Journal of Cellular Biochemistry 101:1074–1086 (2007)

Dissecting the Molecular Mechanisms of Cancer Through Bioinformatics-Based Experimental Approaches

Ashley G. Rivenbark and William B. Coleman*

Department of Pathology and Laboratory Medicine, Curriculum in Toxicology, University of North Carolina Lineberger Comprehensive Cancer Center, University of North Carolina School of Medicine, Chapel Hill, North Carolina

Abstract Cancer is a disease of aberrant gene expression characterized by inappropriate (temporal or quantitative) expression of positive mediators of cell proliferation in conjunction with diminished expression of negative mediators of cell growth. Alteration of the normal balance of these positive and negative mediators leads to the abnormal growth of cells and tissues that typify neoplastic disease. Development of a better understanding of the genetic and epigenetic mechanisms that induce neoplastic transformation and drive the cancer phenotype is essential for continued progress towards the design of practical molecular diagnostics and effective treatment strategies. Over the past decades, molecular techniques that facilitate the assessment of gene expression, identification of gene mutations, and characterization of chromosome abnormalities (numeric and structural) have been established and applied to cancer research. However, many of these techniques are slow and labor-intensive. More recently, high-throughput technologies have emerged that generate large volumes of data related to the genetics and epigenetics of cancer (or other disorders). These advances in molecular genetic technology required the development of sophisticated bioinformatic tools to manage the large datasets generated. The combination of high-throughput molecular assays and bioinformatic-based data mining strategies has significantly impacted our understanding of the molecular pathogenesis of cancer, classification of tumors, and now the management of cancer patients in the clinic. This article will review basic molecular techniques and bioinformatic-based experimental approaches used to dissect the molecular mechanisms of carcinogenesis. J. Cell. Biochem. 101: 1074–1086, 2007. © 2007 Wiley-Liss, Inc.

Key words: microarray; bioinformatics; genomics; cancer genetics

Carcinogenesis is a multi-step process driven by changes in gene expression that leads to the abnormal growth of cells and tissues that characterize the neoplastic phenotype of cancer. It is now well-recognized that cancer in its simplest form is a genetic disease, or more specifically, a disease of aberrant gene expression. The processes of tumor formation and progression are driven by chromosomal alterations, gene mutations, and epigenetic alterations of DNA that most often affect the expression of genes controlling cell proliferation and survival. The genes that are important for cancer development are frequently classified as proto-oncogenes and tumor suppressor genes, but also include growth factors and their receptors, cell signaling molecules, and nuclear transcription factors. Most proto-oncogenes function to regulate cell growth or survival. Activated forms of proto-oncogenes are termed oncogenes, and usually represent gain of function mutants. Conversely, tumor suppressor genes function in normal cells to suppress cell growth, and are usually recessive genes that can be passed through the heterozygous germline, leading to increased susceptibility to cancer. Other important changes in gene expression associated with cancer include autocrine production of growth factors, abnormal

Grant sponsor: Susan G. Komen Breast Cancer Foundation; Grant number: BCTR0100575; Grant sponsor: National Cancer Institute (NIH); Grant number: CA78343; Grant sponsor: Medical Alumni Endowment Fund of the University of North Carolina at Chapel Hill.

*Correspondence to: William B. Coleman, Department of Pathology and Laboratory Medicine, University of North Carolina School of Medicine, 515 Brinkhous-Bullitt Building, CB#7525, Chapel Hill, NC 27599.
E-mail: william.coleman@pathology.unc.edu

Received 3 January 2007; Accepted 4 January 2007
DOI 10.1002/jcb.21283

© 2007 Wiley-Liss, Inc.

regulation of genes that encode proteins important for apoptosis (loss of pro-apoptotic or gain of anti-apoptotic), as well as alterations associated with invasion/metastasis (proteases, etc.). Thus, carcinogenesis is a multi-step process resulting from the sequential perturbation of both positive and negative mediators of cell proliferation and other regulatory genes. The molecular basis of cancer has been the focus of intense investigation for many years, and without an understanding of the molecular events that drive neoplastic transformation, the development of human cancer will remain ill-defined. Although we have cataloged many of the molecular genetic and epigenetic consequences of cellular transformation, we are just beginning to understand the complicated events that lead to the neoplastic phenotype.

Bioinformatics represents a relatively new field of scientific inquiry that emerged to fill the need created by advances in high-throughput molecular techniques. The construction and advancement of computational and statistical techniques to solve biological problems posed by the analysis of molecular biology data sets is bioinformatics. For example, recent advances in molecular technology related to the Human Genome Project gave rise to the rapid sequencing of large portions of the human genome (and others) [Shih Ie and Wang, 2005]. The resulting surge of DNA sequence information necessitated the development of an information science able to organize large and complex volumes of data [Segal et al., 2005]. Thus, the field of bioinformatics was born. During the subsequent interval of time, emerging technologies continue to simplify the generation of large genomic data sets, requiring the development of additional bioinformatic solutions and tools for analyzing and comparing these data sets [Travis et al., 2006]. It is widely believed that genomics coupled with bioinformatics will revolutionize our understanding of the complex molecular pathways that constitute the fundamental mechanisms of cancer development, and that this understanding will transform the diagnosis and treatment of disease as the pathways and specific genes that comprise the molecular pathogenesis of cancer become targets for new therapeutic interventions [Segal et al., 2005].

This article will review basic molecular techniques as well as newer array-based technologies, and will discuss bioinformatic-based experimental approaches used to investigate the molecular pathogenesis of cancer. The intent of this review is to provide the reader with a description of the available tools for dissecting mechanisms of carcinogenesis, as well as to provide illustrations of how these tools have been applied to cancer research to date. The literature on this subject is extensive. Therefore, we have chosen to cite recent reviews and pertinent earlier literature. We encourage the reader to consult the recent primary literature on topics of interest.

BASIC CONCEPTS AND TECHNOLOGIES

The establishment of molecular mechanisms governing cancer pathogenesis emerged over the last 50 years from investigations that employed a number of basic molecular methods, many of which have recently given way to high-throughput array-based technologies. The following sections will briefly review important developments in molecular technology related to cancer research, which formed the basis for the newer (and now widely-utilized) array-based techniques and applications.

