For reprint orders, please contact reprints@future-drugs.com

Advancement in characterization of genomic alterations for improved diagnosis, treatment and prognostics in cancer

Xiaofeng Zhou, Tianwei Yu, Steven W Cole and David TW Wong[†]

Most human cancers are characterized by genetic instabilities. These instabilities manifest themselves as a series of genetic alterations, including discrete mutations and chromosomal aberrations. With the human genome deciphered, high-throughput technologies are rapidly advancing the field to generate genome-wide gene expression and mutation profiles that are highly correlative of biologic and disease phenotypes. While recent advancement in comprehensive genomic characterization presents an unprecedented opportunity for advancing the treatment of cancer, there are still many challenges that need to be overcome before we can fully utilize genomic markers and targets for cancer prediction, diagnostics, treatment and prognostics. This review describes recent advances in comprehensive genomic characterization at the DNA level, and considers some of the challenges that remain for defining the precise genomic portrait of tumors. Potential solutions that may help overcome these challenges are also offered.

Expert Rev. Mol. Diagn. 6(1), 39–50 (2006)

CONTENTS

Approaches to genomic profiling

Challenges for genomic studies of cancer

Comprehensive genomic approaches

Statistical considerations

Expert commentary & five-year view

Key issues

References

Affiliations

[†]*Author for correspondence*
Dental Research Institute, School of Dentistry & Division of Head & Neck Surgery/Otolaryngology, David Geffen School of Medicine & Henry Samueli School of Engineering & Jonsson Comprehensive Cancer Center, University of California at Los Angeles, Los Angeles, CA, USA
Tel.: +1 310 206 3048
Fax: +1 310 794 7109
dtww@ucla.edu

KEYWORDS:
comparative genomic hybridization, loss of heterozygosity, multicolor fluorescence *in situ* hybridization, personalized therapies, single nucleotide polymorphism array, spectral karyotyping, tiling array

Tumors develop through the combined processes of genetic instability and selection, resulting in clonal expansion of cells that have accumulated the most advantageous set of genetic aberrations. Many types of instability can contribute to neoplastic development, including point mutations, chromosomal rearrangements, DNA dosage abnormalities (amplifications or deletions), alteration of microsatellite sequences and epigenetic changes.

During the 1970s and 1980s, several genome-wide approaches were developed to measure these tumor genomic alterations including loss of heterozygosity (LOH) analysis and comparative genomic hybridization (CGH). Advances in genetics and bioengineering have refined these techniques over the past two decades, and the recent development of multicolor staining-based cytogenetic techniques such as multicolor (M)-fluorescence *in situ* hybridization (FISH) and spectral karyotyping (SKY) have further improved the ability to analyze the tumor genome [1].

Knowledge of genomic aberrations can have clinical value in diagnosis, treatment and prognostics in cancer. Four decades ago, the milestone discovery of Philadelphia chromosome (a translocation between chromosome 9 and 22 that fuses the Bcr gene and the Abl tyrosine kinase gene) led to one of the first effective targeted therapies for cancer: treatment of chronic myelogenous leukemia (CML) with the tyrosine kinase inhibitor imatinib (Gleevec®) [2]. Since then, many exciting clinical advances have been made, based on increasing knowledge of the tumor genome. The completion of the human genome project now makes it possible to query the cancer genome systematically in ways that were hitherto impossible [3,4]. Microarrays designed to analyze targeted genomic regions relevant to chronic lymphocytic leukemia have been produced for use in clinical trials to determine the relationship between therapeutic options and genomic aberrations [5]. The association of genomic aberrations with prognosis has been found for a variety of

10.1586/14737159.6.1.39
© 2006 Future Drugs Ltd
ISSN 1473-7159

tumor types, including prostate cancer [6], breast cancer [7], gastric cancer [8], head and neck cancer [9] and lymphoma [10,11]. Many more studies are in progress or near completion. These findings provide a new paradigm for cancer treatment that is fundamentally guided by genomic characterizations of the disease [12–16].

In this review, recent technological advances in the characterization of tumor genomes at the DNA level are surveyed, and some of the major challenges that remain in defining genomic markers and targets for cancer prediction, diagnostics, treatment and prognostics are considered. Potential solutions that may overcome these outstanding challenges to improve the diagnosis and treatment of solid tissue malignancies will be identified.

Approaches to genomic profiling

Chromosomal aberrations can be analyzed using a variety of high-throughput genetic and molecular technologies, including the analysis of chromosome banding, LOH, CGH, digital karyotyping [17], FISH, restriction landmark genome scanning (RLGS) [18], representational difference analysis (RDA) [19] and statistical inference of chromosomal changes from gene expression data [20–22]. These analyses enable the identification of a broad range of chromosomal abnormalities in cancer.

Comparative genomic hybridization: systematic copy-number analysis

CGH was developed to survey gene copy-number abnormalities (amplifications and deletions) across a whole genome [23]. In a typical CGH analysis, differentially labeled test (disease) and reference genomic DNAs are cohybridized to the normal metaphase chromosomes to generate fluorescence ratios along the length of chromosomes that provide a cytogenetic representation of DNA copy-number variation. CGH was the first effective approach to scan the entire genome for variations in DNA content [24,25]. However, chromosome-based CGH has a limited mapping resolution (~20 Mb). Array-based CGH is a second-generation approach in which fluorescence ratios on microarrayed DNA elements provide a locus-by-locus measure of gene copy-number variation [26,27]. Although this approach can potentially increase mapping resolution, most array CGH methods have utilized large genomic clones (e.g., bacterial artificial chromosomes), which limits spatial sensitivity. In addition, large genomic clones suffer from reduced specificity due to their inclusion of common repeats (e.g., Alu and long interspersed nuclear elements), redundant sequences (e.g., low copy repeats, also known as segmental duplications) and segments of extensive sequence similarity (pseudogenes or paralogous genes) [28]. Recently, several additional higher density tools for CGH analysis have become available with the completion of the human genome sequence. These include cDNA array-based CGH [29,30], oligonucleotide array-based CGH [31,32], tiling array-based CGH [27] and copy-number analysis using high-density single nucleotide polymorphism (SNP) microarrays [33–36]. Tiling and SNP array-based approaches have drawn most attention due to their high resolution. Tiling arrays have the potential to resolve small (gene level) gains and losses (at a resolution of ~40 kb) that might be missed by marker-based genomic arrays that contain a large number of gaps due to the distance between the targeted probes [27,37]. In the near future, it is hoped that the ability to survey copy-number changes at close to base-pair resolution, using tiling arrays that contain billions of overlapping probes covering the entire genome, will be realized. The SNP array-based approach provides the unique advantage of concurrent CGH and LOH analysis, which is discussed in further detail later [34,35].

An alternative approach to measure copy-number abnormalities is digital karyotyping, which was first introduced in 2002 [17]. This method is similar to the serial analysis of gene expression (SAGE) approach [38], where short sequence tags containing sufficient information of specific genomic loci are first obtained from the whole genome. These tags are then amplified *en masse*, concatenated, cloned, sequenced and computationally ordered in sequence along each chromosome. Digital enumeration of tag observations along each chromosome can then be used to quantitatively evaluate DNA content with high resolution. This method has been successfully utilized in several studies [39–41], thereby demonstrating the capacity to identify copy-number abnormalities at high resolution [42,43].

