Carcinogenesis vol.21 no.3 pp.443–452, 2000

Genome changes and gene expression in human solid tumors

Joe W.Gray[1] and Colin Collins

UCSF Cancer Center, 2340 Sutter Street, University of California San Francisco, San Francisco, CA 94143-0808, USA

[1]To whom correspondence should be addressed
Email: jgray@cc.ucsf.edu

Genome-wide analysis techniques such as chromosome painting, comparative genomic hybridization, representational difference analysis, restriction landmark genome scanning and high-throughput analysis of LOH are now accelerating high-resolution genome aberration localization in human tumors. These techniques are complemented by procedures for detection of differentially expressed genes such as differential display, nucleic acid subtraction, serial analysis of gene expression and expression microarray analysis. These efforts are enabled by work from the human genome program in physical map development, cDNA library production/sequencing and in genome sequencing. This review covers several commonly used large-scale genome and gene expression analysis techniques, outlines genomic approaches to gene discovery and summarizes information that has come from large-scale analyses of human solid tumors.

Introduction

Tumors progress through the continuous accumulation of genetic and epigenetic changes that enable escape from normal cellular and environmental controls. These aberrations may involve genes that affect cell cycle control, apoptosis, angiogenesis, adhesion, transmembrane signaling, DNA repair and genomic stability. A substantial number of such oncogenes and tumor suppressor genes has already been discovered. However, large-scale genome analysis techniques suggest that the number of such genes may be large, perhaps strikingly so, and many important cancer-related genes remain to be discovered. Indeed, it is common to find as much as 30% of a solid tumor genome to be present at abnormal copy number, or otherwise aberrant (1), and the number of genomic mutations may be as large as 10^3–10^5 per tumor (2–4). Of course, many of these are likely to reflect 'noise' accumulated as genomically unstable tumors progress. The challenge is to distinguish the important genomic aberrations from the genomic noise, identify the affected genes, and clarify their roles in tumorigenesis, progression and/or response to therapy.

Definition of regions of recurrent genomic aberration is one historically important route to the identification of genes that play a role in cancer. A number of such regions have been identified in human cancers but the functional consequences of most of these abnormalities are not yet known. Identification of the affected genes in these regions, elucidation of their functions and association of these genes with tumor progression are required to fully understand tumorigenesis and progression. Regions of recurrent abnormality are typically defined through comprehensive analysis of many tumors with the goal of finding a few tumors carrying informative aberrations that can be used to narrowly define the extent of the region. Large-scale genomic analysis, such as chromosome painting, comparative genomic hybridization (CGH), representational difference analysis (RDA), restriction landmark genome scanning (RLGS) and high-throughput analysis of loss of heterozygosity (LOH) are now accelerating genome aberration localization.

Identification of genes that are expressed differently in normal tissues and the cancers that originate in these tissues is another important approach to identification of cancer-related genes. Differential display, nucleic acid subtraction and serial analysis of gene expression (SAGE) have been applied effectively to discover such genes. More recently, expression microarray analysis techniques have been developed that promise to allow quantitative, very large-scale analysis of gene expression.

These efforts have been enabled by work from the human genome program in physical map development, cDNA library production/sequencing and in genome sequencing. For example, cDNA and genomic clones serve as targets for microarray analysis technologies; interpretation of CGH, RDA and RLGS data depends on physical map information, and gene discovery has been speeded considerably by genome sequencing and cDNA library characterization.

This review covers several commonly used large-scale genome and gene expression analysis techniques, outlines genomic approaches to gene discovery and summarizes information that has come from large-scale analyses of human solid tumors.

Genome analysis techniques

Remarkable progress has been made in recent years in the development of technologies for definition of regions likely to harbor genes important in tumorigenesis or progression. These include: (i) metaphase chromosome analysis; (ii) genome copy number mapping; (iii) high-throughput polymorphism analysis; and (iv) multiplex PCR analysis.

Metaphase chromosome analysis

Analysis of metaphase chromosomes has been a cornerstone of cancer genetics since the introduction of modern banding analysis techniques. Banding analysis has been especially useful in identifying causative chromosome rearrangements in leukemias and lymphomas (5). However, this approach has been less successful in solid tumors because of the difficulty of obtaining high quality, representative metaphase chromosome preparations and because the high level of chromosomal rearrangement complicates karytoype interpretation. Fluorescence *in situ* hybridization (FISH) with chromosome specific

Abbreviations: CGH, comparative genomic hybridization; FISH, fluorescence *in situ* hybridization; LOH, loss of heterozygosity; RDA, representational difference analysis; RLGS, restriction landmark genome scanning; SAGE, serial analysis of gene expression.

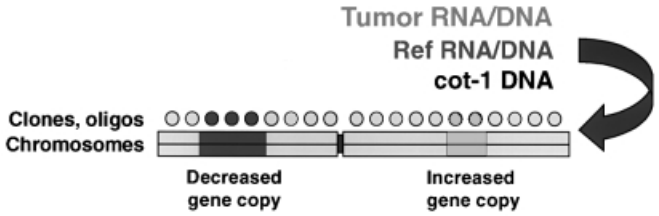

Fig. 1. Fluorescence ratio hybridization is a key component of comparative genomic hybridization and expression microarray analysis. In these processes, two nucleic acid samples to be compared are differentially labeled with reagents that fluoresce at different wavelengths. They are then hybridized along with excess, unlabeled repeat rich DNA, to the representation of the genome onto which information is to be mapped. In CGH, the representation may be either metaphase chromosomes or arrays of cloned probes. In expression microarray analysis, the representation may be arrays of cDNA clones or oligonucleotides.

probes (6,7) has significantly improved chromosome classification by increasing the specificity with which chromosomes or subregions thereof can be recognized. Staining patterns that can be achieved range from whole chromosome staining to precise staining of specific loci. Recently, combinatorial labeling strategies (e.g. probe 1 stained red, probe 2 stained green, probe 3 stained red + green, etc.) have been introduced that allow dozens of spatially separate loci or chromosomes to be simultaneously visualized (8–11). These techniques have dramatically increased the accuracy and sensitivity with which chromosome aberrations can be detected and classified, and clearly demonstrate the limitations of conventional banding analysis in analysis of human malignancies (12). Several recent reviews cover developments in FISH-based chromosome analysis (9,13–17).

Genome copy number mapping

Although FISH has substantially improved metaphase chromosome classification, its application in solid tumors is still limited by the difficulty of interpreting the complex karyotypes. CGH, developed in 1992, partially overcomes this by mapping changes in relative DNA sequence copy number onto normal metaphase chromosomes (18). In CGH, total genome DNAs from tumor and reference samples are labeled independently with different fluorochromes or haptens and co-hybridized to normal chromosome preparations along with excess unlabeled Cot-1 DNA to inhibit hybridization of labeled repeated sequences. The concept of CGH is illustrated schematically in Figure 1 and the results of a CGH hybridization are shown in Figure 2. The ratio of the amounts of the two genomes that hybridize to each location on the target chromosomes is an indication of the relative copy number of the two DNA samples at that point in the genome. Figure 3 illustrates the application of CGH to analysis of an advanced breast cancer. The remarkable level of genomic abnormality is apparent. The principle advantages of CGH are that it maps changes in copy number throughout a complex genome onto a normal reference genome so the aberrations can be easily related to existing physical maps, genes and genomic DNA sequence, and it employs genomic DNA so that cell culture is not required. Technical aspects of CGH are covered in several reviews (19–21).

The main limitations of chromosome-based CGH are that it is limited in resolution to 10–20 Mb, it does not provide quantitative information about gene dosage and it is insensitive to structural aberrations that do not result in a DNA sequence copy number change. Replacing metaphase chromosomes as the substrate onto which aberrations are mapped with arrays of well-mapped cloned nucleic acid sequences can eliminate some of these limitations, as illustrated in Figure 1. The arrays are constructed using a robot to place clone DNA in high-density arrays on glass substrates. Array densities as high as 10^4/cm can now be achieved. This approach has now been demonstrated in several laboratories (22–24). Initial work involved CGH to arrays comprised of targets spanning >100 kb of genomic sequence such as BACs (22,23). Most recently, CGH to cDNA arrays has been demonstrated (24). cDNA arrays are attractive for CGH since they are increasingly available and carry a very large number of clones. In addition, the same array can be used to assess expression and copy number. However, the sensitivity of cDNA clone-based CGH for detection of low-level copy number changes is likely to be less than that for BAC-based CGH due to the decreased hybridization signal to the smaller clones. At present, both approaches appear to be useful and clearly demonstrate that changes in genome copy number can be detected and mapped at a resolution defined by the genomic spacing of the clones used to form the array. Furthermore, array CGH allows quantitative assessment of DNA sequence dosage from one copy per test genome to hundreds of copies per genome (23). Figure 4 from Pinkel *et al.* (23) shows an analysis of the amplicon structure on Chr 20q for a human breast cancer. Both the increased resolution of array CGH compared with chromosome CGH and the opportunities for quantitative aberration definition are apparent in this analysis.

