

Discovering Tumor Suppressor Genes Through Genome-Wide Copy Number Analysis

S. Michael Rothenberg and Jeff Settleman*

Massachusetts General Hospital Cancer Center and Harvard Medical School, 149, 13th Street, Charlestown, MA 02129, USA

Abstract: Classical tumor suppressor gene discovery has largely involved linkage analysis and loss-of-heterozygosity (LOH) screens, followed by detailed mapping of relatively large chromosomal regions. Subsequent efforts made use of genome-wide PCR-based methods to detect rare homozygous deletions. More recently, high-resolution genomic arrays have been applied to cancer gene discovery. However, accurate characterization of regions of genomic loss is particularly challenging due to sample heterogeneity, the small size of deleted regions and the high frequency of germline copy number polymorphisms. Here, we review the application of genome-wide copy number analysis to the specific problem of identifying tumor suppressor genes.

Received on: February 18, 2010 - Revised on: April 07, 2010 - Accepted on: April 07, 2010

Keywords: Array CGH, copy number analysis, cancer, tumor suppressor genes.

INTRODUCTION

Ever since the seminal demonstration by Bishop and Varmus more than 30 years ago that transforming oncogenes arise from normal cellular genes, the discovery of additional cancer-specific mutations has fueled the hope of diagnosing and treating cancer with tumor-specific modalities [1]. However, although identifying the driving genetic lesions for select cancers has led to improved diagnostics and therapies in some cases, the total number of patients who benefit from these approaches remains a relatively small fraction of all patients suffering from cancer [2-7]. To identify additional cancer genes, a variety of unbiased, genome-wide approaches have emerged that can detect a large fraction of the genetic differences between human tumors and their normal tissue counterparts. Among these, the development of array-based genomic hybridization technologies has paved the way for interrogating an ever increasing density of probes for markers distributed across the cancer genome for increases and decreases in intensity suggestive of somatic alterations causally related to tumor formation. To facilitate the analysis of the very large datasets generated by such methodologies, several automated approaches have been developed to systematically refine the number of candidate cancer genes for further genetic and functional characterization. These have led to the nomination of numerous candidates whose biological significance requires deeper genetic analysis and robust functional validation. Here, we review the application of array-based gene copy number analysis to the specific problem of identifying tumor suppressor genes. We start by reviewing the discovery of *bona fide* tumor suppressor genes for common themes that may guide the interpretation of

candidates identified through genomic approaches. Next, we review the development of genome-wide approaches aimed at the global identification of copy number alterations in cancer. We then focus on array-based copy number analysis and review the major steps in converting probe-level intensity measurements to deleted regions harboring candidate tumor suppressor genes. Finally, we discuss specific examples of new candidate tumor suppressor genes identified through genome-wide copy number analysis and we summarize the evidence for their involvement in human cancer.

DISCOVERING *BONA FIDE* TUMOR SUPPRESSOR GENES

Historically, genetic approaches for discovering tumor suppressor genes (TSGs) have required validating several candidate genes contained within a large chromosomal region implicated by linkage analysis in families afflicted with cancer predisposition syndromes or targeted by tumor-specific allelic loss (so called loss-of-heterozygosity/LOH). Several common themes have emerged from these efforts (Table 1).

1. Rare Cases Associated with Focal (Homozygous) Deletions can be Informative

In the “pre-genomic” era, the chromosomal regions implicated in cancer by linkage analysis or LOH screens were usually very large (on the order of many megabases), and would therefore normally require many iterative steps of genotyping with many more genetic markers, and eventually, the sequencing of many candidate genes to detect inactivating mutations indicative of the targeted TSG. In most cases, uncovering the targeted gene relied on the discovery of rare cases with focal deletions (germline heterozygous deletions for genes predisposing to cancer and somatic tumor-specific homozygous deletions for sporadic tumors) encompassing one or a few genes (Table 1). In fact, the recognition of the

*Address correspondence to this author at the Massachusetts General Hospital Cancer Center and Harvard Medical School, 149, 13th Street, Charlestown, MA 02129, USA; Tel: 617-724-9556; Fax: 617-643-5410; E-mail: settleman@helix.mgh.harvard.edu