Cytogenetics

Cytogenetics is used to investigate gross structural abnormalities of human chromosomes and their contribution to disease. Nowell and Hungerford [1960] identified an abnormal chromosome 22 in select chronic myelogenous leukemia patients, calling this chromosome the Philadelphia (Ph) chromosome. This was the first consistent chromosomal abnormality identified in human cancer. With the development of high-resolution chromosome banding techniques [Sumner et al., 1971], it was determined that the Ph chromosome results from a reciprocal translocation between chromosome 22 and chromosome 9 [Rowley, 1973]. Prior to the development of chromosome banding techniques, the nature of this translocation could not be determined [Rowley, 1973]. Following the discovery of the Ph chromosome, numerous other chromosomal abnormalities in cancer were identified and characterized using G-banding and R-banding techniques [Mitelman, 1986]. More recently, fluorescent chromosome painting techniques were described. These techniques facilitate the easy and accurate identification of each chromosome pair, making

the detection and characterization of chromosomal abnormalities relatively straight forward [Sumner et al., 1971]. These methods can (i) detect numerical changes of chromosomes, (ii) identify large deletions, (iii) detect and characterize complex chromosomal rearrangements. Despite the power of chromosomal analysis (through traditional banding techniques or fluorescent methods), a number of technical limitations affect its general application to cancer research. Foremost, cytogenetic analysis has limited sensitivity. These methods can only detect large genomic alterations. In addition, cytogenetic methods require actively dividing cells to facilitate capture of cells in metaphase. Thus, cytogenetic laboratories must transiently culture tumor cells to enable good chromosomal preparations. For many diseases with chromosome abnormalities, cytogenetics remains the gold standard for diagnosis. However, Polymerase chain reaction (PCR)-based and array-based methods have (or will soon) replaced many of these.

Genomics

Southern blotting. The technique of Southern blotting was introduced in 1975 [Southern, 1975], revolutionized early molecular biology, and continues to be widely used today. Applications of the Southern blot include evaluation of gene copy number, assessment of genetic alterations (such as chromosomal rearrangements), and gene/chromosome mapping. Southern blotting has been widely applied to the analysis of chromosomal alterations in cancer. Numerous studies have reported proto-oncogene amplification detected by Southern analysis. Amplification of the *c*-ras and *c*-myc proto-oncogenes has been reported in many forms of cancer [Taya et al., 1984; Tashiro et al., 1986]. Additional examples of gene amplification can be found for specific cancers, including Her2 amplification in breast cancer [Hynes et al., 1989; Slamon et al., 1989]. This approach is useful in characterization of chromosomal rearrangements (reciprocal and non-reciprocal translocations) as well as large-scale chromosomal deletions (often measured as loss of heterozygosity). All of the early genome mapping studies utilized Southern blotting, prior to the advent of PCR-based mapping techniques. While Southern blotting represents a powerful method for investigation of the molecular pathogenesis of cancer, its application is hampered by certain limitations. Specifically, Southern blotting is slow and labor-intensive, is limited by the number of samples that can be simultaneously analyzed, and requires high quality, high molecular weight DNA. Nevertheless, Southern blotting remains a useful technique that continues to find application in both research and clinical laboratories.

Polymerase chain reaction (PCR). PCR represents one of the most valuable advances in molecular technology ever. This method facilitates the rapid, sensitive, and specific amplification of target DNA (or cDNA) sequences. A large number of molecular applications have been developed that are based upon PCR amplification of target sequences. These include applications for genomic sequencing, cloning, analysis of gene expression, chromosomal mapping, and others. Through microsatellite PCR, chromosomal loss or retention can be determined, localizing tumor suppressor genes to specific chromosomal regions [Negrini et al., 1994; Reid et al., 1996a; Ricketts, 2006]. Quantitative PCR is used for the evaluation of gene copy number and can provide an alternative to Southern blotting. Quantitative PCR is useful when reproducibility is important since the accumulation of amplified gene products occurs exponentially and follows a conventional curve [Ricketts, 2006]. PCR has become an important method for research and clinical molecular biology applications. The advantages of PCR include speed, sensitivity (only a small amount of template DNA is needed), and specificity. This technique is also very flexible and can be adapted to a variety of analyses.

Mutation detection. It is well-known that cancer cells contain numerous DNA mutations representing either chromosomal aberrations or sequence alterations. DNA sequence alterations often affect gene expression or result in the expression of a defective gene product. Thus, methods to identify gene mutations are very important [Highsmith, 2006]. DNA sequencing represents the gold standard for identification and characterization of gene mutations. The common availability of automated fluorescent DNA sequencers has increased the ease, speed, and accuracy of DNA sequencing compared to manual sequencing methods. Nevertheless, DNA sequencing is inefficient when applied to mutation screening of large numbers of

individual DNA samples and when used as a primary means of mutation discovery [Highsmith, 2006]. Thus, DNA scanning techniques were developed to identify DNA regions that are likely to harbor a sequence alteration. These methods include single-strand conformation polymorphism (SSCP) analysis and heteroduplex analysis. SSCP is utilized to detect mutations in a DNA sequence when the mutation influences the secondary structure of the DNA or disrupts single stranded DNA conformation [Hayashi, 1992]. SSCP for capillary electrophoresis identified mutations in the *p53* gene in many hereditary and spontaneous cancers [Bosserhoff et al., 2000]. Likewise, heteroduplex analysis exploits the altered migration of DNA heteroduplexes containing an altered sequence to identify putative mutants [Highsmith, 2006]. In both cases, the screening analysis is followed by DNA sequencing of the DNA segment to identify the precise location and nature of the sequence alteration. In contrast to mutation screening, methods for detection of known mutations have also been developed. These methods include allele-specific PCR and modified restriction fragment length polymorphism (RFLP) analysis [Highsmith, 2006]. These methods are useful when a specific mutation is suspected. In both cases, the nature of the specific mutation must be known, enabling the design of appropriate PCR primers. These methods are most frequently employed in the analysis of hot-spot (or other frequently occurring) mutations, like those identified in *p53* [Gao et al., 2006] and *BRCA1* [Yim et al., 2005]. Most of these methods are still used commonly, but often to verify results obtained using discovery-based high-throughput technologies.

Gene Expression Analysis

Northern blotting. Northern blot analysis is a method for quantitative analysis of gene expression [Amiss and Presnell, 2006]. Using Northern blot analysis, various characteristics of mRNA can be examined, including: mRNA abundance and mRNA size. Thus, northern blots are able to assess gene deletions affecting the coding region, post-transcriptional modifications such as splicing variations or abnormalities, and production of mutated RNA resulting from premature termination of transcription [Amiss and Presnell, 2006]. Perhaps the most important contribution of Northern blotting to the analysis of cancer was in the generation of early gene expression signatures. The advantages of Northern blot include sensitivity and quantitive estimation of mRNA abundance. The limitations of Northern blot analysis relates to the need for relatively large amounts of RNA and the fairly slow and labor-intensive methodology. Nevertheless, Northern blots have been and continue to be useful for certain specific applications.