Quantitative analysis of genome copy number also can be accomplished using real-time, quantitative PCR (25–27). In this procedure, PCR is carried out in a 96-well format using a PCR reaction containing 'TaqMan' reporter oligonucleotides (carrying a reporter fluorescence molecule and a quencher molecule) as indicators of DNA sequence copy number. Exonuclease activity during PCR digests the Taqman probes resulting in liberation of the fluorescent reporter from the quencher. As a result, reporter fluorescence can be measured to detect the extent of the PCR reaction. This technique may be an alternative to analysis of LOH in cases where LOH is the result of a physical deletion, since polymorphism of the alleles being measured is not necessary. The copy number at each locus is measured relative to others in the same 96-well plate so that all sites measured are informative. The major drawbacks of this procedure are the labor and cost associated with synthesis of the TaqMan probes and the relatively small number (hundreds) of loci that can be conveniently analyzed.

RLGS is another genome scanning technique that allows identification of genome copy number differences (28,29). In addition, it enables identification of mutations and polymorphisms. In RLGS, a test genome is digested with a rare cutting enzyme like *Not*I and the resulting fragments are radioactively labeled. The DNA is then digested with a second restriction enzyme and electrophoretically separated on an agarose gel. Finally, the DNA fragments in the gel are digested with a third enzyme and electrophoretically separated in a second dimension by placing a strip from the agarose gel along the top of an acrylamide gel. The radioactively labeled landmark DNA fragments are then detected using autoradiography or by phosphoimaging. Up to 2000 separate loci can be analyzed in a single experiment. Applications of RLGS include high-speed construction of linkage maps (30,31), quantitative analysis of copy number at each landmark locus (32) and detection of mutations involving the restriction sites used to prepare the two-dimensional (2D) electropherograms. Landmark fragments

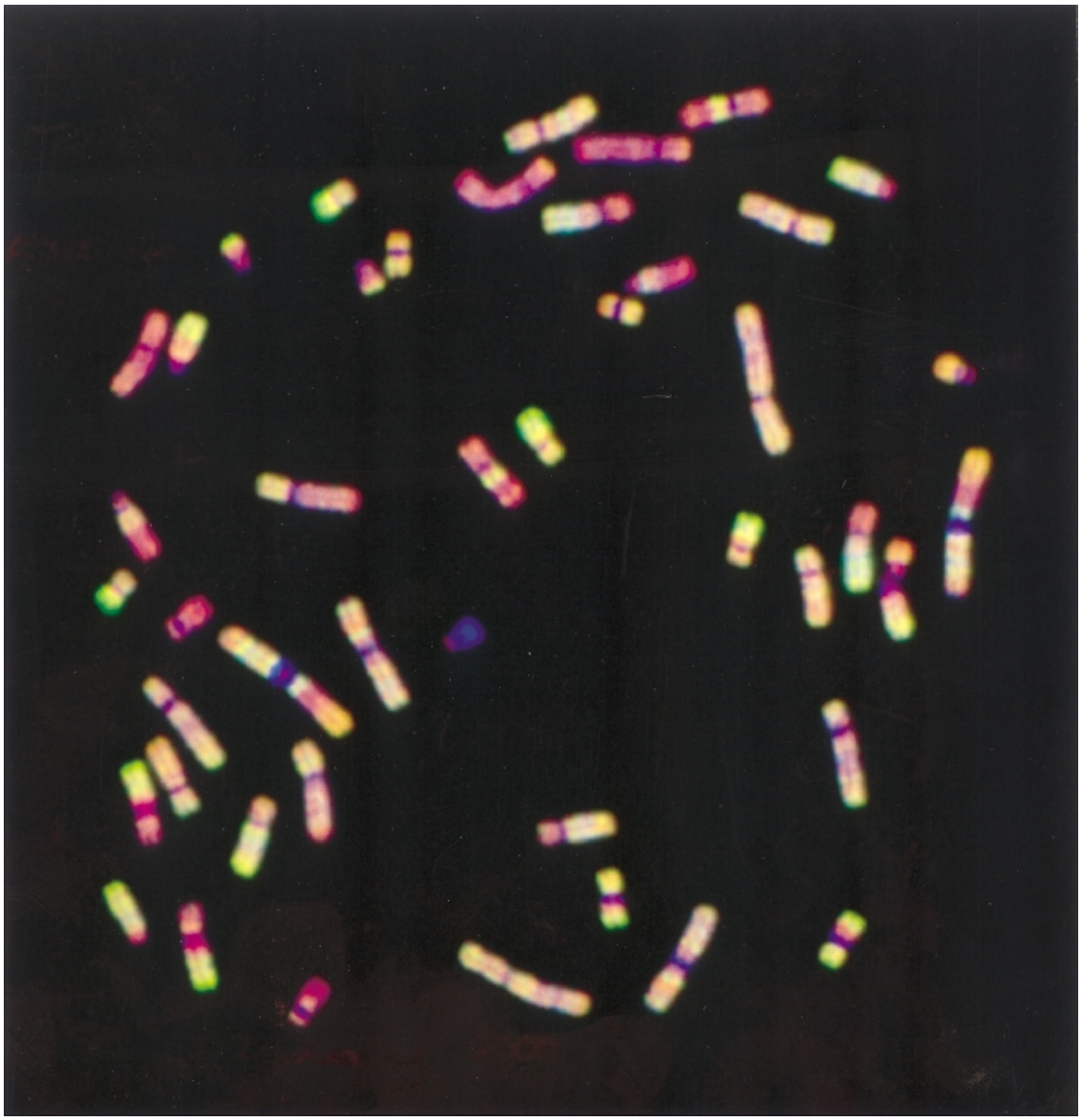

Fig. 2. Fluorescence photomicrograph showing the results of a CGH analysis of the human breast cancer cell line, MCF7. MCF7 DNA was labeled green and normal reference DNA was labeled red. The chromosomes were counterstained with DAPI. Thus, regions of weak hybridization appear blue, regions of increased copy number appear green and regions of decreased copy number appear red.

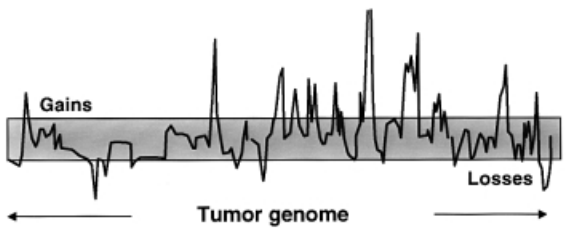

Fig. 3. A CGH analysis of an advanced breast cancer. The data are arranged along the *x*-axis with chromosome 20pter to the left and chromosome 22qter to the right. The green:red CGH ratio is plotted along the *y*-axis. The gray band indicates the region of normal variability. Thus, values above the band show significant increases in copy number and values below the band show significant decreases in copy number.

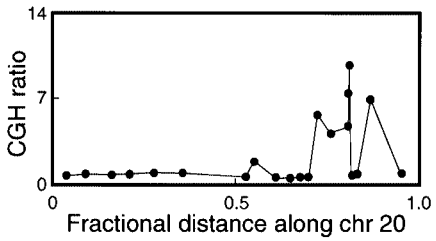

Fig. 4. An array CGH analysis of genome copy number changes along chromosome 20 in a human breast cancer (adapted from ref. 23). The resolution and dynamic range is much improved compared with chromosome CGH.

of interest may be cloned by excising the DNA from that part of the 2D gel (33). Alternately, cloned *Not*I/*Eco*RV boundary library elements can be mixed with the genomic DNA during RLGS so that landmark loci present in the library appear to be increased in intensity after RLGS (34). The corresponding clones can then be obtained from the library for further analysis. RLGS can be made sensitive to methylation differences by

employing methylation-sensitive restriction enzymes (29,34). The principal limitations of RLGS are that many landmark loci are not associated with existing physical maps and it does not permit high-resolution analysis of interesting regions.

Arbitrarily primed PCR (AP-PCR) also has been used to detect somatic genetic alterations in human tumors (35). In this approach, DNA fingerprints generated by PCR amplification of normal and tumor tissue using single arbitrary primers are separated electrophoretically and compared. Differences in PCR fragment band intensity (visible as AP-PCR bands with decreased or increased intensities in tumor tissue DNA relative to normal) indicate regions of the tumor genome that are present in altered copy number. Sequences located at these sites can be cloned after reamplification with the same arbitrary primer. This approach also has been applied to detection of gene expression differences between tumor and normal tissue (36) (see below).

High-throughput analysis of polymorphisms

The adaptation of automated DNA sequencers for the analysis of DNA fragment length polymorphisms has facilitated linkage studies and analysis of LOH by allowing rapid, inexpensive measurements of a large number of loci in multiple samples (37). Automated analysis is accomplished by PCR amplifying loci containing length polymorphisms such as those produced by simple sequence repeat length variations. To allow automated analysis, one member of each primer pair is synthesized to contain a fluorescent label (38). Amplification results in allele-specific products that can be distinguished during electrophoresis. Use of separable fluorescent labels allows analysis of several different loci in each lane in the gel. Careful selection of PCR primers to amplify regions of different size allows simultaneous analysis of even more loci in each lane.

Single nucleotide polymorphisms (SNPs) also can be detected efficiently by hybridization of fluorescently labeled, PCR-amplified representations of the genome to arrays comprised of oligonucleotides (39,40). Both alleles of each of several thousand SNP markers and single base mismatch targets may be represented on an array. The stringency of the hybridization reaction is adjusted so that hybridization is diminished if a single base mismatch exists between the probe and oligonucleotide substrate. Thus, the presence or absence of an allele in the hybridization mixture can be determined by its hybridization signature. This technique is highly parallel and scales well to genome wide assessments of linkage or LOH. However, its robustness for analysis of formalin-fixed, paraffin-embedded samples from pathology archives remains to be assessed.