Table 1. Examples of *Bona fide* Tumor Suppressor Genes Identified by Focal Deletions

| Gene | Methodology | Size of minimal deletion | Reference |
|-----------------|------------------|----------------------------|-----------|
| <i>CDKN2A/B</i> | Deletion mapping | intragenic | [17,22] |
| <i>Rb1</i> | Linkage* | intragenic | [13] |
| <i>APC</i> | Linkage | three genes | [14,16] |
| <i>WT1</i> | Linkage | 25 nucleotides | [12,15] |
| <i>SMAD4</i> | LOH | 2Mb | [21] |
| <i>PTEN</i> | RDA, LOH | intragenic | [18,20] |
| <i>WTX</i> | array CGH | 1-3 oligonucleotide probes | [19] |

*Homozygous deletions in sporadic tumors.

value of tumor-specific focal homozygous deletions to implicate new TSGs provided much of the impetus for the development of unbiased, genome-wide approaches for detecting them (see discussion below) [8-11].

2. Very Small Deletions can Inactivate Genes

In theory, deletion of a single coding nucleotide will cause a frameshift mutation that can lead to gene inactivation through protein truncation. Although detecting single nucleotide deletions essentially requires single nucleotide sequencing, very small deletions encompassing one or a few genes have been instrumental in cloning many important TSGs (Table 1) [12-22]. Identifying the *APC* gene relied on an analysis of two cases with focal deletions spanning just three genes [14,16]. Focal homozygous deletions in cancer cell lines were instrumental in the identification of *INK4A/B* as one of the most common targets of genetic inactivation in human cancers [17,22]. Loss of heterozygosity affecting nearly the entire long arm of chromosome 10 is extremely common in primary gliomas, and the identification of *PTEN* was made possible by the discovery of rare cell lines with intragenic homozygous deletions affecting *PTEN* and no other flanking genes [18,20]. A deletion of just 25 nucleotides (spanning an exon-intron boundary and leading to aberrant splicing) was critical to implicating *WT1* in Wilms' tumor [12,15]. As can be seen in Fig. (1), a high resolution scan of genomic loci known to be targeted by homozygous deletions reveals both broad deletions encompassing several genes and very focal, intragenic deletions disrupting just one or a few coding exons and therefore directly implicating the individual, targeted gene.

3. Detecting Homozygous Deletions is Extremely Sensitive to Background Effects and Stromal Admixture

Each of the examples described above relied on the presence of the deletion in every cell within the analyzed samples, either because the deletion was present in the germline (e.g. for cancer predisposing TSGs) [12,14-17] or because cancer cell lines or primary cultures (presumably, clonally-derived) were utilized (for somatically inactivated TSGs) [17,20]. For both PCR-based approaches and, as discussed below, array-based copy number analyses, the detection of deletions may be particularly confounded by the presence of mixed tumor-associated stroma, since the presence of non-cancer diploid genomes can lead to preferential detection of the non-deleted, normal DNA [8,23]. In addition, accurately measuring genomic losses is inherently more difficult than detecting amplifications, since losses are limited to only one or two copies, and any residual signal will be extremely close to the lower limits of detection of the platform being utilized.

Each of these features may significantly impact the discovery of candidate TSGs by genome-wide copy number analysis. Small deletions consisting of one or a few genomic probes with hybridization signals barely above background may easily be obscured within the significant noise associated with the simultaneous measurement of literally hundreds of thousands of genetic markers. The invariable presence of stromal cells in primary tumor specimens may further decrease the ability to detect discrete losses indicative of novel TSGs. Therefore, the specific consideration of background effects and stromal admixture by individual genomic platforms and computational algorithms is critical for estab-