Loss of heterozygosity: systematic allelic imbalance analysis

Chromosomal aberrations include segments of allelic imbalance identifiable by LOH at polymorphic loci, which can be used to identify regions harboring tumor-suppressor genes. Allelic losses, which are caused by mitotic recombination, gene conversion or nondisjunction, cannot be detected by CGH and, thus, require LOH analysis for their identification. This approach is favored by the Knudson two-hit hypothesis for hunting the tumor-suppressor genes [44,45]. The discovery of the first tumor-suppressor gene, *RB1* [46], followed the Knudson two-hit hypothesis that tumor-suppressor genes are inactivated by a recessive mutation in one allele followed by the loss of the other wild-type allele, which can be detected by LOH. Traditionally, polymorphic markers, such as restriction fragment length polymorphisms (RFLPs) and microsatellite markers, have been used to detect LOH through allelotypic comparisons of DNA from a cancer sample and a matched normal sample [47]. However, this approach is tedious, labor intensive and requires a large amount of sample DNA, thus allowing only a modest number of markers to be screened. High-density, whole-genome allelotyping cannot be readily performed. The mapping of the human genome has allowed for the identification of millions of SNP loci [101], which makes them ideal markers for various genetic analyzes, including LOH. Due to their abundance, regular spacing and stability across the genome, SNPs have significant advantages over RFLPs and microsatellite markers as a basis for high-resolution, whole-genome allelotyping with accurate copy-number measurements. High-density oligonucleotide arrays have recently been generated to

support large-scale, high-throughput SNP analysis [48]. It is now possible to genotype over 100,000 SNP markers using the Affymetrix Mapping 100K SNP oligonucleotide array (soon to be 500K). LOH patterns generated by SNP array analysis have a high degree of concordance with previous microsatellite analyses of the same cancer samples [49]. Additionally, shared regions of LOH from SNP arrays can cluster lung cancer samples into subtypes [50], and distinct patterns of LOH are found to associate with clinical features in primary breast, bladder, head and neck, and prostate tumors [35,51–54]. A unique advantage of this SNP array-based approach is that the intensity of sample hybridization to the array probes can also be used to infer copy-number changes (similar to CGH) [33–35]. This unique feature has been explored by algorithms implemented in several independent bioinformatics and statistical software packages, including dChip-SNP [34] and Copy Number Analysis Tool (Affymetrix) [55]. Use of these novel analytic tools to analyze data from high-density SNP arrays now allows the DNA copy-number analyses to be combined with LOH analysis to distinguish copy-number gains, copy-number neutral LOH and copy-number losses, to comprehensively map the configuration of tumor genomes [34].

Cytogenetics-based approaches: old techniques with new twists

Cytogenetics has flourished since the introduction of chromosome banding techniques in 1969 [56,57]. One major drawback of these approaches is the requirement of *in vitro* culture and metaphase preparation of the cells of interest, which limits its application for many studies of solid cancers. Nevertheless, cytogenetic approaches will always have their place in the genomic profiling due to their ability to directly visualize chromosomal abnormalities. More importantly, these cytogenetic techniques complement CGH and LOH by providing information on chromosomal structural rearrangements that are not resolved by DNA copy-number analyses. For example, translocations are one of the most common genomic abnormalities in cancer, but they cannot be detected by CGH or LOH [13]. However, an experienced cytogeneticist can readily detect many forms of chromosomal translocations using classic cytogenetic techniques, such as chromosome banding technique (also known as karyotyping). A banding analysis usually involves blocking cells in mitosis, staining the condensed chromosomes with Giemsa dye (this dye stains regions of chromosomes that are rich in the base pairs adenine and thymine, which produces a dark band), and visualization under a light microscope. Karyotype analysis is performed over 500,000-times per year in the USA and Canada as part of standard clinical tests for prenatal and postnatal screening, as well as for the diagnosis of cancers (hematologic malignancies in particular). However, many cancer cells have complex karyotypes that are difficult to interpret (FIGURE 1). Recently, several new labeling techniques have been introduced in the field of molecular cytogenetics, including SKY, M-FISH, cross-species

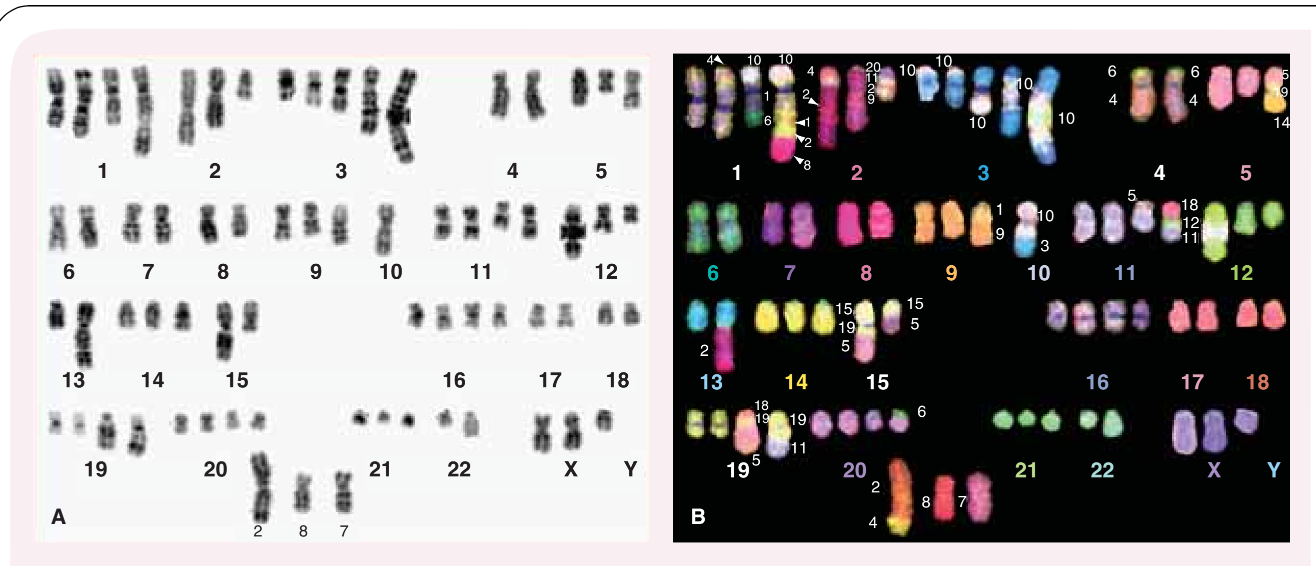


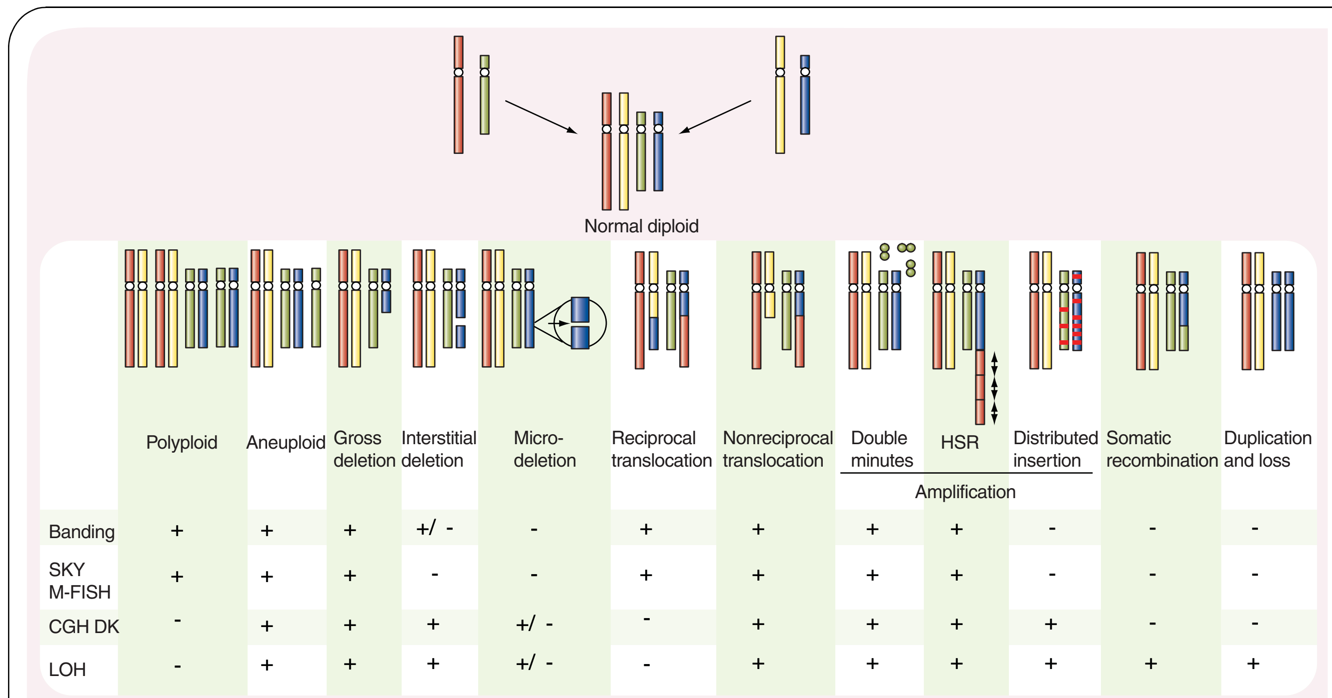
Figure 1. Comprehensive cytogenetic characterization of prostate cancer cell line CL-1. (A) Standard trypsin Giemsa banding technique-based karyotype analysis was performed on CL-1 cell line, a fast growing, highly tumorigenic and androgen-independent prostate cancer cell line [61]. **(B)** Multicolor fluorescence *in situ* hybridization (M-FISH) was performed for the CL1 cells using SpectraVysion assay according to manufacturer's instructions (Abbot-Vysis, IL, USA). The hybridized cells are visualized by fluorescence microscopy using a Zeiss Axiophot with an M-FISH filter configuration. The images are captured with a charge-coupled-device camera (Cohu Inc.) and analyzed using telomeric M-FISH software (Applied Imaging Corp., CA, USA). The precise karyotype for the CL-1 cells was defined by combining the information generated from both G-banding and M-FISH using the ISCN 1995 nomenclature [62]: 68, XX, del(X)(q11), der(1)(1;4)(p36;q31), +der(1)(1;6;10)(10pter→10p11.2::1p31–1q21::6p11→6pter), +der(1)t(1;8;10;12)(10pter→10p11.2::1p31→1q21::12q13→q15::8q22→qter), der(2)t(2;4)(p23:q31)X2,+der(2)(20::11::2::9),der(3)t(3;10)X2,+der(3)t(3;10)(3p21→q21::10q22→qter),der(4)t(4;6)(p16;q25)X2,+der(5)t(5::19::18)X2,+7,+8,+der(9)t(1;9)(1pter→1p32::9p13→9qter);der(10)t(3;10)(3qter→q25::10q22→pter), +der(10)t(3;10)X2, +der(11)t(5;11)(5::11q13→qter), iso(12q), +del(12)(q13), der(13)t(2;13)(q11.2;p11), +14, der(15)t(5;15)(5::q24→qter)X2, +16, +der(19)t(11;19)(11qter→11q23::19qter→pter), +der(19)t(5;19)(5q33→q13::19q13→p13), +der(20)t(6;20)(6::20qter→p11.2)X2, +21, del(22)(13), der(22)[CP5].