Amplification-based gene expression analysis. Following the development of PCR in the mid-1980s [Mullis and Faloona, 1987], a number of amplification-based applications for gene expression analysis emerged and were widely adapted by research and clinical laboratories. The most common of these is reverse transcription-PCR (RT-PCR), a method based upon amplification of target sequences contained in a cDNA template population. RT-PCR is rapid, extremely sensitive and specific, requires very little template RNA for the initial reverse transcription reaction, and can be adapted for multiplex amplification. While this method is still widely used, quantitative PCR (qPCR) has become the more popular version of this technique. qPCR has all of the advantages of RT-PCR, along with several additional features, including a quantitative output and an automated process. RT-PCR and qPCR are employed to assess the expression of specific genes, but are of limited value in a study aimed at gene discovery. In contrast, differential display RT-PCR (RT-PCR/DD) was developed to facilitate gene discovery. Using RT-PCR/DD, the gene expression pattern corresponding to two or more populations of cells (or tissues) can be compared to identify differentially expressed genes that may be of significance (such as when a cancer cell is compared to a normal cell). RT-PCR/DD has been successful in identifying genes differentially expressed in several cancer systems [Liang and Pardee, 1992]. The limitations associated with RT-PCR/DD include (i) the large number of false positives identified that fail validation with Northern blot or quantitative RT-PCR (are not truly differentially expressed), and (ii) the 3′-untranslated regions of genes are favorably amplified and may not match the database sequence for that gene [Sunday, 1995]. Nevertheless, RT-PCR/DD is still widely used for comparing molecular gene expression patterns between cancer and normal cells.

BIOINFORMATIC-BASED EXPERIMENTAL APPROACHES

The molecular biology methods developed over the last several decades allow specific questions to be asked about genomic DNA and the genes encoded, as well as facilitating gene discovery and mutation screening. However, the major drawback for all of these techniques relates to the quantities of starting material needed, the time required to complete the procedure (labor-intensive), and limited amount of information generated (mostly reflecting limited numbers of samples analyzed). The methods that now represent the state-of-the-art are characterized by large-scale high-throughput testing. These methods (many of which are array-based) enable investigators to concurrently conduct thousands of parallel experiments (tests) starting with very small quantities of DNA or RNA through largely automated processes. The following sections describe some of these newer technologies and how they have been applied to cancer research.

Gene Target and Biomarker Discovery

In order to elucidate the complex molecular pathogenesis of cancer, whole genome analysis is required to catalog changes associated with cancer initiation, progression, and metastasis. Whole genome analyses may ultimately uncover novel gene targets for drug development and disease biomarker discovery for diagnostic applications. Genomic DNA microarrays coupled with bioinformatic approaches have been used as a tool for the identification of gene expression alterations and genomic modifications that correspond to various stages of neoplastic transformation and tumor development. Consequently, DNA microarray technology has been exceptionally important to the field of molecular carcinogenesis research. There are a number of bioinformatic-based approaches for mining data derived from high-throughput information-rich methods utilized for identification of molecular targets in cancer and elucidation of molecular pathways that account for the neoplastic phenotype.

Comparative genomic hybridization. It is well-known that cancer cells carry numerous chromosomal abnormalities, and that some of these abnormalities are causally related to cancer induction or progression. The chromosomal changes associated with neoplasia manifest as gains (gene amplification) or losses (large-scale deletions) of chromosomal segments. Methods to characterize these alterations have been applied for many years, but recent developments have automated this process and opened the analysis to difficult samples (solid tumors). Array-based comparative genomic hybridization (CGH; [Kallioniemi et al., 1992]) is a technique that analyzes global genomic changes by documenting gains and losses of genes in diseases such as cancer [Albertson et al., 2000; Jain et al., 2001; Bignell et al., 2006]. CGH arrays are constructed using thousands of probes that correspond to defined and mapped chromosomal segments. To facilitate analysis of cancer cell genomes, total genomic DNA from a tumor of interest and normal (control) cell population are labeled with different fluorochromes and hybridized to the CGH array [Albertson et al., 2000]. The ratio of fluorescence intensity for a given probe (or set of probes) is proportional to the copy number of the corresponding sequences in the tumor and can be measured over the length of the chromosomes to determine regions of gain or loss [Albertson et al., 2000; Medvedovic and Wiest, 2006]. CGH arrays have been applied to the analysis of chromosomal alterations in gastric cancer [Weiss et al., 2003a], chronic lymphocytic leukemia [Schwaenen et al., 2004], breast cancer [Albertson et al., 2000; van Beers and Nederlof, 2006], oral squamous carcinoma [Garnis et al., 2004], bladder cancer [Veltman et al., 2003], pancreatic cancer [Holzmann et al., 2004], colon cancer [Kleivi et al., 2004], lung cancer [Choi et al., 2006b], and prostate cancer [Saramaki et al., 2006]. Chromosomal alterations identified by CGH array can uncover large-scale deletions (possibly associated with tumor suppressor genes), gene amplifications (possibly associated with proto-oncogenes), and/or may identify alterations that predict tumor behavior of patient outcome. Albertson et al. [2000] identified recurrent amplification at 20q13.2 in breast cancer, which mapped to the *CYP24* gene. In addition, CGH data from a set of 52 human breast tumors showed two loci (8q24 and 9q13) where copy number abnormalities were correlated with poor survival and also identified a relationship between *p53* mutation status and two loci (8q24 and 5q15-5q21) [Jain et al., 2001]. Array CGH has significant advantages over conventional cytogenetic approaches including high-throughput, high-resolution,

and the ability to analyze interphase cells, although it has been limited by the availability of genomic clones that can be spotted as targets [Ishkanian et al., 2004; Shih Ie and Wang, 2005]. Recently, high-density (1 Mb) CGH arrays and new analysis tools have been established in order to provide a more uniform and convenient analysis platform for array CGH [Wang et al., 2004; Shih Ie and Wang, 2005].

Representational oligonucleotide microarray analysis. Representational oligonucleotide microarray analysis (ROMA) is an array-based technique used for the detection of genomic aberrations in cancer based upon copy number variation [Lucito et al., 2003; Shih Ie and Wang, 2005]. ROMA measures the DNA concentration of two samples, generated by restriction digestion, by hybridizing two differently labeled samples to a set of oligonucleotide microarray probes (designed in silico) from the human genome sequence [Sebat et al., 2004; Shih Ie and Wang, 2005]. Using ROMA, chromosomal regions with copy number variations have been identified between cancer and normal genomes [Sebat et al., 2004; Shih Ie and Wang, 2005]. The resolution of ROMA is between 30 kb to 35 kb, and with further refinement ROMA promises to contribute more substantially to the identification of genes involved in disease [Shih Ie and Wang, 2005].

Tiled oligonucleotide microarray. Tiled oligonucleotide microarrays allow investigators to interrogate previously unexplored regions of chromosomal DNA at 35-nucleotide resolution using 25 bp probes [Carroll et al., 2005]. Carroll et al. used tiled microarrays that cover the entire sequence of chromosomes 21 and 22 in combination with chromatin immunoprecipitation (ChIP) analyses in order to identify regulatory regions that may play a role in ER-mediated transcription. The Forkhead protein FoxA1 was subsequently identified in estrogen signaling which exhibits the power and importance of microarray-based approaches to identify novel regulatory domains that prove to be significant in human cancer [Carroll et al., 2005]. Selzer et al. [2005] used fine-tiling arrays in conjunction with oligonucleotide array CGH to map chromosomal imbalances found in neuroblastoma tumors and cell lines. First, CGH analysis was performed to map chromosomal alterations using whole-genome oligonucleotide or BAC arrays at 50 kb or 1 Mb resolution. The results of the CGH analyses were utilized to design fine-tiling oligonucleotide microarrays spanning the chromosome region of interest [Selzer et al., 2005]. Thus, the combination of CGH to localize chromosomal regions of interest and the high level of resolution of fine tiling array CGH facilitates detailed mapping of chromosomal abnormalities in cancer cells.

