needs to be examined for particular data sets, and more sophisticated analysis than is available in these software packages is necessary for extracting the maximum amount of information from the data. Microarray data present many unresolved statistical issues, and therefore new methodology to remediate this is being developed by many researchers in the field of bioinformatics.

Quality of the laboratory work also is important to minimize the noise level in the data that are generated, which can substantially decrease the number of false positive gene expression changes. "Noise" is an informal term widely used in microarray data analyses and experiments to refer to parameters extraneous to the aspect of a system one is investigating, which eventually can interfere with data analysis. Performing a microarray experiment and the follow-up validations for genes with statistically significant result requires expertise in the experimental procedures, including the preservation of biological specimens in a manner that protects the integrity of mRNA.[18]

While DNA microarrays have become mainstream, the development of protein microarrays is still in its infancy. The microarray format is similar to that of oligonucleotide and cDNA arrays for gene expression profiling, but instead of DNA sequences, proteins are deposited in specified spots on the chip. There are many formidable technical hurdles at this point, but this technology has considerable potential, as it would allow us to measure the protein abundance directly rather than measuring the abundance of the intermediary mRNA molecules. It is well-known that mRNA levels do not always accurately reflect the level or activity of proteins, the molecules that directly govern cell function.[19] Protein arrays can reveal information about post-translational modifications or protein-protein interactions,[20] and the information obtained from them in the future will complement transcriptional data.[21] Among the many applications will be a rapid approach to identifying small molecules with desirable drug-protein interactions. This would lead to a more direct and efficient method of finding drug candidates.

The molecular and biochemical activity associated with oral health and disease has the potential to be understood and classified by the 'signature' of the DNA or the proteins characterized using microarray technologies.

APPLICATIONS OF MICROARRAYS TO DENTISTRY

The molecular and biochemical activity associated with oral health and disease has the potential to be understood and classified by the "signature" or "fingerprint" of the DNA or the proteins characterized using these technologies. Areas that can be coupled with microarray technologies include classification of diseases, or molecular phenotyping[22]; the study of gene function in relation to gene regulatory networks, or functional

genomics[23]; drug development and prediction of efficacy and toxicity, or pharmacogenomics[24]; and developmental biology.[25]

Antibiotic treatment. Antibiotic treatment and prevention of infection have become increasingly complex with the rise of antibiotic-resistant bacterial strains. While certain local and systemic factors increase the likelihood of postoperative infections, little is known about virulent strains that may increase a patient's potential susceptibility to orofacial infections. Many anaerobic bacteria might be the infective agent, but they often are not easily culturable; organisms such as actinomyces require long incubation periods and so must be treated empirically. DNA microarray analysis may offer a more sensitive and accurate endpoint for several reasons. First, bacterial genomic DNA often outlasts the viability of the bacteria. Second, a diagnosis can be made using a small amount of DNA, as opposed to the large numbers of bacteria needed for culture. One day, an abscess specimen might be sent not for culture and sensitivity testing but rather for DNA microarray analysis.

Oral lesions. Leukoplakia or white lesions of the oral cavity may result from a myriad of reversible conditions. Currently, microscopic examination fails to identify the small subset of these lesions that progress to oral cancer. Identification of gene expression profiles or "genomic fingerprints" will allow clinicians to differentiate harmless white lesions (leukoplakias) from precancerous lesions or from very early cancer. Recent studies have illustrated the effectiveness of microarrays in oral cancer.[26,27] In the future, samples taken from an incisional biopsy or brush biopsy may be sent to a laboratory for gene expression analysis. Early diagnosis and management of oral cancer is correlated with increased survival. Identification and treatment of premalignant and early cancerous oral lesions may become one of the most valuable services dentistry performs.

TMD. TMD is a pervasive and challenging problem across the disciplines of dentistry. Today, diagnosis of and treatment planning for patients with TMD are based largely on anatomical changes associated with pain, functional complaints or both. Treatment of TMD can be controversial and unsuccessful, even if the right "diagnosis" is obtained. However, because TMD is a complex, multifactorial problem, researchers are studying the mediators and byproducts of the disease process within the joint fluid to develop new, biologically based diagnostic and therapeutic approaches.

Ultimately, protein arrays may prove to be important tools used in identifying biomarkers associated with TMD. In the future, clinicians may be able to better classify and treat TMD by microarray analysis of joint aspirates.

CONCLUSION

Microarrays hold much promise for the analysis of diseases in the oral cavity. Classification of oral disease by DNA, RNA or protein profiles will greatly enhance our ability to diagnose, prevent, monitor and treat our patients. Currently, microarrays are primarily a research tool. However, in the future, these highly sophisticated screening tools or their derivatives may provide an accurate, simple, rapid and inexpensive means for better managing the care of dental patients. Microarrays promise a more biologically based, individualized and vastly improved standard for oral care that will have great clinical impact on the way dentistry will be practiced in the future. ■

Dr. Kuo is an oral medicine and dental informatics fellow, Department of Oral Medicine, Infection and Immunity, Harvard School of Dental Medicine, Boston; a biomedical informatics research fellow, Children's Hospital Informatics Program, Children's Hospital, Harvard Medical School, Boston; a biomedical informatics research fellow, Decision Systems Group, Brigham and Women's Hospital, Harvard Medical School, Boston; a biomedical informatics research fellow, Department of Genetics, Cepko Laboratory, Harvard Medical School, Boston; and a biomedical informatics research fellow, Division of Health Sciences and Technology, Harvard University and Massachusetts Institute of Technology, Cambridge, Mass. Address reprint requests to Dr. Kuo at Harvard School of Dental Medicine, Department of Oral Medicine, Infection, and Immunity, 188 Longwood Ave., Boston, Mass. 02115, e-mail "winston_kuo@hms.harvard.edu".