color banding (Rx-FISH) and multicolor chromosome banding. These techniques permit the simultaneous visualization of all chromosomes in different colors and, thus, considerably improve the detection of translocations or deletions. For example, both SKY and M-FISH use a combinatorial labeling scheme with spectrally distinguishable fluorochromes. The chromosome-specific probe pools (chromosome painting probes) are generated from flow-sorted chromosomes and then amplified and fluorescently labeled by degenerate oligonucleotide-primed PCR. The comprehensive analysis of complex chromosomal rearrangements present in tumor karyotypes was greatly improved through the introduction of these techniques in 1996 (FIGURE 1) [58–60].

Challenges for genomic studies of cancer

Three major biologic difficulties confront the identification and eventual translation of genomic markers and targets for cancer prediction, diagnostics, treatment and prognostics. The first is the diversity of genetic alterations that can contribute to malignant cell growth. Among these are germline variations that lead to hereditary cancer predispositions, the acquisition of transforming DNA or RNA sequences from cancer viruses, somatic mutations in the cancer genome and epigenetic mechanisms (such as DNA methylation or histone modification) that promote oncogenesis by modifying cancer-related genes. Somatic genomic alterations, such as point mutations, genomic amplifications or deletions, loss of allelic heterozygosity and chromosomal translocations, are believed to play a central role in the development of most solid tumors [63]. All of these mechanisms result in dysregulated expression of oncogenes and tumor-suppressor genes, but none of the existing genomic techniques can capture all of these genetic changes in a single analysis (FIGURE 2) [64]. This represents a major obstacle to the comprehensive analysis of tumor genomes and their relationship to clinical phenotypes.

The second challenge is the multifactorial nature of oncogenesis (i.e., the multi-hit model). Most tumors, especially those in adults, result from an interdependent series of genetic alterations, rather than a single decisive event. This complicates the task of prediction because a single abnormality may lead to cancer in some cases (e.g., those in which a complementary genetic alteration already exists), but not in others (e.g., in cases where a tumor-suppressor gene is mutated but no oncogene has been altered). Multifactorial etiology implies that the causal (or predictive) event is actually a combination of individual events, any one of which may forecast a benign outcome in isolation, but the combination of which may forecast malignancy. This complicates the development of prediction models because any single alteration is only inconsistently associated with malignancy. Since only certain combinations of complementary alterations consistently result in cancer, the discovery of predictive combinations is a major objective of analysis. However, the identification of these etiologic combinations from diverse streams of genomic data poses a major challenge. For example, a given genomic abnormality may not be functionally oncogenic if the damaged gene is not expressed. As a result, DNA structural analyses might incur noise by forecasting malignant



	Polyploid	Aneuploid	Gross deletion	Interstitial deletion	Micro-deletion	Reciprocal translocation	Nonreciprocal translocation	Double minutes	HSR	Distributed insertion	Somatic recombination	Duplication and loss
Banding	+	+	+	+/ -	-	+	+	+	+	-	-	-
SKY M-FISH	+	+	+	-	-	+	+	+	+	-	-	-
CGH DK	-	+	+	+	+/ -	-	+	+	+	+	-	-
LOH	-	+	+	+	+/ -	-	+	+	+	+	+	+

Figure 2. Detection and mapping of chromosomal abnormalities using different genomic and cytogenetic approaches. '+' and '-' denote effective and ineffective methods for the detection of a specific abnormality.
Banding: Chromosome banding or karyotyping; CGH: Comparative genomic hybridization; DK: Digital karyotyping; LOH: Loss of heterozygosity; M-FISH: Multicolor fluorescence *in situ* hybridization; SKY: Spectral karyotyping.

cell growth (e.g., based on gene copy-number analyses) when it does not, in fact, occur (e.g., because transcription factors or epigenetic influences prevent the expression of the aberrant gene). In this case, the addition of mRNA indications of gene expression and alterations in protein concentration would enhance the predictive significance of that genetic abnormality. Thus, the etiologic structure of cancer is combinatorial at two levels: one involving interactions among multiple genes (and/or biologic pathways) [65,66], and another involving interactions among multiple levels of analysis for a single gene (DNA structure, RNA expression and potential post-translational modifications). A powerful prediction model needs to account for all these factors and, as a result, there is a great need for comprehensive expression and proteomic analyses to complement the genome-wide DNA structural analyses outlined earlier.

A third biologic issue complicating the development of oncology prediction models is heterogeneity in oncogenic pathways. Distinct genetic lesions may give rise to a common malignant phenotype (i.e., there may be several different ways to contract oral cancer). Indeed, based on different genomic make-ups, cancer geneticists often consider a clinically identical cancer type as multiple distinct diseases that share a similar clinical phenotype. If there are many distinct etiologic possibilities, and only one is necessary to produce cancer, then many other etiologic variables may remain normal, despite the fact that a malignancy develops. Again, this complicates the development of prediction and classification models because no single etiologic event is consistently present in all cases. Furthermore, if there are many distinct etiologic pathways, no single etiologic event may even be present in the majority of cases. Thus, the prediction model needs to seek a set of potential classifiers, any one of which is sufficient to forecast malignancy. Traditional statistical classification models are essentially voting schemes in which each measured variable adds an independent indication of malignant potential that is then averaged together with all other measured variables' votes to produce an overall prediction of malignant potential [67]. When there are many variables measured, a large number of votes for 'no cancer' may dilute a small number of 'yes' votes, even though those yes votes are entirely sufficient to cause cancer at the biologic level.

The following sections suggest potential solutions that may overcome the challenges that hinder the identification and eventual translation of genomic markers and targets for cancer prediction, diagnostics, treatment and prognostics. In addition to these analytic difficulties, several technical obstacles also confront genomic studies of cancer, including limited quantities of tumor specimens, potential compromise of the genetic materials due to the archival procedures, and heterogeneity in tumor samples (i.e., the existence of multiple clones of tumor cells with different genomic characteristics in the same tumor sample). New approaches are being optimized to overcome these hurdles, including laser-capture microdissection (LCM) of single cells from clinical specimens [53,68–70], and whole-genome amplification techniques that can provide over 1000-fold amplification of the target nucleic acids with minimum bias [71–73].