Single-nucleotide polymorphism arrays. The identification of both loss-of-heterozygosity events and copy-number alterations in cancer, previously accomplished by PCR and Southern blot techniques, can be identified by high-density single-nucleotide polymorphism (SNP) arrays [Garraway et al., 2005]. Garraway et al. [2005] used 100 K SNP arrays, containing 124,000 SNP alleles spaced with a median intermarker distance of 8.5 kb, to evaluate the genomes of cell lines that represent tumors from nine different tissue types. Hierarchical clustering was used to determine patterns of chromosomal copy number alterations and identified distinct genetic subgroups between the tumor cell lines. Within one amplified chromosomal region, the melanocyte master regulator (*MITF*) was overexpressed and may offer a therapeutic target for melanoma [Garraway et al., 2005]. SNP arrays afford the opportunity for parallel analysis of thousands of markers using very small amounts of starting material, resulting in a more rapid and information-rich assessment of genomic changes through an automated process.

Molecular Classification of Cancer

Improvements in molecular tumor classification have been essential to the advances seen in cancer treatment. Historically, the classification of cancer types has been primarily (or exclusively) based upon morphological appearance of the tumor [Golub et al., 1999]. However, traditional approaches to tumor classification have serious limitations given that tumors with similar histopathological appearance can result in different clinical courses and exhibit a wide range of responses to therapy [Golub et al., 1999]. It is now recognized that tumors with common behavior (phenotype) express a common gene expression signature and that classification of tumors based upon their molecular signatures is much more useful for predicting patient outcome and response to therapy than morphological characterization [Chung et al.,

2002]. Numerous analytical methods have been used to study tumors and classify them into similar groups to predict clinical behavior [Chung et al., 2002]. DNA microarrays have revolutionized and significantly improved this field through the simultaneous analysis of numerous gene expression patterns that detect similarities and differences among tumors [Chung et al., 2002].

DNA microarrays coupled with statistical analysis have provided investigators tools in which to develop molecular classifications for many types of cancer [Chung et al., 2002], including brain [Pomeroy et al., 2002], breast [Perou et al., 2000; Sorlie et al., 2001; van 't Veer et al., 2002], colon [Zou et al., 2002], stomach [Hippo et al., 2002], leukemia [Yeoh et al., 2002], lung [Beer et al., 2002], lymphoma [Alizadeh et al., 2000; Rosenwald et al., 2002; Shipp et al., 2002], and kidney [Takahashi et al., 2001]. Golub et al. [1999] recognized that there was a lack of a general approach for identifying new cancer types, or for assigning tumors to known classes (called class discovery and class prediction, respectively). Therefore, gene expression analysis using Affymetrix microarrays were applied to human acute leukemias as a test case and the results demonstrated that cancer classification based on gene expression was attainable [Golub et al., 1999]. Thus, DNA microarrays proved to be useful as a tool for discovering and predicting cancer classes.

Several DNA microarray studies have demonstrated that different types of cancer represent functionally distinct types of cells and each type of cell expresses a unique set of genes that is needed to produce the cancer phenotype [Chung et al., 2002]. Using between 10 and 19 different tissues, tumor- and tissue-specific portraits of gene expression have been defined by microarray technology [Ramaswamy et al., 2001; Chung et al., 2002]. Ramaswamy et al. [2001] examined 144 primary tumors representing 14 different tumor types using Affymetrix Genechips (Hu6800 and Hu35k-subA) and defined a set of genes that identified each of the 14 tissues separately. Therefore, each tumor type was represented by a gene signature pattern, consequently, classifying each cancer by a molecular pattern. Perou et al. [2000] was the first to propose a phenotypic diversity of breast tumors and examined the gene expression patterns of 65 surgical specimens of human breast tissue from 42 different individuals using custom cDNA microarrays representing 8,102 human genes. The human breast tumors were classified into subtypes (luminal, normal-breast-like, basal-like, HER2+) distinguished by differences in their gene expression signatures [Perou et al., 2000]. Head and neck squamous cell carcinoma (HNSCC) clinical status is mostly based upon the tumor size and the presence and location of metastasis of lymph nodes [Chung et al., 2004]. Utilizing an Agilent Human 1 cDNA microarray, Chung et al. [2004] analyzed the gene expression patterns of 60 HNSCC tumors and identified four distinct subtypes. Additionally, patterns of gene expression that are associated with prediction of metastatic disease in breast cancer also predicts the presence of lymph node metastases in HNSCC [Chung et al., 2004]. Interestingly, in a previous study by Chung et al. [2002] hierarchical clustering analysis of lung and breast [Perou et al., 2000; Sorlie et al., 2001] tumor data sets were analyzed using a combined lung and breast "intrinsic" gene sets and gene expression patterns identified sets of genes involved in regulating the cell cycle, DNA replication, and genes that encode proteins responsible for chromosome dynamics. The application of DNA microarrays to tumor classification shows great promise, but is only in the beginning stages. The studies that have classified various types of tumors based on gene expression patterns are ready to explore the important quest of whether tumor classes based upon molecular patterns can be used to improve cancer treatment and patient outcome [Chung et al., 2002]. Microarrays will certainly continue to be used in cancer research to explore molecular portraits of tumors, although it is uncertain whether microarrays could be used routinely in the clinic for diagnostic purposes [Chung et al., 2002]. Nevertheless, the knowledge gained through microarray studies may be harnessed through other methods (like immunohistochemistry) to accomplish tumor class predictions.

Improving Cancer Diagnosis and Predicting Clinical Outcomes

Whole genome genotyping using DNA microarrays hold promise in identifying gene expression signatures and locating gene variants that distinguish cancer subtypes, offering important prognostic indications for cancer outcome, as well as patient response to clinical treatment.

Studies on breast cancer [Lonning et al., 2001; Sorlie et al., 2001; van 't Veer et al., 2002; van de Vijver et al., 2002; Glinsky et al., 2004; Weigelt et al., 2005], diffuse large B-cell lymphoma (DLBCL) [Chung et al., 2002], HNSCC [Chung et al., 2004], mesothelioma [Gordon et al., 2003], renal cell carcinoma (RCC) [Zhao et al., 2006], and lymphoblastic leukemia [Armstrong et al., 2002; Teuffel et al., 2004] are a few examples of cancers in which gene expression patterns have been developed, and utilized to predict prognosis and treatment options. These insights into the molecular mechanisms of cancer have opened the door for better prognostic tests and treatment that can be used to customize therapy for better patient care.