Representational difference analysis (RDA)

Techniques like array CGH or analysis of LOH are limited at present since they only detect aberrations at the test loci. Nucleic acid subtraction strategies overcome this limitation (41–43). These techniques allow detection and cloning of segments of DNA that differ in copy number between two complex genomes. In this approach, the two genomes to be compared are enzymatically restricted and ligated to linker oligonucleotides. This process creates 'representations' of the two genomes to be compared that can be amplified using primers for the adapter oligonucleotides. Differences between two genomes are detected by denaturing representations from both genomes and hybridizing the 'tester' genome against excess amounts of the 'driver' genome. The representation elements common to each will form driver–tester heteroduplexes, whereas elements that are present only in 'tester' DNA form tester–tester homoduplexes. Special linkers attached to the representations before hybridization allow PCR amplification of homoduplexes but suppress the amplification of heteroduplexes. Thus, the products of the amplification are strongly enriched in sequences present only in the tester DNA. These strategies reveal differences that result from deletion, amplification or mutation of the test genome. However, since they survey representations of the genome, several representations must be tested for comprehensive difference discovery.

Genome sequence analysis

Once regions of abnormality or susceptibility have been defined to within ~1 Mb, it becomes feasible to assess genes and associated regulatory sequences in these regions. This process typically begins with analysis of candidate genes already mapped to these regions. This is increasingly productive as the number of mapped ESTs in public databases increases. The work of the CGAP (http://www.ncbi.nlm.nih.gov/CGAP/), TIGR (http://www.tigr.org/) and the IMAGE programs (http://www-bio.llnl.gov/bbrp/image/image.html) are especially important in this regard. However, there is risk with the candidate gene approach of missing crucial genes. Moreover, the candidate gene approach reveals little about the regulatory elements of the candidate genes. Thus, comprehensive techniques for gene discovery and characterization are needed. Large-scale genomic sequencing is now sufficiently advanced that it is reasonable to consider genomic sequencing as a practical tool for gene discovery. The worldwide effort is scheduled to complete a 'draft' sequence of the human genome by mid-2000 and to have the complete 'Bermuda quality' sequence by 2003 (44). This will dramatically speed discovery of candidate genes in regions of recurrent abnormality or linkage to disease susceptibility.

Human, human sequence by itself is not adequate for comprehensive biological annotation since computational tools are not yet sufficiently robust to accurately identify exons, predict their boundaries and to group them into functional genes. In addition, regulatory regions may be missed and the false positive rate for transcription factor binding site predictions remains too high. Comparative genomic sequencing of syntenic regions in other organisms promises to substantially enhance assessment of gene function and discovery of coding and regulatory sequences (45). The power of this approach has already been demonstrated in comparisons of *Caenorhabditis elegans* and *Caenorhabditis briggsae* genomic sequences (46,47) and through comparisons of human genomic sequences with those from model organisms such as *C.elegans* (48,49), *Drosophila* (50), *Saccharomyces cerevisiae* (51) and rodents (52). Comparison between human and mouse appears especially useful for genomic sequence interpretation because of the increasing importance of mouse models in the elucidation of human cancer gene function. The finding of conserved regulatory sequences in introns in mid-gene (53,54) is intriguing since it raises the possibility that comparative genomic sequencing may reveal previously unsuspected regulatory regions.

Gene expression

Remarkable progress has been made in recent years in the development of techniques to identify differences in gene expression between cell populations. At least three approaches

are now in widespread use: (i) differential display; (ii) nucleic acid subtraction; (iii) analysis using expression microarrays.

Differential display

This well-established technique is used to identify and isolate genes that are differentially expressed between two cell populations (55). In this approach, mRNA sequences from cell populations to be compared are reverse transcribed and amplified by PCR using a set of oligonucleotide primers, one anchored to the poly(A) tail and the other to a short arbitrary oligonucleotide that binds at varying distances from the poly(A) tail for the various RNA molecules. For some RNA molecules, the separation between the two primer sequences is too large to allow PCR amplification so that only a subset of RNA molecules are amplified. Separation of the amplified sequences on a DNA sequencing gel allows visualization of each of the amplified sequences. Comparison of gels for two cell populations reveals sequences that are abundant in one but not the other. Use of several different primer sets allows analysis of a larger number of genes. Sequences of interest may be excised from the gel and cloned. The advantages of differential display include its ease of use and its power to discover previously unknown differences. Its principal disadvantages are that not all differences are discovered using a single arbitrary primer, recovery of interesting DNA fragments is somewhat time consuming and differences in levels of expression are difficult to quantify. Nonetheless, this technique has been widely and successfully applied to analysis of human malignancies.

Nucleic acid subtraction

Techniques to clone differences between two mRNA populations are well developed. The principles are similar to those for RDA. The process begins with reverse transcription of the mRNA from two populations to form cDNA. In one approach, the 'driver' cDNA is labeled to allow affinity separation of the labeled driver sequences. The driver cDNA is then hybridized in excess to 'tester' cDNA from the other population and the driver–driver and tester–driver hybrid molecules are removed by affinity separation (56). Alternately, the driver cDNA and hybrid molecules are enzymatically removed by digestion with exonucleases rather than by physical partitioning (57).

Serial analysis of gene expression (SAGE)

The relative frequency of gene expression can also be determined by sequencing a large number of cDNA fragments in a library prepared from the cells or tissue of interest (58,59). This is accomplished by ligating together short ~10 bp long sequence 'tags' from the 3′-most *Nla*III restriction sites of multiple genes. The tags are separated by distinctive linker sequences so the various sequences can be distinguished. The ligated sequences from many different concatimers are then sequenced and the results compiled to form a distribution showing the frequencies of the various gene-associated tags. This process is sufficiently efficient that 10^4–10^5 tags can be sequenced from each library. The main advantage of SAGE is its unbiased assessment of the frequencies with which genes are expressed. Disadvantages include the lack of clones from novel tags that may appear during sequencing and the need for extensive sequencing to accurately assess levels of expression of weakly expressed genes. NCI resources for SAGE analysis as well as a more detailed description can be found at http://www.ncbi.nlm.nih.gov/SAGE/.

Expression microarrays

Enormous progress has been made in recent years in the development and DNA sequence characterization of cDNA clones from the human, mouse and other model organisms. In humans, these data have been computationally assembled into over 8000 genes and 83 000 clusters. The cDNA clones associated with these sequences are publicly available. These clones and their associated sequences form the basis for a powerful microarray approach to large-scale analysis of gene expression (60–64). In this approach, labeled mRNA samples are hybridized to arrays of cDNA clones or oligonucleotides derived from the associated sequences. The arrays may be on silicon or membrane substrates. The labeled probes may be labeled radioactively or with fluorescent reagents so that the resulting hybridization signals can be detected using autoradiography, phosphorimaging or fluorescence imaging. cDNA and oligonucleotides arrays have been made using robots to move DNA from microtiter trays to silicon substrates or to nylon membranes (62,63,65). This approach is flexible and is especially well suited to production of custom arrays but also been applied to make large-scale arrays carrying as many as 40 000 different clones. An alternative is to synthesize oligonucleotide arrays directly on silicon substrates using photolithographic approaches (61,66). These techniques work by projecting light through a photolithographic mask onto the synthesis substrate. The light 'deprotects' the surface so a nucleoside carrying a photolabile protecting group can be added. The synthesis proceeds by deprotecting all the areas that are to receive a common nucleoside, coupling that nucleoside, deprotecting areas to receive another nucleoside and so on. Thus, four cycles and four photolithographic masks are required to add one base to the entire array. The photolithographic approaches allow production of large numbers of very high-density arrays. However, the initial setup costs are high so this approach is best suited to production of large numbers of 'standard' arrays. One alternative is to use a scanned light spot to deprotect each 'feature' independently (67). However, this approach has not yet been fully developed. Single oligonucleotide arrays on silicon substrates have been constructed with elements representing more than 40 000 genes/ESTs while densities on membranes are somewhat lower (60). Several review articles on microarray technologies and their applications can be found in 'The Chipping Forecast' (68).

Large-scale analyses of solid tumors

Genome scanning techniques such as chromosome painting, spectral karyotyping, analysis of LOH and CGH clearly demonstrate a remarkably high degree of chromosome rearrangement in human solid tumors (Figures 3 and 4). Analyses of LOH and CGH and interphase FISH are especially useful since they can be applied to uncultured primary tumors and thus give a relatively unbiased view of the spectrum of abnormalities. In addition, they map abnormalities onto existing physical maps of the genome. A comprehensive review of the literature is impossible in this publication because of the remarkably large number of tumor analyses that have been published in recent years thanks to these more efficient analysis technologies. However, several general features of solid tumors are becoming apparent. Several of these are reviewed below. The associated references are necessarily anecdotal and only intended as an entry point to the literature.