Comprehensive genomic approaches

As discussed earlier, one of the major challenges in cancer research is to precisely delineate the complex genomic aberrations that shape tumor cell behavior and clinical outcomes. A feasible approach to overcome this problem is to combine a selective set of molecular genetic technologies such as CGH, LOH and various molecular cytogenetic analyses for comprehensive screening of genomic alterations with high resolution. Each of these techniques has its own unique advantages, but also its limitations, which have motivated efforts to combine these approaches as shown in FIGURE 2. In this instance, the SNP array-based LOH and CGH analyses provide a high-resolution mapping of copy-number abnormalities, but offer little information on chromosomal structure or spatial changes (e.g., translocations: the most common class of somatic mutation registered in the cancer-gene census [13]). On the contrary, modern cytogenetic techniques provide a clear picture of the gross chromosomal structure and spatial alterations, but have limited resolution. This is clearly illustrated in FIGURE 3, where concurrent cytogenetic and SNP array-based LOH analysis were performed on three myelodysplastic syndrome (MDS) samples. Two of the three cases exhibited concordant results between karyotyping and SNP array-based LOH. However, for case 1, the SNP array-based approach identified the loss of chromosome arm 5q, but failed to identify the translocation of chromosome 14 to chromosome 5 at the pericentromeric region FIGURE 3A. This translocation was identified by karyotyping and further confirmed by whole-chromosome paint (FIGURES 3B & C). These results illustrate the advantage of a multimodal approach to tumor genome analysis that combines the complementary strengths of array-based and cytogenetic approaches.

As aforementioned, this multimodal approach can be extended to combine DNA structural analyses with additional high-resolution mapping genome functional activity at the RNA and/or protein levels. Recent technical advances in microarray-based gene expression analysis have offered substantial improvement in diagnosis, treatment and prognosis of cancer patients. This continuous progress in microarray-based expression analysis, and the large public depositories of microarray data, have motivated new efforts to extract additional biologic information from these data in addition to the static RNA transcript levels. One such attempt involves inferring the chromosomal structural changes from spatially linked changes in microarray expression data [20–22,75]. Several array-CGH studies have shown a genome-wide correlation of gene expression with copy-number alterations, and have proved useful in individual amplicon refinement [76,77]. For example, through tissue microarray FISH and reverse transcriptase PCR, a minimally amplified region around *ERBB2* was identified in a large number of breast tumors; in addition, gene amplification was found to be correlated with increased gene expression in a subset of those samples [78]. Recently, several groups have observed that chromosomal alterations can lead to regional gene expression biases in human tumors and tumor-derived cell lines [20–22,79,80]. A recent study also demonstrated the correlation between SNP

array-based LOH profiles and expression profiles [51]. These studies suggest that a fraction of gene expression values (15–25%) are regulated in concordance with chromosomal DNA content [20–22,79,80]. Several statistical methods have been developed and have shown promising results for detecting DNA copy-number abnormalities based on differential gene expression [20–22,75]. As shown in FIGURE 4, using the authors' recently developed statistical model, a cytogenetically identified 10p chromosomal deletion was refined, based on the microarray expression data, and the boundaries of the deletion were mapped specifically between bands 10p14 and 10p12 [21]. These results were further confirmed by subtelomere FISH. These data demonstrate that it is feasible to use microarray differential expression data to identify significant DNA copy-number abnormalities, and that RNA-based gene expression analyses are concordant with DNA-based measures of chromosomal structural alteration. The development of bioinformatics techniques for reverse inference of DNA alterations from RNA expression data offer a new approach for genomic profiling that can provide cross validation of functional genomic alterations at multiple biologic levels when combined with DNA-based approaches such as CGH and LOH.

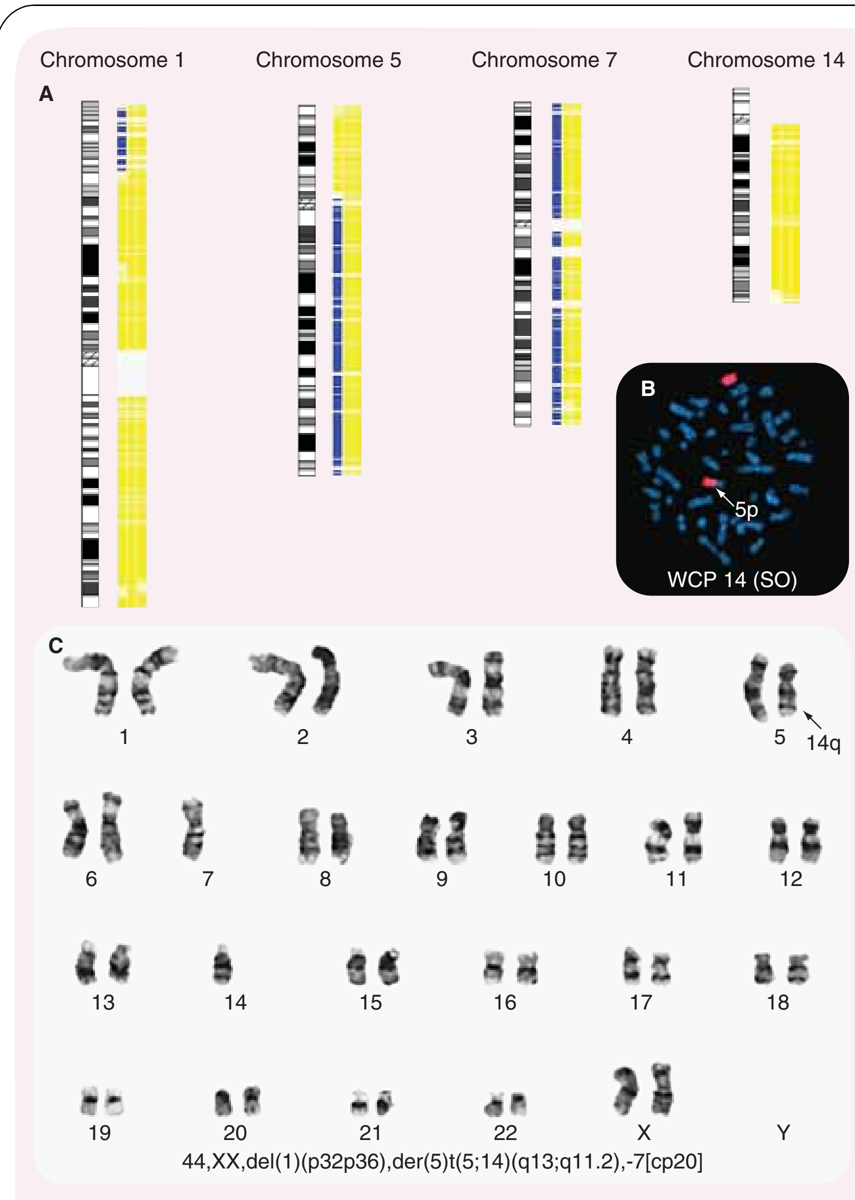


Figure 3. Comprehensive genomic analyses of a myelodysplastic syndrome using single nucleotide polymorphism (SNP) array-based approach and complementary cytogenetic approaches. (A) Three myelodysplastic syndrome cases were analyzed with a 10K SNP mapping array. The loss of heterozygosity (LOH) regions were detected and demarcated as described in [21,22]. The LOH patterns for chromosome 1, 5, 7 and 14 are shown. Concordant results between the SNP array-based LOH and karyotyping were observed for two of the three cases. In case 1, SNP array-based LOH demonstrated no loss of chromosome 14 material and a more extensive 5q deletion than interpreted by the karyotype. **(B)** Whole Chromosome Paint of chromosome 14 was performed to verify the results from LOH analysis in this case [74]. Two signals for chromosome 14 (red) were clearly identified, with one chromosome 14 translocated to 5q close to the pericentromeric region. **(C)** The karyotype for case 1 is presented. Together with the results from **A** and **B**, these data indicate a karyotype of 44,XX,del(1)(p32p36),der(5)t(5;14)(q13;q11.2),-7[cp20].