Breast cancer. Breast cancer has been studied extensively, although the biology of breast cancer remains poorly understood. Lymph node metastases [Fisher et al., 1993], histologic grade [Elston and Ellis, 1991], expression of steroid and growth factor receptors [Vollenweider-Zerargui et al., 1986; Torregrosa et al., 1997], expression status of *ERBB2* [Slamon et al., 1989], mutations in the *TP53* gene [Bergh et al., 1995; Borresen et al., 1995], and estrogen-inducible genes [Foekens et al., 1999] have all been associated with prognosis of breast cancer. In a large custom cDNA microarray study, Sorlie et al. [2001] analyzed 78 breast cancers, three fibroadenomas, and four normal breast tissues by hierarchical clustering and classified the tumors based on gene expression in the previously identified subgroups [Perou et al., 2000]: basal epithelial-like group, an *ERBB2*-overexpressing group, luminal-like/ER+ group, and a normal breast-like group. These breast tumor subgroups have prognostic value with respect to overall and relapse-free survival in a subset of patients that had been given a uniform therapy [Sorlie et al., 2001]. The *ERBB2*+ group was associated with poor survival, the basal-like subtype was associated with shorter survival times and a higher frequency of *TP53* mutations, and the luminal C subtype is associated with a gene expression pattern similar to the *ERBB2*+ group, which results in the worst outcome [Sorlie et al., 2001].

It has been shown that chemotherapy or hormonal therapy reduces the risk of distant metastases by about one-third [van 't Veer et al., 2002]. A study by van 't Veer et al. [2002] performed microarray (Affymetrix Hu25k) analysis on 117 young patients with breast cancer, applied supervised tumor classification, and identified a gene expression pattern that strongly predicts poor prognosis (indicated by a short interval to distant metastases) in patients that were initially lymph node negative. Additionally, patients that are mutant *BRCA1*carriers exhibited a distinct gene expression signature [van 't Veer et al., 2002]. The gene signature of poor prognosis tumors consisted of 70 genes regulating cell cycle, angiogenesis, and metastasis [van 't Veer et al., 2002]. Recently, a number of investigators [van de Vijver et al., 2002; Glinsky et al., 2004; Weigelt et al., 2005] have characterized the 70-gene prognosis profile in greater depth. van de Vijver et al. [2002] evaluated the predictive power of the prognosis profile using statistical methods (univariable and multivariable) and a DNA microarray study consisting of a series of 295 consecutive patients less than 53 years old, and found that the 70-gene profile is a more powerful predictor of the outcome of disease in young patients with breast cancer than clinical and histologic criteria. Glinsky et al. [2004] described a breast cancer classification algorithm that takes into account estrogen receptor status, lymph node status, and the 70-gene expression signature, and suggested that the small gene cluster may be useful in stratification of breast cancer patients into subgroups with positive outcomes and helpful in selection of optimal treatment strategies. Using this signature of 70-genes defined by DNA microarrays [van 't Veer et al., 2002], recent data suggest that expression of these genes reflects a distinct disease signature that is continuous throughout the metastatic process [Weigelt et al., 2005], adding to the prognostic power of the molecular signature.

Diffuse large B-cell lymphoma. Microarray analysis has provided molecular insight into the differential stages of many hematological malignancies [Chung et al., 2002]. Both Rosenwald et al. [2002] and Alizadeh et al. [2000] utilized a custom array (Lymphochip) which produced a molecular signature from hundreds of DLBCL samples, and identified three subtypes including a germinal center B cell-like (GCBL) subtype, and activated B cell-like (ABL) subtype, and a subtype lacking high expression of either the GCBL- or the ABL-defining genes. Sixteen genes were recognized for their ability to accurately predict patient outcome and survival, when used to

supervise tumor cluster analysis [Rosenwald et al., 2002]. In an additional study by Shipp et al. [2002], a 13-gene outcome predictor pattern was identified using Affymetrix Hu6800 microarray technology with DLBCLs.

Head and neck squamous cell carcinoma. HNSCC is treated aggressively with surgery, chemotherapy, and radiation therapy, but 40–50% of patients with advanced disease develop recurrent disease. Therefore, gene biomarkers that predict poor clinical outcome are essential in assisting treatment of HNSCC patients. Chung et al. [2004] analyzed gene expression patterns of 60 HNSCC tumors, categorizing these tumors into four distinct subtypes with different clinical outcomes, including (i) an EGFR-pathway signature, (ii) a mesenchymal-enriched subtype, (iii) a normal epithelium-like subtype, and (iv) a subtype with high levels of antioxidant enzymes. An HNSCC poor clinical outcome subgroup showed high expression of some of the same genes as the basal-like subgroup identified in breast cancer [Perou et al., 2000], including *Bullous Pemphogoid Antigen 1*, *P-Cadherin*, *Laminin γ 2*, and *Collagen XVII-α* [Sorlie et al., 2001]. Tumors in both breast and HNSCC showing the high expression of these four genes exhibit poor patient outcomes [Chung et al., 2004]. It is interesting to observe the possibility that these genes may be causative of poor outcome, or that they may be expressed in association with an aggressive tumor phenotype [Chung et al., 2004].

Mesothelioma. There are no prognostic molecular markers or genetic abnormalities that are predictive of patient outcome for mesothelioma. Production of survival among mesothelioma patients is based upon histological appearance of the tumor, which can be somewhat subjective [Gordon et al., 2003]. Gordon et al. [2003] used gene expression profile data from a Affymetrix U95A microarray study previously performed on 17 mesothelioma tumors to predict treatment-related outcome. A significant four-gene expression ratio test predicted mesothelioma treatment-related patient outcome [Gordon et al., 2003]. Application of this four-gene expression ratio test in the routine clinical workup of mesothelioma could impact the clinical treatment of these patients.

Renal cell carcinoma. Microarrays have been useful in identifying gene expression patterns that correlate with survival of RCCs. RCC is responsible for the majority of deaths due to kidney cancer, with tumor stage, grade, and patient performance status currently used to predict survival after surgery [Zhao et al., 2006]. Zhao et al. [2006] used comprehensive gene expression profiling (Stanford Functional Genomics Facility) to identify gene expression patterns that correlate with survival. A number of primary RCCs were subdivided into five gene expression subgroups that correlated with survival in long-term follow-up [Zhao et al., 2006]. Using a semisupervised learning algorithm, Zhao et al. [2006] identified a group of 259 genes that accurately predict disease-specific survival among patients after surgery independent of clinical prognostic factors.