Recurrent genome aberrations

Cytogenetic and genome analysis techniques have revealed numerous regions that are frequently abnormal in tumors of the same type. The number of such regions in any given tumor type can be large, presumably the result of malfunction of components of damage surveillance systems, DNA repair and/or mitotic apparatus that lead to chromosomal and genetic instability. As an example, over 30 regions of abnormal copy number or LOH have been identified in breast cancers (69,70). In addition, many of these abnormalities appear abnormal in multiple tumor types. Several well established oncogenes, tumor suppressor genes and genes associated with cancer susceptibility have been mapped to regions of recurrent abnormality. These include *MYC* (amplification at 8q24), *AKT2* (amplification at 19q13), *ERBB2* (amplification at 17q21.2), *CCND1* (amplification at 11q13), *p53* (LOH at 17p13), *BRCA1* (LOH at 17q21) and *BRCA2* (LOH at 13q12). In addition, many well mapped regions of genetic susceptibility in mouse tumors localize to regions of LOH/genome copy number abnormality (71). These data strongly support the notion that regions of recurrent abnormality encode genes that contribute to cancer progression when differentially expressed because of mutation, loss or amplification.

Several cancer-related genes have been identified based on their locations in regions of recurrent LOH including the multiple endocrine neoplasia type 1 gene (*MEN-1*) at 11q13 (72), the fragile histidine triad (*FHIT*) gene at 3p14.2 (73), and a cell adhesion associated gene [*CDH1* at 16q22 (74)]. Genes discovered in regions of abnormality identified using CGH include the androgen receptor (*AR*) gene, located in a region of increased copy number at Xq12 in hormone-refractory prostate cancers (75–77), and *PIK3CA*, located in a region of recurrent abnormality at 3q26 in ovarian cancer (78). In addition, several genes have been identified in regions of amplification at 20q including the steroid receptor co-activator, *AIB1* (79,80); a putative Zn-finger transcription factor, *ZNF217* (81), associated with instability and immortalization (82), and a centrosome-associated serine/threonine protein kinase *STK15* (83). The protein tyrosine phosphatase, *PTEN* was discovered using RDA and has been implicated as a tumor suppressor gene. Comprehensive reviews of regions of common cytogenetic and genomic abnormalities by Mitelman *et al.* (84) and Knuutila *et al.* (85,86), respectively, illustrate the vast number of regions that remain to be explored. Discovery and functional characterization of the genes in these regions and determination of the order, if any, in which they occur, should lead to a much improved understanding of carcinogenesis and progression.

Progression

Association of specific abnormalities with progression is important since this may lead to identification of early events that may guide the development of markers for early detection, shed light on the mechanisms associated with carcinogenesis, provide information about the mechanisms of progression and lead to development of improved prognostic or predictive markers. Several studies have investigated progression related events. This process is best worked out in colon cancer where lesions at various stages of progression are readily apparent (87). Unfortunately, such visual clues are not as apparent in many other tumor types, so events associated with progression are much less well established. Nonetheless, genome scanning techniques have revealed several general aspects of progression in other tumor systems. As expected, the overall number of abnormalities increases with indicators of progression such as grade and/or stage (88–90). Somewhat surprising, however, is the finding that genomic aberrations occur early during progression and are often well established prior to the onset of invasion (91). Finally, comparison of invasive cancers with distant metastases reveals remarkable genomic similarities between these two disease stages in the majority of cancers (90,92–94). In fact, related tumors are generally sufficiently similar that genome scanning techniques may well be able to distinguish *de novo* from recurrent tumors (95). Numerous regions of specific abnormalities also have been associated with progression. However, the details are sufficiently tumor-type-specific that their description is beyond the scope of this review.

Although many recurrent abnormalities have been identified, the exact spectrum of aberrations often varies according to tumor histology, genetic or ethnic background. Invasive ductal and lobular breast carcinomas, for example, show distinct patterns of recurrent abnormalities with 16q loss more frequent, and 8q and 20q gains less frequent in lobular cancers than in ductal cancers (96). Likewise, in brain tumors, LOH involving D17S379 (17p13.3) is associated with high-grade malignancies such as anaplastic astrocytoma and glioblastoma multiforme (97). Genetic background also influences the spectrum of accumulated abnormalities. The most extreme example involves patients with an inherited predisposition to microsatellite instability (98). Tumors from these patients show few chromosomal or LOH abnormalities while those from patients lacking this mutator phenotype show a much larger number of chromosomal abnormalities; consistent with the idea that these two classes of tumor evolve through different genetic mechanisms (99). Genetic-background-specific differences in aberration spectrum also have been observed between sporadic breast cancers and those arising in patients with *BRCA1* and *BRCA2* (100). Both the number and the spectrum of abnormalities differ between these groups of patients with tumors arising in patients carrying predisposing genetic lesions having more accumulated abnormalities. Given this influence of genetic background on the spectrum of abnormalities, it would not be surprising to find that the spectrum of abnormalities differs according to ethnic background. CGH studies (101) suggest such differences. Specifically, gain of the 4q25–q28 region appears to be much more common in prostate tumors from African-American patients compared with those from Caucasian patients. Taken together, these data support the hypothesis that recurrent abnormalities encode genes that play important roles in cancer progression. They also indicate that tumors of different histology and genetic background may be significantly different at the genetic level and thus likely to differ in their biological characteristics and clinical behavior including response to therapy.

Inter-tumor heterogeneity, prognostication and prediction of response to therapy

Substantial genomic and expression differences also can exist between tumors that appear clinically similar. This is likely one reason why tumors that appear similar can progress and respond to therapy in dramatically different ways. The existence of these differences and the availability of convenient technologies for their detection has stimulated an international effort to identify molecular determinants of tumor behavior with the goal of improving the precision with which tumors can be classified according to outcome (102–104). Numerous gene

specific studies support the concept of molecular tumor classification. Particularly strong associations with response to chemotherapy have been established between genes in the *p53*, signaling and apoptosis pathways (105–108). Large-scale genome and gene expression profiling studies suggest the existence of a much larger number of therapy-associated genes. Several reports have associated clinical outcome with genome copy number increases and/or decreases with clinical outcome (109–115), LOH (73,116–121), general level of genomic disarray [e.g. the total number of abnormalities as measured using genome scanning techniques such as CGH, LOH or RLGS (1,122,123)] and differential gene expression (124–126). These studies suggest the possibility of using individual tumor genetic profiles to predict responses to specific therapies if applied before the start of treatment or to detection of conditions of resistance if applied during the course of treatment.

Differential gene expression

Differential display, SAGE and subtraction techniques have been applied to solid tumors and tumor cell lines, although the total number of samples analyzed is far fewer than have been analyzed using genome analysis technologies. The smaller number is mostly due to the difficulty of obtaining tumor material prepared in a manner that preserves mRNA. Nonetheless, these studies have revealed numerous differentially expressed genes, typically dozens to hundreds per analysis. This is not surprising, given the large number of genomic aberrations that exist in most solid tumors. SAGE has been applied to gene expression differences between normal human bronchial/tracheal epithelial cell cultures and non-small cell lung cancers (127) and to assess and study differences in gene expression between cells with functional and non-functional *p53* (128). Differential display has been more extensively applied. Recent studies include analyses gene expression changes associated with quiescence or in late G_1 phase of the cell cycle in human breast cancer cells (129), to identify genes regulated by androgen in an androgen-responsive prostate cancer cell line (130), and to identify genes that are differentially expressed in prostate cancers (131). Differential display also has been used to compare gene expression in normal and tumor-derived human mammary epithelial cells (132) to identify genes associated with progression in breast cancer (133). Expression microarrays have not yet been applied extensively to human tumors. However, analyses of cell lines and model organisms illustrate its potential. Recent applications include analysis of radiation stress response (134,135), analysis of response of human fibroblasts to serum (136), detection of differences in gene expression between normal and neoplastic human ovarian tissues (137) and analysis of gene expression in alveolar rhabdomyosarcoma cell lines (138). These studies show the potential of expression microarrays both for detection of differences in gene expression and in quantitative analysis of gene expression levels. However, they also reveal the dramatic changes in gene expression that can result from sample handling and cell culture as well as the difficulty of interpreting gene expression changes in hundreds or thousands of genes. Success in this venture will depend critically on the development of new information management and interpretation techniques (139–141).