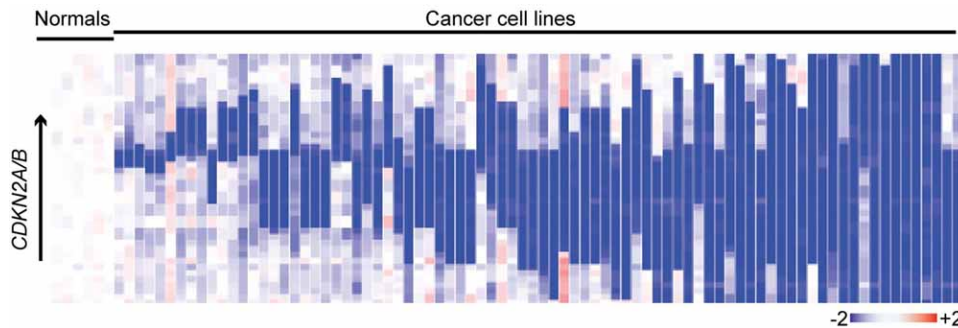


Fig. (1). Homozygous deletions of *CDKN2A/B* in human cancer cell lines. Heatmap of log[2] copy number demonstrates targeting of this locus by both broad homozygous deletions in blue (red indicates gains) affecting neighboring genes (rightmost samples) and intragenic homozygous deletions (leftmost samples) that remove individual coding exons without impacting neighboring genes. This figure was adapted from [85].

lishing candidate TSGs proposed on the basis of genome-wide copy number analyses. The sensitivity of deletions to these confounding effects may be contrasted with chromosomal regions of copy number gain/amplification that may harbor novel oncogenes. Unlike inactivating deletions, gains that may lead to oncogene activation through increased expression must extend at least the entire length of the targeted gene (and in practice often span several genes). In addition, the degree of amplification may be many times the normal diploid copy number and therefore will be less subject to the confounding effects of background signal and admixed stroma.

IDENTIFYING TUMOR SUPPRESSOR GENES BY GENOME-WIDE COPY NUMBER ANALYSIS

PCR-Based Approaches

Initial unbiased efforts to screen the cancer genome for regions of genomic loss that potentially harbor tumor suppressor genes utilized PCR to enrich the DNA sequences present in normal genomic DNA that were absent from tumor DNA. Among several methods developed to identify the differences between two complex genomes, representation differentiation analysis (RDA) has been most successfully used to identify novel TSGs [8,9]. The RDA methodology consists of restriction digestion, subtractive hybridization and selective PCR amplification to generate simplified “representations” of the genome using mixed tumor and matched normal samples. RDA is extremely powerful for detecting rare differences in largely identical genomes (e.g. viral sequences, tumor samples with matched normal DNA available). RDA was instrumental in the cloning of *PTEN*, *BRCA2* and other candidate tumor suppressor genes [18,24,25]. However, several features have complicated its use as a screening tool for cancer gene identification: (1) technical complexity (which limits the number of samples that can be analyzed); (2) genomic instability in solid tumors, leading to the detection of many candidate regions that appear to be decreased in tumor DNA (compared to normal); (3) copy number polymorphisms, which, in the absence of matched normal tissue DNA, lead to many false positives; and (4) the fact that candidate regions are identified *en masse*, and must be subsequently deconvoluted to permit validation of individual loci. Nevertheless, several improvements have been made to increase the power of RDA. Using tumor genomic DNA from genetically engineered mouse models and strain-matched normal DNA can decrease the impact of germline copy number variation, while combining the initial subtraction and amplification steps with array-based hybridization permits rapid deconvolution of complex mixtures of differentially amplified loci (representational oligonucleotide array analysis/ROMA—see below) [25,26].