Additional functional genomic information can be derived from microarray gene expression data using bioinformatics analyses of upstream transcription factor dynamics. Several tools have recently been developed to identify aberrant transcription factor activity based on sequence similarities in the promotors of large groups of genes showing altered expression [81,82]. Aberrant transcription factor activity plays a central role in many solid tumors, and reverse inference of such alterations from microarray gene expression data provides another mechanism for cross validating the results of structural genomic surveys, thus suggesting that a particular transcription control pathway might be altered in a tumor.

Statistical considerations

Array-based techniques offer a promising approach for several types of DNA structural mapping (e.g., CGH and LOH), but these techniques also pose some new analytic challenges. The increasing use of array-based analyses stems, in part, from the phenomenal success of expression microarrays [83]. However, array-based CGH and LOH analyses require more complex preprocessing than expression data, because spatially adjacent probes can be expected to yield highly correlated signals (which may undermine assumptions of statistically independent signals made

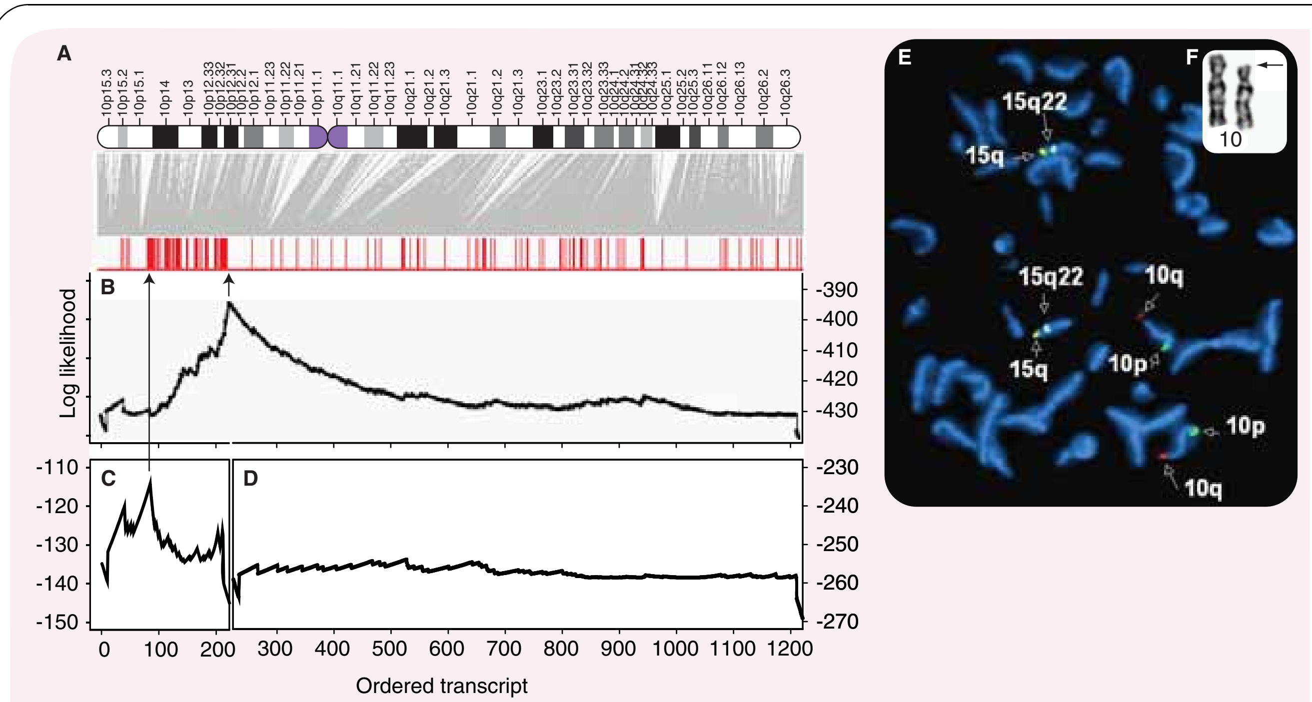


Figure 4. Mapping the boundaries of chromosomal deletion by differential expression. (A) Human Genome U133 Plus 2.0 array (Affymetrix) was used to generate expression profiles for del(10) cells (GM03047) and matching control. Underexpressed transcripts in del(10) cells were identified using Microarray Suite 5.0 (Affymetrix), with decreased transcription declared when the change p-value was less than 0.002. The transcripts were ordered according to sequence on chromosome 10, with red bars indicating the transcription start site of genes identified as significantly underexpressed in del(10) cells relative to a tissue-matched normal control cell. **(B)** As detailed in the authors' recent study, a single break-point model allowing differential density of underexpression was fit by maximum likelihood [21,22]. The log likelihood associated with breakpoints at each ordinal position on chromosome 10 is plotted (black line) with the maximum likelihood value serving as the estimated origin of copy-number abnormality. Grey lines map ordinal positions of each assayed transcript to its chromosomal location. Significant change in the prevalence of underexpressed transcripts was identified at ordered transcript 224 (28.1 Mb from 10pter), and agrees with the previously defined origin of deletion by cytogenetic analyses. **(C)** To determine whether deletion extended to the p-terminus, transcripts 1–223 were rescanned, and a second significant change in the prevalence of underexpressed transcripts was identified at ordered transcript 85 (12.2 Mb from 10pter). **(D)** No significant change in the prevalence of underexpressed transcripts was identified in the region ranging from ordered transcript 224 to 10qter. Together with the results from **B**, these data indicate a single partial deletion of chromosome 10p, spanning the region from 10p14–10p12. **(E)** Subtelomere fluorescence *in situ* hybridization verified results from the maximum likelihood expression-based analysis by confirming that the 10p deletion was interstitial with the intact subtelomere regions. Probes used are: 10ptel006 (10pter probe, green); 10qtel24 (10qter probe, red); PML (15q22 probe, aqua) and AFMA224XHI (15qter probe, yellow). Two normal signals for both 10p and 10q subtelomeres were clearly identified. **(F)** G-banded chromosome 10 of del(10) cell showing the deletion of the p arm of chromosome 10. Adapted with permission from [21] © 2005, British Medical Journal Publishing Group.

in many statistical analyses). New statistical approaches are required to accurately estimate low-level copy-number and heterozygosity changes [55,84–87]. Once these low-level structural changes have been mapped, higher order information can be extracted using many statistical techniques originally developed for expression arrays. For example, unsupervised clustering methods can discover novel subclasses by identifying groups of statistically related samples and/or genomic changes [67,88]. Supervised methods, which are more common in cancer research, are used to find features that could help classify samples into known classes [89]. At the center of this approach is the selection of features (e.g., specific chromosomal alterations) that are tumor related. The most common approach simply selects features that differ significantly across classes in univariate analyses (e.g., considered in isolation from other features) [90,91]. However, this first-order approach lacks the power to discover etiologically relevant features that are present only in subsets of the samples. As discussed earlier, the complexity of oncogenesis requires the identification of combinations and temporal relationships among features that link to cancer development. In principle, two interrelated components (parameters) are sought: classifiers that can discriminate disease status, and a set of roles for compiling these classifiers into a classification model. Although the magnitude of CGH and LOH data is lower than that of expression data, the small numbers of samples available in most experiments limits the performance of higher order analysis. In addition, the small sample size makes feature selection vulnerable to the instability caused by experimental and biologic variability [92]. These facts argue for the merit of establishing centralized databases and performing meta-analysis across data sets.

As discussed in the previous section, the expression level of a gene is often associated with its copy number. Lipson and coworkers argue that copy-number change, coupled with differential expression, may be a stronger indication of tumor relatedness than copy-number change alone [93]. It may also be

possible to incorporate information about a gene's functional characteristics in data analyses focusing on growth control pathways [94,95]. This approach is appealing when most genes in a functional module are differentially expressed at a moderate level, but few of the genes show drastic change. Examining groups of genes may also help suppress the effects of noise in the measurement of individual genes. However, this pathway approach to analysis is undermined by the fact that functional annotations are not complete, and tumors may involve only one or two decisive alterations, rather than coordinated changes in many genes in a pathway. The feasibility and clinical utility of functional module analyses in CGH and LOH data remain to be seen, but the more general principle of combining multiple streams of information to enhance analytic resolution represents a promising approach for enhancing interpretation of genome-wide surveys of DNA structural alterations in cancer. The optimal analysis may involve a synthesis of information on DNA structure, RNA expression, proteomic characteristics and functional (pathway) or regulatory (transcription factor) themes, which unite the ensembles of genes and proteins that show aberrant activity in a given tumor.