References

1. Iwabuchi,H., Sakamoto,M., Sakunaga,H., Ma,Y.Y., Carcangiu,M.L., Pinkel,D., Yang-Feng,T.L. and Gray,J.W. (1995) Genetic analysis of benign, low-grade and high-grade ovarian tumors. *Cancer Res.*, **55**, 6172–6180.
2. Loeb,L.A. and Christians,F.C. (1996) Multiple mutations in human cancers. *Mutat. Res.*, **350**, 279–286.
3. Cheng,K.C. and Loeb,L.A. (1997) Genomic stability and instability: a working paradigm. *Curr. Top. Microbiol. Immunol.*, **221**, 5–18.
4. Ionov,Y., Peinado,M.A., Malkhosyan,S., Shibata,D. and Perucho,M. (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. *Nature*, **363**, 558–561.
5. Rowley,J.D. (1998) The critical role of chromosome translocations in human leukemias. *Annu. Rev. Genet.*, **32**, 495–519.
6. Pinkel,D., Landegent,J., Collins,C., Fuscoe,J., Segraves,R., Lucas,J. and Gray,J. (1988) Fluorescence *in situ* hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. *Proc. Natl Acad. Sci. USA*, **85**, 9138–9142.
7. Lichter,P., Cremer,T., Borden,J., Manuelidis,L. and Ward,D.C. (1988) Delineation of individual human chromosomes in metaphase and interphase cells by *in situ* suppression hybridization using recombinant DNA libraries. *Hum. Genet.*, **80**, 224–234.
8. Speicher,M.R., Gwyn Ballard,S. and Ward,D.C. (1996) Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genet.*, **12**, 368–375.
9. Speicher,M.R. and Ward,D.C. (1996) The coloring of cytogenetics. *Nature Med.*, **2**, 1046–1048.
10. Liyanage,M., Coleman,A., du Manoir,S., Veldman,T., McCormack,S., Dickson,R.B., Barlow,C., Wynshaw-Boris,A., Janz,S., Wienberg,J., Ferguson-Smith,M.A., Schrock,E. and Ried,T. (1996) Multicolour spectral karyotyping of mouse chromosomes. *Nature Genet.*, **14**, 312–315.
11. Schrock,E., du Manoir,S., Veldman,T., Schoell,B., Wienberg,J., Ferguson-Smith,M.A., Ning,Y., Ledbetter,D.H., Bar-Am,I., Soenksen,D., Garini,Y. and Ried,T. (1996) Multicolor spectral karyotyping of human chromosomes. *Science*, **273**, 494–497.
12. Fleischman,E.W., Reshmi,S., Sokova,O.I., Kirichenko,O.P., Konstantinova, L.N., Kulagina,O.E., Frenkel,M.A. and Rowley,J.D. (1999) Increased karyotype precision using fluorescence *in situ* hybridization and spectral karyotyping in patients with myeloid malignancies. *Cancer Genet. Cytogenet.*, **108**, 166–170.
13. Fletcher,C.D. (1997) Soft tissue tumours: the impact of cytogenetics and molecular genetics. *Verh. Dtsch Ges. Pathol.*, **81**, 318–326.
14. Lichter,P. (1997) Multicolor FISHing: what's the catch? *Trends Genet.*, **13**, 475–479.
15. Ried,T., Liyanage,M., du Manoir,S., Heselmeyer,K., Auer,G., Macville,M. and Schrock,E. (1997) Tumor cytogenetics revisited: comparative genomic hybridization and spectral karyotyping. *J. Mol. Med.*, **75**, 801–814.
16. Ried,T., Schrock,E., Ning,Y. and Wienberg,J. (1998) Chromosome painting: a useful art. *Hum. Mol. Genet.*, **7**, 1619–1626.
17. Raap,A.K. (1998) Advances in fluorescence *in situ* hybridization. *Mutat. Res.*, **400**, 287–298.
18. Kallioniemi,A., Kallioniemi,O.P., Sudar,D., Rutovitz,D., Gray,J.W., Waldman,F. and Pinkel,D. (1992) Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, **258**, 818–821.
19. Kallioniemi,O.P., Kallioniemi,A., Sudar,D., Rutovitz,D., Gray,J.W., Waldman,F. and Pinkel,D. (1993) Comparative genomic hybridization: a rapid new method for detecting and mapping DNA amplification in tumors. *Semin. Cancer Biol.*, **4**, 41–46.
20. James,L.A. (1999) Comparative genomic hybridization as a tool in tumour cytogenetics. *J. Pathol.*, **187**, 385–395.
21. DeVries,S., Gray,J.W., Pinkel,D., Waldman,F.M. and Sudar,D. (1995) *Comparative Genomic Hybridization, Current Protocols in Human Genetics.* John Wiley, New York, Vol. 6, pp. 1–18.
22. Solinas-Toldo,S., Lampel,S., Stilgenbauer,S., Nickolenko,J., Benner,A., Dohner,H., Cremer,T. and Lichter,P. (1997) Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*, **20**, 399–407.
23. Pinkel,D., Segraves,R., Sudar,D., Clark,S., Poole,I., Kowbel,D., Collins,C., Kuo,W.L., Chen,C., Zhai,Y., Dairkee,S.H., Ljung,B.M., Gray,J.W. and Albertson,D.G. (1998) High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genet.*, **20**, 207–211.
24. Pollack,J.R., Perou,C.M., Alizadeh,A.A., Eisen,M.B., Pergamenschikov,A., Williams,C.F., Jeffrey,S.S., Botstein,D. and Brown,P.O. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.*, **23**, 41–46.
25. Heid,C.A., Stevens,J., Livak,K.J. and Williams,P.M. (1996) Real time quantitative PCR. *Genome Res.*, **6**, 986–994.
26. Srivastava,M., McBride,O.W., Fleming,P.J., Pollard,H.B. and Burns,A.L.