MICROARRAY-BASED COMPARATIVE GENOMIC HYBRIDIZATION (CGH)

The initial demonstration of tumor-specific genomic imbalance by traditional metaphase karyotyping has led to increasingly sensitive methods for detecting cytogenetic abnormalities in cancer. With comparative genomic hybridization (CGH), differentially fluorophore-labelled tumor (e.g. fluorescein) and normal (e.g. rhodamine) DNA are co-

hybridized to normal metaphase spreads in the presence of unlabelled Cot-1 DNA (to block off-target binding to repetitive sequences), permitting quantitative detection of discrete regions of chromosomal gain or loss (as determined by the ratio of fluorescein-to-rhodamine fluorescence) [27]. CGH has permitted the cataloguing of human tumors by recurrent regions of gains and losses that correlate with clinical features and have helped to pinpoint causative cancer genes [28]. However, the use of metaphase chromosomes to localize genomic alterations and the high complexity of the probes limit its detection resolution to ~20Mb. A related methodology, spectral karyotyping (SKY), combines multi-color fluorescence in situ hybridization with spectral analysis that has been particularly valuable for detecting chromosomal translocations, though still with limited resolution (1-2Mb) [28].

The development of DNA microarrays, initially for mRNA expression analysis, provided two significant advantages over traditional CGH: (1) increased resolution, limited only by the density of genomic probes on the array (expected to increase further with the completion of the human genome sequencing project) [29,30]; and (2) highly parallel analysis, permitting the identification of individual candidate regions (for subsequent validation) in a single step. Initial applications of array-based CGH used large insert genomic DNA (e.g. BAC) clones that provided higher resolution (~40kB) than traditional CGH and the quantitative measurement of DNA gains and losses. However, increasing the resolution further was limited by probe complexity as well as the chemistry used to couple the genomic probes to the array surface, such that smaller probes were not bound effectively [10,31,32]. The development of complementary DNA (cDNA) microarrays and their application to the determination of DNA copy-number variation in human tumor samples represented a significant advance in genomic analysis of tumors by providing the first genome-wide view of copy number alterations at gene resolution [11,30]. In addition, the demonstrated ability to robustly detect single and two-copy losses (despite the much greater complexity of genomic DNA compared to cellular mRNA) is critical to detecting tumor suppressor genes, since genomic losses are discrete and limited in amplitude (e.g. loss of one or both alleles) compared to gains [11]. Another important advantage of cDNA microarray-based CGH is the ability to characterize both copy number and gene expression patterns in parallel in the same samples and on the same array platform, enabling candidate genes to be narrowed to those whose genomic gain or loss is associated with corresponding changes in gene expression (and thus are the likely targets of the alteration).

Arrays of oligonucleotide probes were developed in parallel to cDNA arrays [26]. The major platforms in current use consist of arrayed probes of 60-70 nucleotides for CGH (e.g. Agilent); or 25 nucleotide probes (e.g. Affymetrix), on which tumor and reference hybridizations are carried out separately, with copy number inferred indirectly by comparison. In addition to determining copy number, Affymetrix arrays are designed to detect single nucleotide polymorphisms (SNPs), and therefore, can be used for simultaneous determination of copy number and LOH at every analyzed locus from a single hybridization, even in the absence of a matched normal sample [23,33]. This may permit determina-

tion of the actual genetic mechanism underlying LOH events (e.g. copy-reducing from copy-neutral), and could reveal additional samples for deeper sequencing to identify inactivating mutations (e.g. point mutations) not apparent from copy number analysis. An important feature of most oligonucleotide probe-based methodologies is the requirement that the complexities of the analyzed genomes be significantly reduced (by restriction digest and PCR) to permit accurate hybridization. These approaches may therefore be quite sensitive to SNPs (that prevent restriction enzyme cleavage) and PCR artifacts that can contribute to misidentification of altered loci. Despite these potential limitations, oligonucleotide arrays provide increasingly high resolution and are capable of detecting alterations located within intergenic and nongenic regions of the genome, and have therefore become the preferred platform for gene expression and copy number analysis.

Additional methodologies for copy number analysis include bead-based oligonucleotide arrays (e.g. Illumina) and digital karyotyping [34]. The former is very similar to arrays

of chemically immobilized oligonucleotides but uses silica bead-based capture technology with more robust and uniform display of probe oligonucleotides, and therefore yields more accurate and complete measurement of probe hybridization (and furthermore does not require PCR-based representation generation or pre-labeling of genomic DNA). The latter approach, which utilizes combinations of restriction enzymes and PCR to generate 21-bp tags from genomic DNA that are concatamerized and sequenced to determine location and tag density and then computationally mapped, has been used to isolate candidate cancer genes, but may be limited in scope by the requirement for labor-intensive deep sequencing and decreased mapping resolution [35]. Finally, the feasibility of using single-molecule sequencing for identifying focal gains and losses based on density of reads has recently been demonstrated [36], although, like digital karyotyping, given cost and labor, is currently limited to analysis of only a few samples in any given study. A summary of the major array copy number analysis platforms is provided in Fig. (2).