Expert commentary & five-year view

Microarray-based technologies for genome-wide assays of DNA and RNA characteristics have developed more rapidly than the ability to synthesize those assessments into a coherent functional portrait of cancer pathogenesis. Over the next 5 years, the key steps forward in realizing the promise of structural and functional genomics in cancer will come from the development of new informatics tools for combining multiple assays of genomic structure, functional genomic activity and their functional proteomic impacts.

Marriage of the top-down & bottom-up approaches

The genome-wide assay technologies outlined earlier are part of the growing number of top-down approaches that provide comprehensive genomic profiles that can be correlated with biologic and clinical status or functional aspects of tumors. In the near future, these analyses may help guide clinical treatment decisions and, in the long term, may substantially advance our fundamental understanding of tumor progression. More traditional bottom-up studies (so-called basic studies), which usually focused on individual genes or proteins, will continue to provide details that are not glimpsed by the top-down (global) approaches. Conversely, focused studies may be misinterpreted due to the lack of global information. Thus, the integration of bottom-up and top-down information will play a critical role in defining the functional pathogenesis of cancer and shaping its treatment. Pathway-based microarray data analysis represents one of the attempts to integrate top-down and bottom-up approaches. For example, the PathwayAssist™ package (Ariadne Genomics) and GeneWays™ (ExerGen Biosciences Inc.) employ language-processing algorithms to extract relationships between molecules by digesting published research literature and incorporating those links into the microarray data analysis algorithm [96,97]. Similar approaches may be employed to identify key functional targets within chromosomal alterations identified by the cytogenetic and array-based structural analyses outlined earlier. Integration of top-down and bottom-up analyses may also speed the identification of interventions that might ameliorate malignant cell growth (e.g., selecting drugs that target growth aberrations downstream of those arising from defined sites of genetic damage).

Data analysis & management

The rapid development of genome-wide profiling technologies has produced a substantial volume of new scientific findings, but it has not yet had a substantial impact on the clinical treatment of cancer. Part of the reason for this disconnection may involve the nature of the statistical analyses employed in most previous studies. As noted earlier, new analytic techniques will be required to integrate multiple streams of data into a coherent functional portrait of tumor pathogenesis. In addition, more attention needs to be paid to the external validity (clinical or biologic significance) of the results produced by various genomic and genetic data analysis tools. When a statistical tool identifies a difference, or a cluster of related genomic alterations or candidate genes, how often does this indication hold up in subsequent validation studies? Development of more sensitive, noise-resistant algorithms might enhance the information yield from massively parallel genomic measurements. There is probably still a long way to go before a satisfactory solution to these analytic problems is found, but substantial progress can be made if resources are focused in the right direction. With the potential to determine a large number of genomic alterations or candidate genes in parallel, new genomics technologies offer tremendous promise for basic biologic sciences and clinical diagnostics. However, their ultimate utility will depend critically on whether or not the search for efficient analytic methods meets with success.

Personalized therapies

Oncologists have long sought for targeted therapies for cancer that focus on the specific genetic lesions present in an individual patient's tumor. The implementation of this concept requires precise characterization of the disease as well as knowledge of patient background (e.g., genetic and environmental characteristics). For tumors with a relatively narrow range of critical genetic defects (e.g., acute promyelocytic leukemia and chronic-phase chronic myeloid leukemia), the development and deployment of targeted therapies has been more easily accomplished than in more complex and heterogeneous tumor types (e.g., breast cancer, and non-small cell lung cancer). The emergence of rapid, high-resolution genome analysis tools provides an opportunity to tackle these more heterogeneous malignancies at both the levels of basic research (e.g., defining pathogenetic mechanisms) and clinical treatment selection (e.g., determining which patients are suitable for therapies targeting a particular growth-control pathway). Advances in genome-wide profiling techniques and the development of new statistical and computational capabilities will play a key role in the expansion of personalized therapy for cancer.

Acknowledgements

This work was supported in part by NIH PHS grants R21 CA97771 and R01 DE015970–01 (to Wong), R21 AI49135 and R01 AI52737 (to Cole), and K22 DE014847, RO3 DE016569, RO3 CA114688, and a TRDRP grant 13KT-0028 (to Zhou). The authors thank Ronald Paquette at the Division of Hematology–Oncology and Nagesh Rao at the Department of Pathology & Laboratory Medicine, David Geffen School of Medicine, UCLA, for contributing to the figures.

Key issues

- There is, and has been, continuous improvement in genomic profiling technologies.
- Selective genomic technologies must be combined strategically to gain comprehensive genomic profiles.
- Statistical methods are being improved to unite multiple streams of data.
- Concurrent analysis of genomic data from multiple experimental platforms from multiple studies and laboratories is required.
- Large-scale discovery and validation studies are underway based on the comprehensive genomic database.
- Integration of basic research results with genome-wide profiling data will lead to a better understanding of the oncogenesis process.
- Genomics-based classifiers and roles hold promise for the implementation of personalized therapies for cancer treatment.

References

Papers of special note have been highlighted as:
• of interest
•• of considerable interest

1 Liehr T, Heller A, Starke H, Claussen U. FISH banding methods: applications in research and diagnostics. *Expert Rev. Mol. Diagn.* 2(3), 217–225 (2002).
• **Good review on fluorescence *in situ* hybridization.**

2 Nowell PC, Hungerford DA. Chromosome studies on normal and leukemic human leukocytes. *J. Natl Cancer Inst.* 25, 85–109 (1960).

3 Lander ES, Linton LM, Birren B *et al.* Initial sequencing and analysis of the human genome. *Nature* 409(6822), 860–921 (2001).

4 Venter JC, Adams MD, Myers EW *et al.* The sequence of the human genome. *Science* 291(5507), 1304–1351 (2001).

5 Schwaenen C, Nessling M, Wessendorf S *et al.* 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(4), 1039–1044 (2004).

6 Paris PL, Andaya A, Fridlyand J *et al.* Whole-genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. *Hum. Mol. Genet.* 13(13), 1303–1313 (2004).

7 Callagy G, Pharoah P, Chin SF *et al.* Identification and validation of prognostic markers in breast cancer with the complementary use of array-CGH and tissue microarrays. *J. Pathol.* 205(3), 388–396 (2005).

8 Weiss MM, Kuipers EJ, Postma C *et al.* Genomic alterations in primary gastric adenocarcinomas correlate with clinicopathological characteristics and survival. *Cell Oncol.* 26(5–6), 307–317 (2004).

9 Rosin MP, Cheng X, Poh C *et al.* Use of allelic loss to predict malignant risk for low-grade oral epithelial dysplasia. *Clin. Cancer Res.* 6(2), 357–362 (2000).

10 Martinez-Climent JA, Alizadeh AA, Segraves R *et al.* Transformation of follicular lymphoma to diffuse large cell lymphoma is associated with a heterogeneous set of DNA copy number and gene expression alterations. *Blood* 101(8), 3109–3117 (2003).

11 Rubio-Moscardo F, Climent J, Siebert R *et al.* Mantle-cell lymphoma genotypes identified with CGH to BAC microarrays define a leukemic subgroup of disease and predict patient outcome. *Blood* 105(11), 4445–4454 (2005).

12 Mundle SD, Sokolova I. Clinical implications of advanced molecular cytogenetics in cancer. *Expert Rev. Mol. Diagn.* 4(1), 71–81 (2004).
• **Review on genomic discoveries and cancer.**

13 Futreal PA, Coin L, Marshall M *et al.* A census of human cancer genes. *Nature Rev. Cancer* 4(3), 177–183 (2004).
• **Latest summary of cancer-related genes.**

14 Granville CA, Dennis PA. An overview of lung cancer genomics and proteomics. *Am. J. Respir. Cell. Mol. Biol.* 32(3), 169–176 (2005).

15 Jeffrey SS, Lonning PE, Hillner BE. Genomics-based prognosis and therapeutic prediction in breast cancer. *J. Natl Compr. Canc. Netw.* 3(3), 291–300 (2005).

16 Avivi I, Rowe JM. Prognostic factors in acute myeloid leukemia. *Curr. Opin. Hematol.* 12(1), 62–67 (2005).

17 Wang TL, Maierhofer C, Speicher MR *et al.* Digital karyotyping. *Proc. Natl Acad. Sci. USA* 99(25), 16156–16161 (2002).