Lymphoblastic leukemia. Subsets of human acute lymphoblastic leukemias (ALL) that contain a chromosomal translocation harboring the mixed-lineage leukemia gene (*MLL*, *HRX*, *ALL1*) have an unfavorable outcome. Clustering algorithms from Affymetrix U95A or U95A V2 gene expression data revealed *MLL* translocations in lymphoblastic leukemias could be separated from conventional acute lymphoblastic and acute myelogenous leukemias [Armstrong et al., 2002]. A study in twins that have ALL and harbor the *TEL-AML1* fusion gene very early in development demonstrate, by Affymetrix HGU133A microarray gene expression profiling, that leukemia in twins show the same subtype-typical feature as *TEL-AML1*-positive leukemia in individuals suggesting that the leukemogenesis model in children might be applicable generally [Teuffel et al., 2004].

Major Implications of DNA Microarray Technology

An important implication of DNA microarray technology is that primary tumor gene expression profiles can be utilized to assess properties of the tumor, such as the propensity to metastasize or patient's treatment response, which can be used as the basis for clinical decision-making. For example, a patient presents with a primary tumor that exhibits a "good prognosis" gene signature and the follow-up may be surgical resection with no chemotherapy, but close observations. Alternatively, a patient with a "poor prognosis" gene pattern may have combined treatments including surgery, radiation, and/or chemotherapy [Chung et al., 2002].

Limitations of Bioinformatic-Based Experiments (DNA Microarrays)

While high-throughput microarray technology confers many advantages and opportunities for advancement in molecular tumor signatures leading to better patient therapy, there are limitations to this experimental approach. DNA microarrays analyze global gene expression changes within a cell population and are limited by (i) cost and access, (ii) quality and amount of RNA, (iii) standardization in sample collection, (iv) cell type heterogeneity, (v) replication to eliminate error, (vi) small tissue size of clinical specimens, and (vii) interpretation of data [Russo et al., 2003]. Significant advances have been made to eliminate these limitations and progress is continuing in order to bring this technology into the clinical laboratories.

CONCLUSIONS

It is the promise of high-throughput molecular technologies (DNA microarray technology) along with the use of bioinformatics that make it possible to tackle and understand the complexity of the molecular basis of human cancers. With the completion of the human genome project, it is estimated that 30,000–40,000 human genes exist within our cells. However, the number of genes expressed within a certain cell type may reflect a much smaller number, making it technically possible to assess most of these genes (transcriptional profiling) with a tumor sample [Pusztai et al., 2003]. High-throughput molecular information is essential in developing gene biomarkers of clinical utility that can be used to formulate approaches to therapy, as well as improvements in cancer diagnosis and prognosis [Mohr et al., 2002; Russo et al., 2003]. A major goal of this technology is to introduce individualized medicine assuring its establishment into clinical practice, and appropriate future clinical strategies for patients to receive suitable therapy that will provide the most beneficial outcome [Mohr et al., 2002; Russo et al., 2003]. With the aid of molecular gene signatures and bioinformatic-based technology, treatment options will be based upon differential gene expression patterns and can be potentially monitored while the patient undergoes therapy (hormonal, chemotherapy and/or radiation). Bioinformatic-based high-throughput technology has been exceptionally important to the field of molecular cancer research, and hopefully will successfully alter the management of cancer.

REFERENCES

Albertson DG, Ylstra B, Segraves R, Collins C, Dairkee SH, Kowbel D, Kuo WL, Gray JW, Pinkel D. 2000. Quantitative mapping of amplicon structure by array CGH identifies CYP24 as a candidate oncogene. Nat Genet 25:144–146.

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Jr., Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. 2000. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403:503–511.

Amiss T, Presnell SC. 2006. Nucleic acid blotting techniques: Theory and practice. In: Coleman WB, Tsongalis GJ, editors. Molecular diagnostics: For the clinical laboratorian. Totowa: Humana Press Inc. p 31–46.

Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, Sallan SE, Lander ES, Golub TR, Korsmeyer SJ. 2002. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet 30:41–47.

Beer DG, Kardia SL, Huang CC, Giordano TJ, Levin AM, Misek DE, Lin L, Chen G, Gharib TG, Thomas DG, Lizyness ML, Kuick R, Hayasaka S, Taylor JM, Iannettoni MD, Orringer MB, Hanash S. 2002. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 8:816–824.

Bergh J, Norberg T, Sjogren S, Lindgren A, Holmberg L. 1995. Complete sequencing of the p53 gene provides prognostic information in breast cancer patients, particularly in relation to adjuvant systemic therapy and radiotherapy. Nat Med 1:1029–1034.

Bignell GR, Huang J, Greshock J, Watt S, Butler A, West S, Grigorova M, Jones KW, Wei W, Stratton MR, Futreal PA, Weber B, Shapero MH, Wooster R. 2006. High-resolution analysis of DNA copy number using oligonucleotide microarrays. Genome Research 14:287–295.

Borresen AL, Andersen TI, Eyfjord JE, Cornelis RS, Thorlacius S, Borg A, Johansson U, Theillet C, Scherneck S, Hartman S, et al. 1995. TP53 mutations and breast cancer prognosis: Particularly poor survival rates for cases with mutations in the zinc-binding domains. Genes Chromosomes Cancer 14:71–75.

Bosserhoff AK, Buettner R, Hellerbrand C. 2000. Use of capillary electrophoresis for high throughput screening in biomedical applications. A minireview. Comb Chem High Throughput Screen 3:455–466.

Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR, Fox EA, Silver PA, Brown M. 2005. Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 122:33–43.

Choi YW, Choi JS, Zheng LT, Lim YJ, Yoon HK, Kim YH, Wang YP, Lim Y. 2006b. Comparative genomic hybridization array analysis and real time PCR reveals genomic