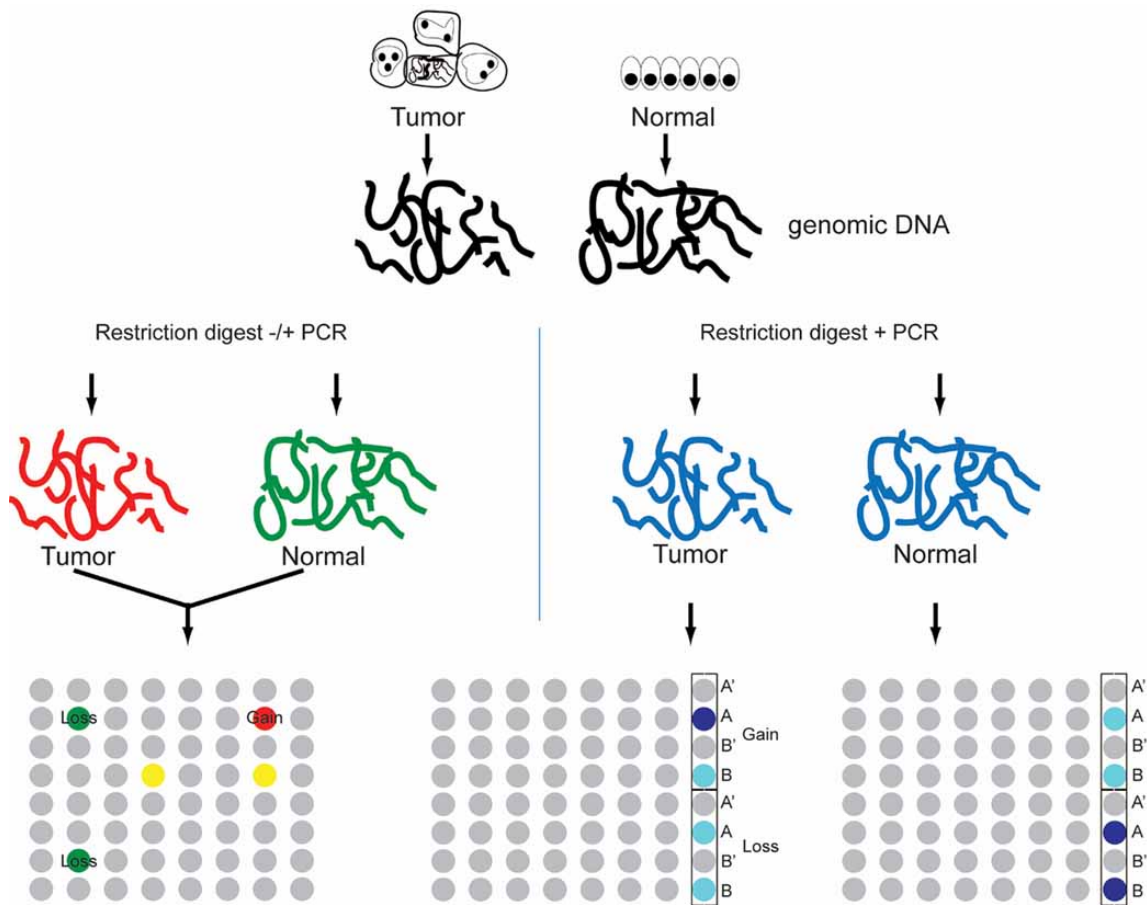


Fig. (2). Comparison of major copy number analysis platforms. With traditional array CGH (left), genomic DNA isolated from tumor and normal samples is labeled with unique fluorophores (e.g. Cy5 (red) for tumor and Cy3 (green) for normal) and then cohybridized to a single array consisting of cDNAs or 60-70mer oligonucleotides. Genomic losses are identified as an abundance of Cy3 (green) fluorescence, gains as an excess of Cy5 (red) fluorescence. With SNP arrays (right), tumor and normal samples are labeled and hybridized in parallel to different arrays. Each SNP probe consists of four sets of 25mer oligonucleotides, two each for the major (A) and minor (B) alleles, one perfectly matched (A,B) and one mismatched (A',B'). Copy number is determined *in silico* by accounting for the total hybridization signal from each perfectly matched probe, normalized to the background hybridization from the mismatched probes and the average hybridization signal from the normal samples. An abundance of signal (increasing blue) in normal relative to tumor marks genomic losses, in tumor relative to normal, gains.

AUTOMATED DISCOVERY OF AMPLIFICATIONS AND DELETIONS

An explosion in the generation of array-based copy number data sets has led to an increasing number of computational approaches for deriving information about candidate cancer genes from probe-level data. The following section reviews the major steps in converting probe-level intensity data into regions of copy number alteration that may harbor cancer genes.

Converting Hybridization Intensity to Copy Number

Methods for converting raw probe intensity data to copy number information depend on the microarray platform (e.g. cDNA or oligonucleotide, CGH versus independent hybridization of controls). In general, following quality control and correction for background array noise, the hybridization signal from each probe on the array is normalized to the corresponding signal from a known or inferred diploid genome(s). SNP arrays provide probe-specific estimates of non-specific hybridization by incorporating single nucleotide mismatched probes (for each matched SNP probe) that permit probe-specific estimation of non-specific hybridization (copy number is determined using model-based estimates that account for the signal from both matched and mismatched probes [37]). To permit more robust comparison between separately arrayed samples (that may be subjected to systematic, array-specific differences in hybridization conditions), the hybridization intensity and/or normalized copy number for each probe in each sample is usually “centered” to the median value of all or select (e.g. “invariant”) probes on the particular array (arrays processed in the same batch may also be centered to the array with the median overall intensity among the arrays in the batch).