18 Imoto H, Hirotsune S, Muramatsu M *et al.* Direct determination of *NotI* cleavage sites in the genomic DNA of adult mouse kidney and human trophoblast using whole-range restriction landmark genomic scanning. *DNA Res.* 1(5), 239–43 (1994).

19 Lisitsyn N, Wigler M. Cloning the differences between two complex genomes. *Science* 259(5097), 946–951 (1993).

20 Crawley JJ, Furge KA. Identification of frequent cytogenetic aberrations in hepatocellular carcinoma using gene-expression microarray data. *Genome Biol.* 3(12), RESEARCH0075 (2002).

21 Zhou X, Cole SW, Chen Z, Li Y, Wong DT. Identification of discrete chromosomal deletion by binary recursive partitioning of microarray differential expression data. *J. Med. Genet.* 42(5), 416–419 (2005).

22 Zhou X, Cole SW, Hu S, Wong DT. Detection of DNA copy number abnormality by microarray expression analysis. *Hum. Genet.* 114(5), 464–467 (2004).

23 Kallioniemi A, Kallioniemi OP, Sudar D *et al.* Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* 258(5083), 818–821 (1992).

24 Pinkel D, Albertson DG. Array comparative genomic hybridization and its applications in cancer. *Nature Genet.* 37(Suppl.), S11–S17 (2005).
• **Good review on comparative genomic hybridization (CGH).**

25 Pinkel D, Albertson DG. Comparative genomic hybridization. *Ann. Rev. Genomics Hum. Genet.* 6, 331–354 (2005).

26 Pinkel D, Segraves R, Sudar D *et al.* High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nature Genet.* 20(2), 207–211 (1998).

27 Ishkanian, AS, Malloff CA, Watson SK *et al.* A tiling resolution DNA microarray with complete coverage of the human genome. *Nature Genet.* 36(3), 299–303 (2004).

28 Mantripragada KK, Buckley PG, de Stahl TD, Dumanski JP. Genomic microarrays in the spotlight. *Trends Genet.* 20(2), 87–94 (2004).

29 Pollack JR, Perou CM, Alizadeh AA *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.* 23(1), 41–46 (1999).

30 Zhou X, Jordan RCK, Mok S, Birrer MJ, Wong DT. DNA copy number abnormality of oral squamous cell carcinoma detected by cDNA array-based CGH. *Cancer Genet. Cytogen.* 151(1), 90–92 (2004).

31 Brennan C, Zhang Y, Leo C *et al.* High-resolution global profiling of genomic alterations with long oligonucleotide microarray. *Cancer Res.* 64(14), 4744–4748 (2004).

32 Lucito R, Healy J, Alexander J *et al.* Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. *Genome Res.* 13(10), 2291–2305 (2003).

33 Bignell GR, Huang J, Greshock J *et al.* High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res.* 14(2), 287–295 (2004).

34 Zhao X, Li C, Paez JG *et al.* An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res.* 64(9), 3060–3071 (2004).

- **Concurrent loss-of-heterozygosity (LOH) and CGH analysis using single nucleotide polymorphism (SNP) arrays.**

35 Zhou X, Mok SC, Chen Z, Li Y, Wong DT. Concurrent analysis of loss of heterozygosity (LOH) and copy number abnormality (CNA) for oral premalignancy progression using the Affymetrix 10K SNP mapping array. *Hum. Genet.* 115(4), 327–330 (2004).

- **Concurrent LOH and CGH analysis using SNP arrays.**

36 Zhao X, Weir BA, LaFramboise T *et al.* Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer Res.* 65(13), 5561–70 (2005).

37 Davies JJ, Wilson IM, Lam WL. Array CGH technologies and their applications to cancer genomes. *Chromosome Res.* 13(3), 237–248 (2005).

38 Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. *Science* 270(5235), 484–487 (1995).

39 Shih Ie M, Sheu JJ, Santillan A *et al.* Amplification of a chromatin remodeling gene, Rsf-1/HBXAP, in ovarian carcinoma. *Proc. Natl Acad. Sci. USA* 102(39), 14004–14009 (2005).

40 Di C, Liao S, Adamson DC *et al.* Identification of OTX2 as a medulloblastoma oncogene whose product can be targeted by all-*trans* retinoic acid. *Cancer Res.* 65(3), 919–924 (2005).

41 Wang TL, Diaz LA Jr, Romans K *et al.* Digital karyotyping identifies thymidylate synthase amplification as a mechanism of resistance to 5-fluorouracil in metastatic colorectal cancer patients. *Proc. Natl Acad. Sci. USA* 101(9), 3089–3094 (2004).

42 Parrett TJ, Yan H. Digital karyotyping technology: exploring the cancer genome. *Expert Rev. Mol. Diagn.* 5(6), 917–25 (2005).

- **Good review on digital karyotyping.**

43 Shih Ie M, Wang TL. Apply innovative technologies to explore cancer genome. *Curr. Opin. Oncol.* 17(1), 33–38 (2005).

44 Knudson AG Jr. Mutation and cancer: statistical study of retinoblastoma. *Proc. Natl Acad. Sci. USA* 68(4), 820–823 (1971).

45 Knudson AG. Hereditary cancer: two hits revisited. *J. Cancer Res. Clin. Oncol.* 122(3), 135–140 (1996).

46 Friend SH, Bernards R, Rogelj S *et al.* A human DNA segment with properties of the gene that predisposes to retinoblastoma and osteosarcoma. *Nature* 323, 643–646 (1986).

47 Vogelstein B, Fearon ER, Kern SE *et al.* Alleotype of colorectal carcinomas. *Science* 244, 201–211 (1989).

48 Wang DG, Fan JB, Siao CJ *et al.* Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280(5366), 1077–1082 (1998).

49 Lindblad-Toh K, Tanenbaum DM, Daly MJ *et al.* Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. *Nature Biotechnol.* 18(9), 1001–1005 (2000).

50 Janne PA, Li C, Zhao X *et al.* High-resolution single-nucleotide polymorphism array and clustering analysis of loss of heterozygosity in human lung cancer cell lines. *Oncogene* 23(15), 2716–2726 (2004).

51 Wang ZC, Lin M, Wei LJ *et al.* Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. *Cancer Res.* 64(1), 64–71 (2004).

52 Hoque MO, Lee CC, Cairns P, Schoenberg M, Sidransky D. Genome-wide genetic characterization of bladder cancer: a comparison of high-density single-nucleotide polymorphism arrays and PCR-based microsatellite analysis. *Cancer Res.* 63(9), 2216–2222 (2003).

53 Lieberfarb ME, Lin M, Lechpammer M *et al.* Genome-wide loss of heterozygosity analysis from laser-capture microdissected prostate cancer using single nucleotide polymorphic allele (SNP) arrays and a novel bioinformatics platform dChipSNP. *Cancer Res.* 63(16), 4781–4785 (2003).

54 Zhou X, Li C, Mok SC, Chen Z, Wong DTW. Whole-genome loss of heterozygosity profiling on oral squamous cell carcinoma by high-density single nucleotide polymorphic allele (SNP) array. *Cancer Genet. Cytogenet.* 151(1), 82–84 (2004).

55 Huang J, Wei W, Zhang J *et al.* Whole-genome DNA copy number changes identified by high-density oligonucleotide arrays. *Hum. Genomics* 1(4), 287–299 (2004).

56 Caspersson T, Zech L, Modest EJ, Foley GE, Wagh U, Simonsson E. Chemical differentiation with fluorescent alkylating agents in *Vicia faba* metaphase chromosomes. *Exp. Cell Res.* 58(1), 128–140 (1969).

57 Caspersson T, Zech L, Modest EJ, Foley GE, Wagh U, Simonsson E. DNA-binding fluorochromes for the study of the organization of the metaphase nucleus. *Exp. Cell Res.* 58(1), 141–152 (1969).

58 Schrock E, du Manoir S, Veldman T *et al.* Multicolor spectral karyotyping of human chromosomes. *Science* 273(5274), 494–497 (1996).

59 Liyanage M, Coleman A, du Manoir S *et al.* Multicolour spectral karyotyping of mouse chromosomes. *Nature Genet.* 14(3), 312–315 (1996).

60 Speicher MR, Gwyn Ballard S, Ward DC. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genet.* 12(4), 368–375 (1996).