(1990) Genomic organization and chromosomal localization of the human nucleolin gene. *J. Biol. Chem.*, **265**, 14922–14931.
27. Bassler,H., Flood,S., Livak,K., Marmaro,J., Knorr,R. and Batt,C. (1995) Use of a fluorogenic probe in a PCR-based assay for the detection of *Listeria monocytogenes*. *Appl. Environ. Microbiol.*, **61**, 3724–3728.
28. Okazaki,Y. and Hayashizaki Y. (1997) High-speed positional cloning based on restriction landmark genome scanning. *Methods*, **13**, 359–377.
29. Nakamura,M., Konishi,N., Tsunoda,S., Hiasa,Y., Takemura,K., Tsuzuki,T., Kobitsu,K. and Sakaki,T. (1997) Genomic alterations of human gliomas detected by restriction landmark genomic scanning. *Brain Tumor Pathol.*, **14**, 13–17.
30. Okuizumi,H., Ohsumi,T., Sakaki,N., Imoto,H., Mizuno,Y., Hanami,T., Yamashita,H., Kamiya,M., Takada,S., Kitamura,A., Muramatsu,M., Nishimura,M., Mori,M., Matsuda,Y., Tagaya,O., Okazaki,Y. and Hayashizaki,Y. (1997) Linkage map of Syrian hamster with restriction landmark genomic scanning. *Mamm. Genome*, **8**, 121–128.
31. Hayashizaki,Y., Hirotsune,S., Okazaki,Y., Shibata,H., Akasako,A., Muramatsu,M., Kawai,J., Hirasawa,T., Watanabe,S., Shiroishi,T. *et al.* (1994) A genetic linkage map of the mouse using restriction landmark genomic scanning (RLGS). *Genetics*, **138**, 1207–1238.
32. Yamamoto,K., Konishi,N., Inui,T., Kitahori,Y., Hiasa,Y., Kirita,T. and Sugimura,M. (1998) Genomic alterations in oral squamous cell carcinoma cell lines detected by two-dimensional gel analysis. *Oral Oncol.*, **34**, 508–512.
33. Hirotsune,S., Shibata,H., Okazaki,Y., Sugino,H., Imoto,H., Sasaki,N., Hirose,K., Okuizumi,H., Muramatsu,M., Plass,C. *et al.* (1993) Molecular cloning of polymorphic markers on RLGS gel using the spot target cloning method. *Biochem. Biophys. Res. Commun.*, **194**, 1406–1412.
34. Smiraglia,D.J., Fruhwald,M.C., Costello,J.F., McCormick,S.P., Dai,Z., Peltomaki,P., O'Dorisio,M.S., Cavenee,W.K. and Plass,C. (1999) A new tool for the rapid cloning of amplified and hypermethylated human DNA sequences from restriction landmark genome scanning gels. *Genomics*, **58**, 254–262.
35. Peinado,M.A., Malkhosyan,S., Velazquez,A. and Perucho,M. (1992) Isolation and characterization of allelic losses and gains in colorectal tumors by arbitrarily primed polymerase chain reaction. *Proc. Natl Acad. Sci. USA*, **89**, 10065–10069.
36. Perucho,M., Welsh,J., Peinado,M.A., Ionov,Y. and McClelland,M. (1995) Fingerprinting of DNA and RNA by arbitrarily primed polymerase chain reaction: applications in cancer research. *Methods Enzymol.*, **254**, 275–290.
37. Canzian,F., Salovaara,R., Hemminki,A., Kristo,P., Chadwick,R.B., Aaltonen,L.A. and de la Chapelle,A. (1996) Semiautomated assessment of loss of heterozygosity and replication error in tumors. *Cancer Res.*, **56**, 3331–3337.
38. Levitt,R.C., Kiser,M.B., Dragwa,C., Jedlicka,A.E., Xu,J., Meyers,D.A. and Hudson,J.R. (1994) Fluorescence-based resource for semiautomated genomic analyses using microsatellite markers. *Genomics*, **24**, 361–365.
39. Sapolsky,R.J., Hsie,L., Berno,A., Ghandour,G., Mittmann,M. and Fan,J.B. (1999) High-throughput polymorphism screening and genotyping with high-density oligonucleotide arrays. *Genet. Anal.*, **14**, 187–192.
40. Gentalen,E. and Chee,M. (1999) A novel method for determining linkage between DNA sequences: hybridization to paired probe arrays. *Nucleic Acids Res.*, **27**, 1485–1491.
41. Lisitsyn,N. and Wigler,M. (1993) Cloning the differences between two complex genomes. *Science*, **259**, 946–951.
42. Diatchenko,L., Lau,Y.F., Campbell,A.P., Chenchik,A., Moqadam,F., Huang,B., Lukyanov,S., Lukyanov,K., Gurskaya,N., Sverdlov,E.D. and Siebert,P.D. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl Acad. Sci. USA*, **93**, 6025–6030.
43. Diatchenko,L., Lukyanov,S., Lau,Y.F. and Siebert,P.D. (1999) Suppression subtractive hybridization: a versatile method for identifying differentially expressed genes. *Methods Enzymol.*, **303**, 349–380.
44. Collins,F.S., Patrinos,A., Jordan,E., Chakravarti,A., Gesteland,R. and Walters,L. (1998) New goals for the U.S. Human Genome Project: 1998–2003. *Science*, **282**, 682–689.
45. Clark,M.S. (1999) Comparative genomics: the key to understanding the Human Genome Project. *Bioessays*, **21**, 121–130.
46. Gilleard,J.S., Henderson,D.K. and Ulla,N. (1997) Conservation of the *Caenorhabditis elegans* cuticle collagen gene col-12 in *Caenorhabditis briggsae. Gene*, **193**, 181–186.
47. Nelson,D.W., Linning,R.M., Davison,P.J. and Honda,B.M. (1998) 5′-Flanking sequences required for efficient transcription *in vitro* of 5S RNA genes, in the related nematodes *Caenorhabditis elegans* and *Caenorhabditis briggsae. Gene*, **218**, 9–16.
48. Plasterk,R.H. (1999) The year of the worm. *Bioessays*, **21**, 105–109.
49. Wilson,R.K. (1999) How the worm was won. The *C.elegans* genome sequencing project. *Trends Genet.*, **15**, 51–58.
50. Rubin,G.M. (1998) The *Drosophila* genome project: a progress report. *Trends Genet.*, **14**, 340–343.
51. Goffeau,A., Aert,R., Agostini-Carbone,M.L., Ahmed,A., Aigle,M. *et al.* (1997) The yeast genome directory. *Nature*, **387** (suppl.), 1–105.
52. Mushegian,A.R., Bassett,D.E.Jr, Boguski,M.S., Bork,P. and Koonin,E.V. (1997) Positionally cloned human disease genes: patterns of evolutionary conservation and functional motifs. *Proc. Natl Acad. Sci. USA*, **94**, 5831–5836.
53. Rohrer,J. and Conley,M.E. (1998) Transcriptional regulatory elements within the first intron of Bruton's tyrosine kinase. *Blood*, **91**, 214–221.
54. Seib,T., Welter,C., Engel,M., Theisinger,B. and Dooley,S. (1994) Presence of regulatory sequences within intron 4 of human and murine c-*myb* genes. *Biochim. Biophys. Acta*, **1219**, 285–292.
55. Martin,K.J. and Pardee,A.B. (1999) Principles of differential display. *Methods Enzymol.*, **303**, 234–258.
56. Lopez-Fernandez,L.A. and del Mazo,J. (1993) Construction of subtractive cDNA libraries from limited amounts of mRNA and multiple cycles of subtraction. *Biotechniques*, **15**, 654–659.
57. Zeng,J., Gorski,R.A. and Hamer,D. (1994) Differential cDNA cloning by enzymatic degrading subtraction (EDS). *Nucleic Acids Res.*, **22**, 4381–4385.
58. Velculescu,V.E., Zhang,L., Vogelstein,B. and Kinzler,K.W. (1995) Serial analysis of gene expression. *Science*, **270**, 484–487.
59. Carulli,J.P., Artinger,M., Swain,P.M., Root,C.D., Chee,L., Tulig,C., Guerin,J., Osborne,M., Stein,G., Lian,J. and Lomedico,P.T. (1998) High throughput analysis of differential gene expression. *J. Cell Biochem. Suppl.*, **31**, 286–296.
60. Lipshutz,R.J., Fodor,S.P., Gingeras,T.R. and Lockhart,D.J. (1999) High density synthetic oligonucleotide arrays. *Nature Genet.*, **21**, 20–24.
61. Fodor,S.P., Read,J.L., Pirrung,M.C., Stryer,L., Lu,A.T. and Solas,D. (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science*, **251**, 767–773.
62. DeRisi,J., Penland,L., Brown,P.O., Bittner,M.L., Meltzer,P.S., Ray,M., Chen,Y., Su,Y.A. and Trent,J.M. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nature Genet.*, **14**, 457–460.
63. Shalon,D., Smith,S.J. and Brown,P.O. (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.*, **6**, 639–645.
64. Duggan,D.J., Bittner,M., Chen,Y., Meltzer,P. and Trent,J.M. (1999) Expression profiling using cDNA microarrays. *Nature Genet.*, **21**, 10–14.
65. Alon,U., Barkai,N., Notterman,D.A., Gish,K., Ybarra,S., Mack,D. and Levine,A.J. (1999) Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. *Proc. Natl Acad. Sci. USA*, **96**, 6745–6750.
66. Maskos,U. and Southern,E.M. (1993) A study of oligonucleotide reassociation using large arrays of oligonucleotides synthesised on a glass support. *Nucleic Acids Res.*, **21**, 4663–4669.
67. Singh-Gasson,S., Green,R.D., Yue,Y., Nelson,C., Blattner,F., Sussman,M.R. and Cerrina,F. (1999) Maskless fabrication of light-directed oligonucleotide microarrays using a digital micromirror array. *Nature Biotechnol.*, **17**, 974–978.
68. (1998) The Chipping Forecast. *Nature Genet.*, **Suppl. 21**, pp. 1–60.
69. Kallioniemi,A., Kallioniemi,O.P., Piper,J., Tanner,M., Stokke,T., Chen,L., Smith,H.S., Pinkel,D., Gray,J.W. and Waldman,F.M. (1994) Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. Natl Acad. Sci. USA*, **91**, 2156–2160.
70. Callahan,R., Cropp,C., Merlo,G.R., Diella,F., Venesio,T., Lidereau,R., Cappa,A.P. and Lisicia,D.S. (1993) Genetic and molecular heterogeneity of breast cancer cells. *Clin. Chim. Acta*, **217**, 63–73.
71. Balmain,A. and Nagase,H. (1998) Cancer resistance genes in mice: models for the study of tumour modifiers. *Trends Genet.*, **14**, 139–144.
72. The European Consortium on MEN-1 (1998) The search for the MEN1 gene. The European Consortium on MEN-1. *J. Intern Med.*, **243**, 441–446.
73. Burke,L., Khan,M.A., Freedman,A.N., Gemma,A., Rusin,M., Guinee,D.G., Bennett,W.P., Caporaso,N.E., Fleming,M.V., Travis,W.D., Colby,T.V., Trastek,V., Pairolero,P.C., Tazelaar,H.D., Midthun,D.E., Liotta,L.A. and Harris,C.C. (1998) Allelic deletion analysis of the FHIT gene predicts poor survival in non-small cell lung cancer. *Cancer Res.*, **58**, 2533–2536. [Published erratum appears in *Cancer Res.*, (1998), **58**, 3488.]
74. Berx,G., Becker,K.F., Hofler,H. and van Roy,F. (1998) Mutations of the human E-cadherin (*CDH1*) gene. *Hum. Mutat.*, **12**, 226–237.
75. Koivisto,P., Kononen,J., Palmberg,C., Tammela,T., Hyytinen,E., Isola,J., Trapman,J., Cleutjens,K., Noordzij,A., Visakorpi,T. and Kallioniemi,O.P.