The choice of reference samples can significantly impact the accuracy of copy number measurements. The use of a single matched (e.g. from the same patient) reference sample to normalize each tumor sample may reduce the frequency of false-positives by controlling for germline copy number variations (see below); however, variable sample quality can make comparisons among different tumor samples more difficult [38]. Therefore, even when matched reference samples are available, a single reference consisting of multiple normal samples is often used. In the case of array CGH, these may be pooled prior to labeling and cohybridization (to each tumor sample). For SNP arrays, the “normals” are arrayed individually and then pooled computationally, most commonly by normalizing each sample probe to the average single intensity of that probe in the reference samples. Some authors have recommended the use of a standardized collection of normal reference samples, the polymorphic features of which have been extensively characterized (e.g., derived from the HapMap Project) [38].

Converting Probe-Level Signals to Genomic Copy Number

Although in theory the copy number at each genetic locus should be defined by integer values (e.g. single copy loss, homozygous deletion or gain), intermediate values are more typically observed in array analyses. Sources of “noisy” copy number measurements include polymorphisms present

in the arrayed DNA, probe sequence-specific variability in hybridization kinetics, incorrectly mapped probe locations and stromal admixture. A variety of computational methods have been developed to “automatically” delineate genomic regions of gains from probe-level data [26,39-49]. Although the mathematics differs among the individual methods, the general approach is: (1) sample by sample and chromosome by chromosome, combine data from neighboring markers with the same underlying copy number; (2) identify statistically significant transitions in copy number for neighboring probes using a reference distribution from normal tissue hybridizations or through permutation testing; and (3) repeat iteratively until significance is maximized for each copy number transition, producing genome “segments” consisting of variable numbers of contiguous probes with statistically similar copy number. Segmentation algorithms are essential for simplifying large datasets consisting of many samples to a manageable number of discrete alterations for subsequent analysis. However, to identify tumor suppressor genes, it is important to consider how each algorithm handles two features of variability in probe-level data, smoothing and small segments.