61 Liu AY, Brubaker KD, Goo YA *et al.* Lineage relationship between LNCaP and LNCaP-derived prostate cancer cell lines. *Prostate* 60(2), 98–108 (2004).

62 International System for Human Cytogenetic Nomenclature. *Guidelines for cancer cytogenetics. Supplement to an International System for Human Cytogenetic Nomenclature.* Mitelman F (Ed.), Karger, Switzerland (1995).

63 Weir B, Zhao X, Meyerson M. Somatic alterations in the human cancer genome. *Cancer Cell* 6(5), 433–438 (2004).

64 Albertson DG, Collins C, McCormick F, Gray JW. Chromosome aberrations in solid tumors. *Nature Genet.* 34(4), 369–76 (2003).

•• **Comprehensive review in cancer genomics.**

65 Vogelstein B, Kinzler KW. *The Genetic Basis of Human Cancer.* McGraw-Hill, Toronto, Canada, (2002).

66 Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nature Med.* 10(8), 789–799 (2004).

67 Golub TR, Slonim DK, Tamayo P *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286(5439), 531–537 (1999).

68 Ohyama H, Mahadevappa M, Luukkaa H, Todd R, Warrington J, Wong DT. Use of LCM-generated targets for hybridization of high-density oligonucleotide arrays. In: *Methods in Enzymology.* Conn PM (Ed.), 356, 323–333 (2002).

69 Ohyama H, Zhang X, Kohno Y *et al.* Laser capture microdissection-generated target sample for high-density oligonucleotide array hybridization. *Biotechniques* 29(3), 530–536 (2000).

70 Simone NL, Remaley AT, Charboneau L *et al.* Sensitive immunoassay of tissue cell proteins procured by laser capture microdissection. *Am. J. Pathol.* 156(2), 445–452 (2000).

71 Zhou X, Teman S, Chen Z, Ye H, Mao L, Wong DT. Allelic imbalance analysis of oral tongue cancer by high-density single nucleotide polymorphism arrays using whole-genome-amplified DNA. *Hum. Genet.* (2005) (In Press).

72 Dean FB, Hosono S, Fang L *et al.* Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl Acad. Sci. USA* 99(8), 5261–5266 (2002).

73 Paez JG, Lin M, Beroukhim R *et al.* Genome coverage and sequence fidelity of ϕ29 polymerase-based multiple strand displacement whole genome amplification. *Nucleic Acids Res* 32(9), e71 (2004).

74 Pettenati MJ, Jackle B, Bobby P *et al.* Unexpected retention and concomitant loss of subtelomeric regions in balanced chromosome anomalies by FISH. *Am. J. Med. Genet.* 111(1), 48–53 (2002).

75 Myers CL, Dunham MJ, Kung SY, Troyanskaya OG. Accurate detection of aneuploidies in array CGH and gene expression microarray data. *Bioinformatics* 20(18), 3533–3543 (2004).

76 Pollack JR, Sorlie T, Perou CM *et al.* Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl Acad. Sci. USA* 99(20), 12963–12968 (2002).

77 Wolf M, Mousses S, Hautaniemi S *et al.* High-resolution analysis of gene copy number alterations in human prostate cancer using CGH on cDNA microarrays: impact of copy number on gene expression. *Neoplasia* 6(3), 240–247 (2004).

78 Kauraniemi P, Kuukasjarvi T, Sauter G, Kallioniemi A. Amplification of a 280-kilobase core region at the ERBB2 locus leads to activation of two hypothetical proteins in breast cancer. *Am. J. Pathol.* 163(5), 1979–1984 (2003).

79 Phillips JL, Hayward SW, Wang Y *et al.* The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. *Cancer Res.* 61(22), 8143–8149 (2001).

80 Virtaneva K, Wright FA, Tanner SM *et al.* Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. *Proc. Natl Acad. Sci. USA* 98(3), 1124–1129 (2001).

81 Cole SW, Yan W, Galic Z, Arevalo J, Zack JA. Expression-based monitoring of transcription factor activity: the TELiS database. *Bioinformatics* 21(6), 803–810 (2005).

82 Frith MC, Fu Y, Yu L, Chen JF, Hansen U, Weng Z. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res* 32(4), 1372–1381 (2004).

83 Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270(5235), 467–470 (1995).

84 Autio R, Hautaniemi S, Kauraniemi P *et al.* CGH-Plotter: MATLAB toolbox for CGH-data analysis. *Bioinformatics* 19(13), 1714–1715 (2003).

85 Eilers PHC, de RX Menezes. Quantile smoothing of array CGH data. *Bioinformatics* 21(7), 1146–1153 (2005).

86 Fridlyand J, Snijders AM, Pinkel D, Albertson DG, Jain AN. Hidden Markov models approach to the analysis of array CGH data. *J. Multivariate Anal.* 90(1), 132–153 (2004).

87 Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* 20(8), 1233–1240 (2004).

• **Useful tool for concurrent LOH and CGH analyses using SNP array.**

88 Armstrong NJ, van de Wiel MA. Microarray data analysis: from hypotheses to conclusions using gene expression data. *Cell Oncol.* 26(5–6), 279–290 (2004).

89 Dudoit S, Fridlyand J, Speed T. Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Am. Stat. Assoc.* 97, 77–87 (2002).

90 Massion PP, Kuo WL, Stokoe D *et al.* Genomic copy number analysis of non-small cell lung cancer using array comparative genomic hybridization: Implications of the phosphatidylinositol 3-kinase pathway. *Cancer Res.* 62(13), 3636–3640 (2002).

91 Veltman JA, Fridlyand J, Pejavar S *et al.* Array-based comparative genomic hybridization for genome-wide screening of DNA copy number in bladder tumors. *Cancer Res.* 63(11), 2872–2880 (2003).

92 Gordon GJ, Rockwell GN, Godfrey PA *et al.* Validation of genomics-based prognostic tests in malignant pleura mesothelioma. *Clin. Cancer Res.* 11(12), 4406–4414 (2005).

93 Lipson D, Ben-Dor A, Dehan E, Yakhini Z. Joint analysis of DNA copy numbers and gene expression levels. *Algorithms in Bioinformatics, Proceedings* 3240, 135–146 (2004).

94 Smid M, Dorssers LCJ. GO-Mapper: functional analysis of gene expression data using the expression level as a score to evaluate gene ontology terms. *Bioinformatics* 20(16), 2618–2625 (2004).

95 Volinia S, Evangelisti R, Francioso F, Arcelli D, Carella M, Gasparini P. GOAL: automated gene ontology analysis of expression profiles. *Nucl. Acids Res.* 32, W492–W499 (2004).

96 Nikitin A, Egorov S, Daraselia N, Mazo I. Pathway studio – the analysis and navigation of molecular networks. *Bioinformatics* 19(16), 2155–2157 (2003).

97 Rzhetsky A, Iossifov I, Koike T *et al.* GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data. *J. Biomed. Inform.* 37(1), 43–53 (2004).

Website

101 NCBI: Single Nucleotide Polymorphism www.ncbi.nlm.nih.gov/SNP

Affiliations

- *Xiaofeng Zhou*
  Dental Research Institute, School of Dentistry & Jonsson Comprehensive Cancer Center, University of California at Los Angeles, Los Angeles, CA, USA
  Tel.: +1 310 825 7210
  Fax: +1 310 794 7109
  xfzhou@ucla.edu
- *Tianwei Yu*
  Dental Research Institute, School of Dentistry, University of California at Los Angeles, Los Angeles, CA, USA
  Tel.: +1 310 206 1139
  Fax: +1 310 794 7109
  tyu@stat.ucla.edu
- *Steven W Cole*
  Division of Hematology–Oncology, Department of Medicine, David Geffen School of Medicine; Jonsson Comprehensive Cancer Center & Molecular Biology Institute, University of California at Los Angeles, Los Angeles, CA, USA
  Tel.: +1 310 267 4243
  Fax: +1 310 825 6192
  coles@ucla.edu
- *David TW Wong*
  Dental Research Institute, School of Dentistry & Division of Head & Neck Surgery/Otolaryngology, David Geffen School of Medicine & Henry Samueli School of Engineering & Jonsson Comprehensive Cancer Center, University of California at Los Angeles, Los Angeles, CA, USA
  Tel.: +1 310 206 3048
  Fax: +1 310 794 7109
  dtww@ucla.edu