alterations in squamous cell carcinomas of the lung. Lung Cancer 55:43–51.
Chung CH, Bernard PS, Perou CM. 2002. Molecular portraits and the family tree of cancer. Nat Genet 32(Suppl):533–540.
Chung CH, Parker JS, Karaca G, Wu J, Funkhouser WK, Moore D, Butterfoss D, Xiang D, Zanation A, Yin X, Shockley WW, Weissler MC, Dressler LG, Shores CG, Yarbrough WG, Perou CM. 2004. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. Cancer Cell 5:489–500.
Elston CW, Ellis IO. 1991. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: Experience from a large study with long-term follow-up. Histopathology 19:403–410.
Fisher ER, Costantino J, Fisher B, Redmond C. 1993. Pathologic findings from the National Surgical Adjuvant Breast Project (Protocol 4). Discriminants for 15-year survival. National Surgical Adjuvant Breast and Bowel Project Investigators. Cancer 71:2141–2150.
Foekens JA, Look MP, Bolt-de Vries J, Meijer-van Gelder ME, van Putten WL, Klijn JG. 1999. Cathepsin-D in primary breast cancer: Prognostic evaluation involving 2810 patients. Br J Cancer 79:300–307.
Gao F, Li C, Li W, Chen L, Tang Z, Tang W. 2006. p53 Gene Mutations in Sporadic Colorectal Carcinoma in Guangxi Region. Cancer Invest 24:689–695.
Garnis C, Coe BP, Zhang L, Rosin MP, Lam WL. 2004. Overexpression of LRP12, a gene contained within an 8q22 amplicon identified by high-resolution array CGH analysis of oral squamous cell carcinomas. Oncogene 23:2582–2586.
Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, Beroukhim R, Milner DA, Granter SR, Du J, Lee C, Wagner SN, Li C, Golub TR, Rimm DL, Meyerson ML, Fisher DE, Sellers WR. 2005. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature 436: 117–122.
Glinsky GV, Higashiyama T, Glinskii AB. 2004. Classification of human breast cancer using gene expression profiling as a component of the survival predictor algorithm. Clin Cancer Res 10:2272–2283.
Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. 1999. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 286:531–537.
Gordon GJ, Jensen RV, Hsiao LL, Gullans SR, Blumenstock JE, Richards WG, Jaklitsch MT, Sugarbaker DJ, Bueno R. 2003. Using gene expression ratios to predict outcome among patients with mesothelioma. J Natl Cancer Inst 95:598–605.
Hayashi K. 1992. PCR-SSCP—rapid and easy detection of DNA-sequence changes. Hum Cell 5:180–184.
Highsmith EW. 2006. Electrophoretic methods for mutation detection and DNA sequencing. In: Coleman WB, Tsongalis GJ, editors. Molecular diagnostics: For the clinical laboratorian. Totowa: Humana Press Inc. p 85–109.
Hippo Y, Taniguchi H, Tsutsumi S, Machida N, Chong JM, Fukayama M, Kodama T, Aburatani H. 2002. Global gene expression analysis of gastric cancer by oligonucleotide microarrays. Cancer Res 62:233–240.
Holzmann K, Kohlhammer H, Schwaenen C, Wessendorf S, Kestler HA, Schwoerer A, Rau B, Radlwimmer B, Dohner H, Lichter P, Gress T, Bentz M. 2004. Genomic DNA-chip hybridization reveals a higher incidence of genomic amplifications in pancreatic cancer than conventional comparative genomic hybridization and leads to the identification of novel candidate genes. Cancer Res 64: 4428–4433.
Hynes NE, Gerber HA, Saurer S, Groner B. 1989. Overexpression of the c-erbB-2 protein in human breast tumor cell lines. J Cell Biochem 39:167–173.
Ishkanian AS, Malloff CA, Watson SK, DeLeeuw RJ, Chi B, Coe BP, Snijders A, Albertson DG, Pinkel D, Marra MA, Ling V, MacAulay C, Lam WL. 2004. A tiling resolution DNA microarray with complete coverage of the human genome. Nat Genet 36:299–303.
Jain AN, Chin K, Borresen-Dale AL, Erikstein BK, Eynstein Lonning P, Kaaresen R, Gray JW. 2001. Quantitative analysis of chromosomal CGH in human breast tumors associates copy number abnormalities with p53 status and patient survival. Proc Natl Acad Sci USA 98:7952–7957.
Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, Pinkel D. 1992. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science 258:818–821.
Kleivi K, Teixeira MR, Eknaes M, Diep CB, Jakobsen KS, Hamelin R, Lothe RA. 2004. Genome signatures of colon carcinoma cell lines. Cancer Genet Cytogenet 155:119–131.
Liang P, Pardee AB. 1992. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257:967–971.
Lonning PE, Sorlie T, Perou CM, Brown PO, Botstein D, Borresen-Dale AL. 2001. Microarrays in primary breast cancer—lessons from chemotherapy studies. Endocr Relat Cancer 8:259–263.
Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, West JA, Rostan S, Nguyen KC, Powers S, Ye KQ, Olshen A, Venkatraman E, Norton L, Wigler M. 2003. Representational oligonucleotide microarray analysis: A high-resolution method to detect genome copy number variation. Genome Res 13:2291–2305.
Medvedovic M, Wiest JS. 2006. DNA microarrays and computational analysis of DNA microarray data in cancer research. In: Warshawsky D, Landolph J, Jr., editors. Molecular carcinogenesis and the molecular biology of human cancer. Boca Raton: CRC Press Taylor & Francis Group. p 243–262.
Mitelman F. 1986. Clustering of breakpoints to specific chromosomal regions in human neoplasia. A survey of 5,345 cases. Hereditas 104:113–119.
Mohr S, Leikauf GD, Keith G, Rihn BH. 2002. Microarrays as cancer keys: An array of possibilities. J Clin Oncol 20:3165–3175.
Mullis KB, Faloona FA. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods Enzymol 155:335–350.
Negrini M, Sabbioni S, Possati L, Rattan S, Corallini A, Barbanti-Brodano G, Croce CM. 1994. Suppression of tumorigenicity of breast cancer cells by microcell-mediated chromosome transfer: Studies on chromosomes 6 and 11. Cancer Res 54:1331–1336.

Nowell PC, Hungerford DA. 1960. A minute chromosome in human chronic granulocytic leukemia. Science 132:1497.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D. 2000. Molecular portraits of human breast tumours. Nature 406:747–752.

Pomeroy SL, Tamayo P, Gaasenbeek M, Sturla LM, Angelo M, McLaughlin ME, Kim JY, Goumnerova LC, Black PM, Lau C, Allen JC, Zagzag D, Olson JM, Curran T, Wetmore C, Biegel JA, Poggio T, Mukherjee S, Rifkin R, Califano A, Stolovitzky G, Louis DN, Mesirov JP, Lander ES, Golub TR. 2002. Prediction of central nervous system embryonal tumour outcome based on gene expression. Nature 415:436–442.

Pusztai L, Ayers M, Stec J, Hortobagyi GN. 2003. Clinical application of cDNA microarrays in oncology. Oncologist 8:252–258.

Ramaswamy S, Tamayo P, Rifkin R, Mukherjee S, Yeang CH, Angelo M, Ladd C, Reich M, Latulippe E, Mesirov JP, Poggio T, Gerald W, Loda M, Lander ES, Golub TR. 2001. Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci USA 98: 15149–15154.

Reid LH, Crider-Miller SJ, West A, Lee MH, Massague J, Weissman BE. 1996a. Genomic organization of the human p57KIP2 gene and its analysis in the G401 Wilms' tumor assay. Cancer Res 56:1214–1218.

Ricketts SL. 2006. Bioinformatics: Computer-based approaches to genetic analysis. In: Coleman WB, Tsongalis GJ, editors. Molecular diagnostics: For the clinical laboratorian. Totowa: Humana Press Inc. p 57–61.

Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltnane JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, Lopez-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt LM. 2002. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 346:1937–1947.

Rowley JD. 1973. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. Nature 243:290–293.

Russo G, Zegar C, Giordano A. 2003. Advantages and limitations of microarray technology in human cancer. Oncogene 22:6497–6507.

Saramaki OR, Porkka KP, Vessella RL, Visakorpi T. 2006. Genetic aberrations in prostate cancer by microarray analysis. Int J Cancer 119:1322–1329.