(1997) Androgen receptor gene amplification: a possible molecular mechanism for androgen deprivation therapy failure in prostate cancer. *Cancer Res.*, **57**, 314–319.
76. Koivisto,P., Hyytinen,E., Palmberg,C., Tammela,T., Visakorpi,T., Isola,J. and Kallioniemi,O.P. (1995) Analysis of genetic changes underlying local recurrence of prostate carcinoma during androgen deprivation therapy. *Am. J. Pathol.*, **147**, 1608–1614.
77. Koivisto,P., Kolmer,M., Visakorpi,T. and Kallioniemi,O.P. (1998) Androgen receptor gene and hormonal therapy failure of prostate cancer. *Am. J. Pathol.*, **152**, 1–9.
78. Shayesteh,L., Yiling,L., Kuo,W.-L., Collins,C., Yang-Feng,T., Pinkel,D., Mills,G.B. and Gray,J.W. (1999) *PIK3CA* is implicated as an oncogene in ovarian cancer. *Nature Genet.*, **21**, 99–102.
79. Anzick,S.L., Kononen,J., Walker,R.L., Azorsa,D.O., Tanner,M.M., Guan,X.Y., Sauter,G., Kallioniemi,O.P., Trent,J.M. and Meltzer,P.S. (1997) AIB1, a steroid receptor coactivator amplified in breast and ovarian cancer. *Science*, **277**, 965–968.
80. Ghadimi,B.M., Schrock,E., Walker,R.L., Wangsa,D., Jauho,A., Meltzer,P.S. and Ried,T. (1999) Specific chromosomal aberrations and amplification of the AIB1 nuclear receptor coactivator gene in pancreatic carcinomas. *Am. J. Pathol.*, **154**, 525–536.
81. Collins,C., Rommens,J.M., Kowbel,D., Godfrey,T., Tanner,M., Hwang,S.I., Polikoff,D., Nonet,G., Cochran,J., Myambo,K., Jay,K.E., Froula,J., Cloutier,T., Kuo,W.L., Yaswen,P., Dairkee,S., Giovanola,J., Hutchinson,G.B., Isola,J., Kallioniemi,O.P., Palazzolo,M., Martin,C., Ericsson,C., Pinkel,D., Gray,J.W. *et al.* (1998) Positional cloning of *ZNF217* and *NABC1*: genes amplified at 20q13.2 and overexpressed in breast carcinoma. *Proc. Natl Acad. Sci. USA*, **95**, 8703–8708.
82. Savelieva,E., Belair,C.D., Newton,M.A., DeVries,S., Gray,J.W., Waldman,F. and Reznikoff,C.A. (1997) 20q gain associates with immortalization: 20q13.2 amplification correlates with genome instability in human papillomavirus 16 E7 transformed human uroepithelial cells. *Oncogene*, **14**, 551–560.
83. Zhou,H., Kuang,J., Zhong,L., Kuo,W.L., Gray,J.W., Sahin,A., Brinkley,B.R. and Sen,S. (1998) Tumour amplified kinase STK15/BTAK induces centrosome amplification, aneuploidy and transformation. *Nature Genet.*, **20**, 189–193.
84. Mitelman,F., Mertens,F. and Johansson,B. (1997) A breakpoint map of recurrent chromosomal rearrangements in human neoplasia. *Nature Genet.*, **15**, 417–474.
85. Knuutila,S., Aalto,Y., Autio,K., Bjorkqvist,A.M., El-Rifai,W., Hemmer,S., Huhta,T., Kettunen,E., Kiuru-Kuhlefelt,S., Larramendy,M.L., Lushnikova,T., Monni,O., Pere,H., Tapper,J., Tarkkanen,M., Varis,A., Wasenius,V.M., Wolf,M. and Zhu,Y. (1999) DNA copy number losses in human neoplasms. *Am. J. Pathol.*, **155**, 683–694.
86. Knuutila,S., Bjorkqvist,A.M., Autio,K., Tarkkanen,M., Wolf,M., Monni,O., Szymanska,J., Larramendy,M.L., Tapper,J., Pere,H., El-Rifai,W., Hemmer,S., Wasenius,V.M., Vidgren,V. and Zhu,Y. (1998) DNA copy number amplifications in human neoplasms: review of comparative genomic hybridization studies. *Am. J. Pathol.*, **152**, 1107–1123.
87. Kinzler,K.W. and Vogelstein,B. (1997) Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature*, **386**, 761–763.
88. Isola,J.J., Kallioniemi,O.P., Chu,L.W., Fuqua,S.A., Hilsenbeck,S.G., Osborne,C.K. and Waldman,F.M. (1995) Genetic aberrations detected by comparative genomic hybridization predict outcome in node-negative breast cancer. *Am. J. Pathol.*, **147**, 905–911.
89. Richter,J., Beffa,L., Wagner,U., Schraml,P., Gasser,T.C., Moch,H., Mihatsch,M.J. and Sauter,G. (1998) Patterns of chromosomal imbalances in advanced urinary bladder cancer detected by comparative genomic hybridization. *Am. J. Pathol.*, **153**, 1615–1621.
90. Nishizaki,T., DeVries,S., Chew,K., Goodson,W.H.III, Ljung,B.M., Thor,A. and Waldman,F.M. (1997) Genetic alterations in primary breast cancers and their metastases: direct comparison using modified comparative genomic hybridization. *Genes Chromosomes Cancer*, **19**, 267–272.
91. O'Connell,P., Pekkel,V., Fuqua,S.A., Osborne,C.K., Clark,G.M. and Allred,D.C. (1998) Analysis of loss of heterozygosity in 399 premalignant breast lesions at 15 genetic loci. *J. Natl Cancer Inst.*, **90**, 697–703.
92. Tarkkanen,M., Huuhtanen,R., Virolainen,M., Wiklund,T., Asko-Seljavaara,S., Tukiainen,E., Lepantalo,M., Elomaa,I. and Knuutila,S. (1999) Comparison of genetic changes in primary sarcomas and their pulmonary metastases. *Genes Chromosomes Cancer*, **25**, 323–331.
93. Bissig,H., Richter,J., Desper,R., Meier,V., Schraml,P., Schaffer,A.A., Sauter,G., Mihatsch,M.J. and Moch,H. (1999) Evaluation of the clonal relationship between primary and metastatic renal cell carcinoma by comparative genomic hybridization. *Am. J. Pathol.*, **155**, 267–274.
94. Al-Mulla,F., Keith,W.N., Pickford,I.R., Going,J.J. and Birnie,G.D. (1999) Comparative genomic hybridization analysis of primary colorectal carcinomas and their synchronous metastases. *Genes Chromosomes Cancer*, **24**, 306–314.
95. Emmert-Buck,M.R., Chuaqui,R., Zhuang,Z., Nogales,F., Liotta,L.A. and Merino,M.J. (1997) Molecular analysis of synchronous uterine and ovarian endometrioid tumors. *Int. J. Gynecol. Pathol.*, **16**, 143–148.
96. Nishizaki,T., Chew,K., Chu,L., Isola,J., Kallioniemi,A., Weidner,N. and Waldman,F.M. (1997) Genetic alterations in lobular breast cancer by comparative genomic hybridization. *Int. J. Cancer*, **74**, 513–517.
97. Chattopadhyay,P., Rathore,A., Mathur,M., Sarkar,C., Mahapatra,A.K. and Sinha,S. (1997) Loss of heterozygosity of a locus on 17p13.3, independent of p53, is associated with higher grades of astrocytic tumours. *Oncogene*, **15**, 871–874.
98. Willenbucher,R.F., Aust,D.E., Chang,C.G., Zelman,S.J., Ferrell,L.D., Moore,D.H.II and Waldman,F.M. (1999) Genomic instability is an early event during the progression pathway of ulcerative-colitis-related neoplasia. *Am. J. Pathol.*, **154**, 1825–1830.
99. Larramendy,M.L., el-Rifai,W., Kokkola,A., Puolakkainen,P., Monni,O., Salovaara,R., Aarnio,M. and Knuutila,S. (1998) Comparative genomic hybridization reveals differences in DNA copy number changes between sporadic gastric carcinomas and gastric carcinomas from patients with hereditary nonpolyposis colorectal cancer. *Cancer Genet. Cytogenet.*, **106**, 62–65.
100. Tirkkonen,M., Johannsson,O., Agnarsson,B.A., Olsson,H., Ingvarsson,S., Karhu,R., Tanner,M., Isola,J., Barkardottir,R.B., Borg,A. and Kallioniemi,O.P. (1997) Distinct somatic genetic changes associated with tumor progression in carriers of *BRCA1* and *BRCA2* germ-line mutations. *Cancer Res.*, **57**, 1222–1227.
101. Cher,M.L., Bova,G.S., Moore,D.H., Small,E.J., Carroll,P.R., Pin,S.S., Epstein,J.I., Isaacs,W.B. and Jensen,R.H. (1996) Genetic alterations in untreated metastases and androgen-independent prostate cancer detected by comparative genomic hybridization and allelotyping. *Cancer Res.*, **56**, 3091–3102.
102. Klausner,R.D. (1999) Cancer, genomics and the National Cancer Institute. *J. Clin. Invest.*, **104**, S15–S17.
103. Klausner,R.D. (1999) Studying cancer in the mouse. *Oncogene*, **18**, 5249–5252.
104. Strausberg,R.L., Dahl,C.A. and Klausner,R.D. (1997) New opportunities for uncovering the molecular basis of cancer. *Nature Genet.*, **15**, 415–416.
105. Bunz,F., Hwang,P.M., Torrance,C., Waldman,T., Zhang,Y., Dillehay,L., Williams,J., Lengauer,C., Kinzler,K.W. and Vogelstein,B. (1999) Disruption of p53 in human cancer cells alters the responses to therapeutic agents. *J. Clin. Invest.*, **104**, 263–269.
106. Serrone,L. and Hersey,P. (1999) The chemoresistance of human malignant melanoma: an update. *Melanoma Res.*, **9**, 51–58.
107. Thor,A.D., Berry,D.A., Budman,D.R., Muss,H.B., Kute,T., Henderson,I.C., Barcos,M., Cirrincione,C., Edgerton,S., Allred,C., Norton,L. and Liu,E.T. (1998) erbB-2, p53 and efficacy of adjuvant therapy in lymph node-positive breast cancer. *J. Natl Cancer Inst.*, **90**, 1346–1360.
108. Elledge,R.M. and Allred,D.C. (1998) Prognostic and predictive value of p53 and p21 in breast cancer. *Breast Cancer Res. Treat.*, **52**, 79–98.
109. Tarkkanen,M., Elomaa,I., Blomqvist,C., Kivioja,A.H., Kellokumpu-Lehtinen,P., Bohling,T., Valle,J. and Knuutila,S. (1999) DNA sequence copy number increase at 8q: a potential new prognostic marker in high-grade osteosarcoma. *Int. J. Cancer*, **84**, 114–121.
110. Lastowska,M., Cotterill,S., Pearson,A.D., Roberts,P., McGuckin,A., Lewis,I. and Bown,N. (1997) Gain of chromosome arm 17q predicts unfavourable outcome in neuroblastoma patients. U.K. Children's Cancer Study Group and the U.K. Cancer Cytogenetics Group. *Eur. J. Cancer*, **33**, 1627–1633.
111. Weber,R.G., Sommer,C., Albert,F.K., Kiessling,M. and Cremer,T. (1996) Clinically distinct subgroups of glioblastoma multiforme studied by comparative genomic hybridization. *Lab. Invest.*, **74**, 108–119.
112. Hermsen,M.A., Baak,J.P., Meijer,G.A., Weiss,J.M., Walboomers,J.W., Snijders,P.J. and van Diest,P.J. (1998) Genetic analysis of 53 lymph node-negative breast carcinomas by CGH and relation to clinical, pathological, morphometric and DNA cytometric prognostic factors. *J. Pathol.*, **186**, 356–362.
113. Larramendy,M.L., Tarkkanen,M., Valle,J., Kivioja,A.H., Ervasti,H., Karaharju,E., Salmivalli,T., Elomaa,I. and Knuutila,S. (1997) Gains, losses and amplifications of DNA sequences evaluated by comparative genomic hybridization in chondrosarcomas. *Am. J. Pathol.*, **150**, 685–691.
114. Visakorpi,T., Kallioniemi,A.H., Syvanen,A.C., Hyytinen,E.R., Karhu,R., Tammela,T., Isola,J.J. and Kallioniemi,O.P. (1995) Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. *Cancer Res.*, **55**, 342–347.
115. Kudoh,K., Takano,M., Koshikawa,T., Hirai,M., Yoshida,S., Mano,Y.,