Dr. Whipple is an assistant professor, Department of Otolaryngology-Head and Neck Surgery, University of Washington School of Medicine, Seattle.

Dr. Jenssen is a research scientist, Department of Tumor Biology, The Norwegian Radium Hospital, Montebello, Norway.

Dr. Todd is an assistant professor, Department of Oral and Maxillofacial Surgery, Massachusetts General Hospital, Harvard Medical School, Boston.

Dr. Epstein is a professor and the head of the department, Department of Oral Medicine and Diagnostic Sciences, University of Illinois, Chicago.

Dr. Ohno-Machado is an associate professor, Decision Systems Group, Brigham and Women's Hospital, Harvard Medical School, Boston; and a member of the HST Affiliated Faculty, Division of Health Sciences and Technology, Harvard University and Massachusetts Institute of Technology, Cambridge, Mass.

Dr. Sonis is the chairman and a professor, Department of Oral Medicine Infection and Immunity, Harvard School of Dental Medicine, Boston.

Dr. Park is an instructor, Children's Hospital Informatics Program, Children's Hospital, Harvard Medical School, Boston.

1. Wang DG, Fan JB, Siao CJ, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 1998;280:1077-82.
2. Lander E, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. Nature 2001;409:860-921.
3. Marshall A, Hodgson J. DNA chips: an array of possibilities. Nat Biotechnol 1998;16:27-31.
4. Schena M, Shalon D, Davis RW, Brown P. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995;270:467-70.
5. Ramsay G. DNA chips: state-of-the art. Nat Biotechnol 1998;16:40-4.
6. Kuo WP, Whipple ME, Sonis ST, Ohno-Machado L, Jenssen TK. Gene expression profiling by DNA microarrays and its application to dental research. Oral Oncol 2002;38:650-6.
7. Wright JT, Hart TC. The genome projects: implications for dental practice and education. J Dent Educ 2002;66:659-71.
8. Slavkin HC. The human genome, implications for oral health and diseases, and dental education. J Dent Educ 2001;65:463-79.
9. Rogers YH, Jiang-Baucom P, Huang ZJ, Bogdanov V, Anderson S, Boyce-Jacino MT. Immobilization of oligonucleotides onto a glass support via disulfide bonds: a method for preparation of DNA microarrays. Anal Biochem 1999;266:23-30.
10. Kato N, Lam E. Detection of chromosomes tagged with green fluorescent protein in live *Arabidopsis thaliana* plants. Genome Biol 2001;2(11):Research0045.1-0045.10. Available at: "genomebiology.com/2001/2/11/research/0045". Accessed Feb. 20, 2003.
11. Shalon D, Smith SJ, Brown PO. A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. Genome Res 1996;6:639-45.
12. Pedrosa AL, Ruiz JC, Tosi LR, Cruz AK. Characterisation of three chromosomal ends of *Leishmania major* reveals transcriptional activity across arrays of reiterated and unique sequences. Mol Biochem Parasitol 2001;114:71-80.
13. Veltman JA, Schoenmakers EF, Eussen BH, et al. High-throughput analysis of subtelomeric chromosome rearrangements by use of array-based comparative genomic hybridization. Am J Hum Genet 2002;70:1269-76.
14. Warrington JA, Shah NA, Chen X, et al. New developments in high-throughput resequencing and variation detection using high density microarrays. Hum Mutat 2002;19:402-9.
15. Lindblad-Toh K, Tanenbaum DM, Daly MJ, et al. Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. Nat Biotechnol 2000;18:1001-5.
16. Jiang WW, Fujii H, Shirai T, Mega H, Takagi M. Accumulative increase of loss of heterozygosity from leukoplakia to foci of early cancerization in leukoplakia of the oral cavity. Cancer 2001;92:2349-56.
17. Kuo WP, Jenssen TK, Butte AJ, Ohno-Machado L, Kohane IS. Analysis of matched mRNA measurements from two different microarray technologies. Bioinformatics 2002;18:405-12.
18. Schuchhardt J, Beule D, Malik A, et al. Normalization strategies for cDNA microarrays. Nucleic Acids Res 2000;28:E47.
19. Gygi SP, Rochon Y, Franza BR, Aebersold R. Correlation between protein and mRNA abundance in yeast. Mol Cell Biol 1999;19:1720-30.
20. MacBeath G, Schreiber SL. Printing proteins as microarrays for high-throughput function determination. Science 2000;289:1760-3.
21. Emili AQ, Cagney G. Large-scale functional analysis using peptide or protein arrays. Nat Biotechnol 2000;18:393-7.
22. Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503-11.
23. Spellman PT, Sherlock G, Zhang MQ, et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. Mol Biol Cell 1998;9:3273-97.
24. Evans WE, Relling MV. Pharmacogenomics: translating functional genomics into rational therapeutics. Science 1999;286:487-91.
25. White KP, Rifkin SA, Hurban P, Hogness DS. Microarray analysis of *Drosophila* development during metamorphosis. Science 1999;286:2179-84.
26. Alevizos I, Mahadevappa M, Zhang X, et al. Oral cancer in vivo gene expression profiling assisted by laser capture microdissection and microarray analysis. Oncogene 2001;20:6196-204.
27. Kuo WP, Jenssen TK, Park PJ, Lingen MW, Hasina R, Ohno-Machado L. Gene expression levels in different stages of progression in oral squamous cell carcinoma. Proc AMIA Symp 2002:415-9.