Schwaenen C, Nessling M, Wessendorf S, Salvi T, Wrobel G, Radlwimmer B, Kestler HA, Haslinger C, Stilgenbauer S, Dohner H, Bentz M, Lichter P. 2004. Automated array-based genomic profiling in chronic lymphocytic leukemia: Development of a clinical tool and discovery of recurrent genomic alterations. Proc Natl Acad Sci USA 101:1039–1044.

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. 2004. Large-scale copy number polymorphism in the human genome. Science 305:525–528.

Segal E, Friedman N, Kaminski N, Regev A, Koller D. 2005. From signatures to models: Understanding cancer using microarrays. Nat Genet 37: (Suppl):S38–S45.

Selzer RR, Richmond TA, Pofahl NJ, Green RD, Eis PS, Nair P, Brothman AR, Stallings RL. 2005. Analysis of chromosome breakpoints in neuroblastoma at sub-kilobase resolution using fine-tiling oligonucleotide array CGH. Genes Chromosomes Cancer 44:305–319.

Shih Ie M, Wang TL. 2005. Apply innovative technologies to explore cancer genome. Curr Opin Oncol 17:33–38.

Shipp MA, Ross KN, Tamayo P, Weng AP, Kutok JL, Aguiar RC, Gaasenbeek M, Angelo M, Reich M, Pinkus GS, Ray TS, Koval MA, Last KW, Norton A, Lister TA, Mesirov J, Neuberg DS, Lander ES, Aster JC, Golub TR. 2002. Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning. Nat Med 8:68–74.

Slamon DJ, Godolphin W, Jones LA, Holt JA, Wong SG, Keith DE, Levin WJ, Stuart SG, Udove J, Ullrich A, et al. 1989. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. Science 244:707–712.

Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lonning P, Borresen-Dale AL. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA 98:10869–10874.

Southern EM. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J Mol Biol 98:503–517.

Sumner AT, Evans HJ, Buckland RA. 1971. New technique for distinguishing between human chromosomes. Nat New Biol 232:31–32.

Sunday ME. 1995. Differential display RT-PCR for identifying novel gene expression in the lung. Am J Physiol 269:L273–L284.

Takahashi M, Rhodes DR, Furge KA, Kanayama H, Kagawa S, Haab BB, Teh BT. 2001. Gene expression profiling of clear cell renal cell carcinoma: Gene identification and prognostic classification. Proc Natl Acad Sci USA 98:9754–9759.

Tashiro F, Morimura S, Hayashi K, Makino R, Kawamura H, Horikoshi N, Nemoto K, Ohtsubo K, Sugimura T, Ueno Y. 1986. Expression of the c-Ha-ras and c-myc genes in aflatoxin B1-induced hepatocellular carcinomas. Biochem Biophys Res Commun 138:858–864.

Taya Y, Hosogai K, Hirohashi S, Shimosato Y, Tsuchiya R, Tsuchida N, Fushimi M, Sekiya T, Nishimura S. 1984. A novel combination of K-ras and myc amplification accompanied by point mutational activation of K-ras in a human lung cancer. EMBO J 3:2943–2946.

Teuffel O, Betts DR, Dettling M, Schaub R, Schafer BW, Niggli FK. 2004. Prenatal origin of separate evolution of leukemia in identical twins. Leukemia 18:1624–1629.

Torregrosa D, Bolufer P, Lluch A, Lopez JA, Barragan E, Ruiz A, Guillem V, Munarriz B, Garcia Conde J. 1997. Prognostic significance of c-erbB-2/neu amplification and epidermal growth factor receptor (EGFR) in primary

breast cancer and their relation to estradiol receptor (ER) status. Clin Chim Acta 262:99–119.

Travis LB, Rabkin CS, Brown LM, Allan JM, Alter BP, Ambrosone CB, Begg CB, Caporaso N, Chanock S, DeMichele A, Figg WD, Gospodarowicz MK, Hall EJ, Hisada M, Inskip P, Kleinerman R, Little JB, Malkin D, Ng AK, Offit K, Pui CH, Robison LL, Rothman N, Shields PG, Strong L, Taniguchi T, Tucker MA, Greene MH. 2006. Cancer survivorship—genetic susceptibility and second primary cancers: Research strategies and recommendations. J Natl Cancer Inst 98:15–25.

van Beers EH, Nederlof PM. 2006. Array-CGH and breast cancer. Breast Cancer Res 8:210.

van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. 2002. A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med 347:1999–2009.

van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415:530–536.

Veltman JA, Fridlyand J, Pejavar S, Olshen AB, Korkola JE, DeVries S, Carroll P, Kuo WL, Pinkel D, Albertson D, Cordon-Cardo C, Jain AN, Waldman FM. 2003. Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. Cancer Res 63:2872–2880.

Vollenweider-Zerargui L, Barrelet L, Wong Y, Lemarchand-Beraud T, Gomez F. 1986. The predictive value of estrogen and progesterone receptors' concentrations on the clinical behavior of breast cancer in women. Clinical correlation on 547 patients. Cancer 57:1171–1180.

Wang J, Meza-Zepeda LA, Kresse SH, Myklebost O. 2004. M-CGH: Analysing microarray-based CGH experiments. BMC Bioinformatics 5:74.

Weigelt B, Hu Z, He X, Livasy C, Carey LA, Ewend MG, Glas AM, Perou CM, Van't Veer LJ. 2005. Molecular portraits and 70-gene prognosis signature are preserved throughout the metastatic process of breast cancer. Cancer Res 65:9155–9158.

Weiss MM, Kuipers EJ, Postma C, Snijders AM, Siccama I, Pinkel D, Westerga J, Meuwissen SG, Albertson DG, Meijer GA. 2003a. Genomic profiling of gastric cancer predicts lymph node status and survival. Oncogene 22:1872–1879.

Yeoh EJ, Ross ME, Shurtleff SA, Williams WK, Patel D, Mahfouz R, Behm FG, Raimondi SC, Relling MV, Patel A, Cheng C, Campana D, Wilkins D, Zhou X, Li J, Liu H, Pui CH, Evans WE, Naeve C, Wong L, Downing JR. 2002. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. Cancer Cell 1:133–143.

Yim SC, Park HG, Chang HN, Cho DY. 2005. Array-based mutation detection of BRCA1 using direct probe/target hybridization. Anal Biochem 337:332–337.

Zhao H, Ljungberg B, Grankvist K, Rasmuson T, Tibshirani R, Brooks JD. 2006. Gene expression profiling predicts survival in conventional renal cell carcinoma. PLoS Med 3:e13.

Zou TT, Selaru FM, Xu Y, Shustova V, Yin J, Mori Y, Shibata D, Sato F, Wang S, Olaru A, Deacu E, Liu TC, Abraham JM, Meltzer SJ. 2002. Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. Oncogene 21:4855–4862.