Yamamoto,K., Ishii,K., Kita,T., Kikuchi,Y., Nagata,I., Miwa,M. and Uchida,K. (1999) Gains of 1q21–q22 and 13q12–q14 are potential indicators for resistance to cisplatin-based chemotherapy in ovarian cancer patients. *Clin. Cancer Res.*, **5**, 2526–2531.
116. Takita,J., Hayashi,Y. and Yokota,J. (1997) Loss of heterozygosity in neuroblastomas–an overview. *Eur. J. Cancer*, **33**, 1971–1973.
117. Steiner,G. and Sidransky,D. (1996) Molecular differential diagnosis of renal carcinoma: from microscopes to microsatellites. *Am. J. Pathol.*, **149**, 1791–1795.
118. Brodeur,G.M. (1990) Neuroblastoma: clinical significance of genetic abnormalities. *Cancer Surv.*, **9**, 673–688.
119. Gasparotto,D., Vukosavljevic,T., Piccinin,S., Barzan,L., Sulfaro,S., Armellin,M., Boiocchi,M. and Maestro,R. (1999) Loss of heterozygosity at 10q in tumors of the upper respiratory tract is associated with poor prognosis. *Int. J. Cancer*, **84**, 432–436.
120. Ragnarsson,G., Eiriksdottir,G., Johannsdottir,J.T., Jonasson,J.G., Egilsson,V. and Ingvarsson,S. (1999) Loss of heterozygosity at chromosome 1p in different solid human tumours: association with survival. *Br. J. Cancer*, **79**, 1468–1474.
121. O'Connell,P., Fischbach,K., Hilsenbeck,S., Mohsin,S.K., Fuqua,S.A., Clark,G.M., Osborne,C.K. and Allred,D.C. (1999) Loss of heterozygosity at D14S62 and metastatic potential of breast cancer. *J. Natl Cancer Inst.*, **91**, 1391–1397.
122. Bocker,T., Schlegel,J., Kullmann,F., Stumm,G., Zirngibl,H., Epplen,J.T. and Ruschoff,J. (1996) Genomic instability in colorectal carcinomas: comparison of different evaluation methods and their biological significance. *J. Pathol.*, **179**, 15–19.
123. Mizoguchi,M., Inamura,T., Ikezaki,K., Iwaki,T., Oda,S., Maehara,Y., Oki,E., Terasaki,M. and Fukui,M. (1999) Patient survival and microsatellite instability in gliomas by high-resolution fluorescent analysis. *Oncol. Rep.*, **6**, 791–795.
124. Moch,H., Schraml,P., Bubendorf,L., Mirlacher,M., Kononen,J., Gasser,T., Mihatsch,M.J., Kallioniemi,O.P. and Sauter,G. (1999) High-throughput tissue microarray analysis to evaluate genes uncovered by cDNA microarray screening in renal cell carcinoma. *Am. J. Pathol.*, **154**, 981–986.
125. Sano,T., Lin,H., Chen,X., Langford,L.A., Koul,D., Bondy,M.L., Hess,K.R., Myers,J.N., Hong,Y.K., Yung,W.K. and Steck,P.A. (1999) Differential expression of MMAC/PTEN in glioblastoma multiforme: relationship to localization and prognosis. *Cancer Res.*, **59**, 1820–1824.
126. Hilsenbeck,S.G., Friedrichs,W.E., Schiff,R., O'Connell,P., Hansen,R.K., Osborne,C.K. and Fuqua,S.A. (1999) Statistical analysis of array expression data as applied to the problem of tamoxifen resistance. *J. Natl Cancer Inst.*, **91**, 453–459.
127. Hibi,K., Liu,Q., Beaudry,G.A., Madden,S.L., Westra,W.H., Wehage,S.L., Yang,S.C., Heitmiller,R.F., Bertelsen,A.H., Sidransky,D. and Jen,J. (1998) Serial analysis of gene expression in non-small cell lung cancer. *Cancer Res.*, **58**, 5690–5694.
128. Madden,S.L., Galella,E.A., Zhu,J., Bertelsen,A.H. and Beaudry,G.A. (1997) SAGE transcript profiles for p53-dependent growth regulation. *Oncogene*, **15**, 1079–1085.
129. Alpan,R.S., Sparvero,S. and Pardee,A.B. (1996) Identification of mRNAs differentially expressed in quiescence or in late G_1 phase of the cell cycle in human breast cancer cells by using the differential display method. *Mol. Med.*, **2**, 469–478.
130. Averboukh,L., Liang,P., Kantoff,P.W. and Pardee,A.B. (1996) Regulation of S100P expression by androgen. *Prostate*, **29**, 350–355.
131. Wang,F.L., Wang,Y., Wong,W.K., Liu,Y., Addivinola,F.J., Liang,P., Chen,L.B., Kantoff,P.W. and Pardee,A.B. (1996) Two differentially expressed genes in normal human prostate tissue and in carcinoma. *Cancer Res.*, **56**, 3634–3637.
132. Liang,P., Averboukh,L., Keyomarsi,K., Sager,R. and Pardee,A.B. (1992) Differential display and cloning of messenger RNAs from human breast cancer versus mammary epithelial cells. *Cancer Res.*, **52**, 6966–6968.
133. Burger,A.M., Zhang,X., Li,H., Ostrowski,J.L., Beatty,B., Venanzoni,M., Papas,T. and Seth,A. (1998) Down-regulation of T1A12/mac25, a novel insulin-like growth factor binding protein related gene, is associated with disease progression in breast carcinomas. *Oncogene*, **16**, 2459–2467.
134. Fornace,A.J.Jr, Amundson,S.A., Bittner,M., Myers,T.G., Meltzer,P., Weinsten,J.N. and Trent,J. (1999) The complexity of radiation stress responses: analysis by informatics and functional genomics approaches. *Gene Expr.*, **7**, 387–400.
135. Amundson,S.A., Bittner,M., Chen,Y., Trent,J., Meltzer,P. and Fornace,A.J.Jr (1999) Fluorescent cDNA microarray hybridization reveals complexity and heterogeneity of cellular genotoxic stress responses. *Oncogene*, **18**, 3666–3672.
136. Iyer,V.R., Eisen,M.B., Ross,D.T., Schuler,G., Moore,T., Lee,J.C.F., Trent,J.M., Staudt,L.M., Hudson,J.Jr, Boguski,M.S., Lashkari,D., Shalon,D., Botstein,D. and Brown,P.O. (1999) The transcriptional program in the response of human fibroblasts to serum. *Science*, **283**, 83–87.
137. Wang,K., Gan,L., Jeffery,E., Gayle,M., Gown,A.M., Skelly,M., Nelson,P.S., Ng,W.V., Schummer,M., Hood,L. and Mulligan,J. (1999) Monitoring gene expression profile changes in ovarian carcinomas using cDNA microarray. *Gene*, **229**, 101–108.
138. Khan,J., Simon,R., Bittner,M., Chen,Y., Leighton,S.B., Pohida,T., Smith,P.D., Jiang,Y., Gooden,G.C., Trent,J.M. and Meltzer,P.S. (1998) Gene expression profiling of alveolar rhabdomyosarcoma with cDNA microarrays. *Cancer Res.*, **58**, 5009–5013.
139. Claverie,J.M. (1999) Computational methods for the identification of differential and coordinated gene expression. *Hum. Mol. Genet.*, **8**, 1821–1832.
140. Zhang,M.Q. (1999) Promoter analysis of co-regulated genes in the yeast genome. *Comput. Chem.*, **23**, 233–250.
141. Eisen,M.B., Spellman,P.T., Brown,P.O. and Botstein,D. (1998) Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*, **95**, 14863–14868.

Received October 13, 1999; accepted October 29, 1999