Smoothing

Smoothing is the process in which each segment is assigned a copy number equal to the average of all of the probes it contains. Smoothing may reduce “noise” in the data by reducing the appearance of underlying probe-to-probe variability in hybridization, thereby increasing sensitivity and specificity [11]. However, smoothing may obscure precise detection of alteration boundaries and very small homozygous deletions comprised of one or a few probes, particularly when a homozygous deletion arises from one large-scale and one focal allelic loss (Fig. 3A).

Small Segments

The problem of individual outlier probes was initially appreciated from the comparison of normal-normal hybridizations using oligonucleotide array-based CGH, in which many stand-alone probes appear as minor losses and gains, manifesting as a “shell” of probes appearing primarily as many single-copy losses and gains [26]. These outlier probes consist of false-positive aberrations—hybridization artifacts, incorrectly mapped probes, SNPs that create or destroy sites for cleavage by the enzyme used to fragment the genomic DNA before hybridization—and potentially true positive, focal losses (and gains). Most algorithms employ methods to filter out potential outliers *via* smoothing, “fusion” to the closest large aberration, or by simply ignoring segments comprised of less than a pre-specified number of probes (usually 5-8) [26,50-53]. However, depending on the density of probes, true copy number losses comprising one or a few probes could in fact span many kilobases, and therefore inactivate the affected gene (Fig. 1).

Scoring an Altered Region as Deleted

Scoring of losses requires a careful consideration of alteration amplitude. Most approaches pre-specify fixed copy number thresholds, either empirically or based on the distribution of probe intensities in reference samples; probes or

segments with copy numbers falling outside of the threshold are considered significantly altered [51,52,54]. Some investigators apply multiple thresholds to distinguish low-level gains from high-level amplification and single copy losses from homozygous deletions [38,50].

Scoring deleted regions with such fixed thresholds is made particularly challenging by the inherent difficulty in accurately measuring the true signal (above background) from low intensity hybridizations. Although a deletion should result in 50% (heterozygous) or total (homozygous) loss of signal from the resident probes, even for homogeneous tumor cell populations (e.g. cell lines), significant probe-to-probe variation in copy number is observed across known genomic deletions, a result of several factors, including noisy measurements at the lower end of the dynamic range, smoothing, tumor heterogeneity, allele-specific copy number changes and negative skew in the distribution of copy number around the diploid peak [38]. Accurate scoring of deletions may be particularly confounded by the presence of contaminating stromal DNA in primary tumor specimens. Although most approaches exclude samples with < ~70% tumor cells, scoring may vary between pathologists and may not account for polyclonal tumor cell populations or extensively admixed stroma. The consequences of stromal admixture have been extensively characterized in “mixing experiments”; the data demonstrate that detection of homozygous

deletions is extremely dependent on tumor purity [23]. Methods to minimize the effect of stromal admixture include tissue micro-dissection or the use of flow cytometry to separate aneuploid from diploid cells in patient samples [26,55].

Germline Copy Number Variants

Germline genomic variation, associated primarily with short (0.1-1kB) insertions and deletions (Indels) and longer (>1kB) gains and losses (copy number variants/CNVs), has been demonstrated in ~12% of the genome, at frequencies as high as 10% of the normal population, and these often involve gene coding sequences (Fig. 3B) [26,55]. A moderate correlation between CNV regions and segmental duplications has led to the suggestion that they may occur in regions of the genome prone to recombination events (and therefore could arise *de novo*). In addition, a recent study suggests that the majority of the copy number differences between any two individuals arise from a limited set of common, inherited polymorphisms [56]. Although germline copy number variations have been demonstrated to contribute to disease susceptibility, the contribution of individual polymorphisms to any particular phenotype is (so far) small [57-59]. Whether germline copy number losses predispose to tumor formation is not yet known, although, as for other diseases, the overall contribution of any individual variant to tumorigenesis may be small.

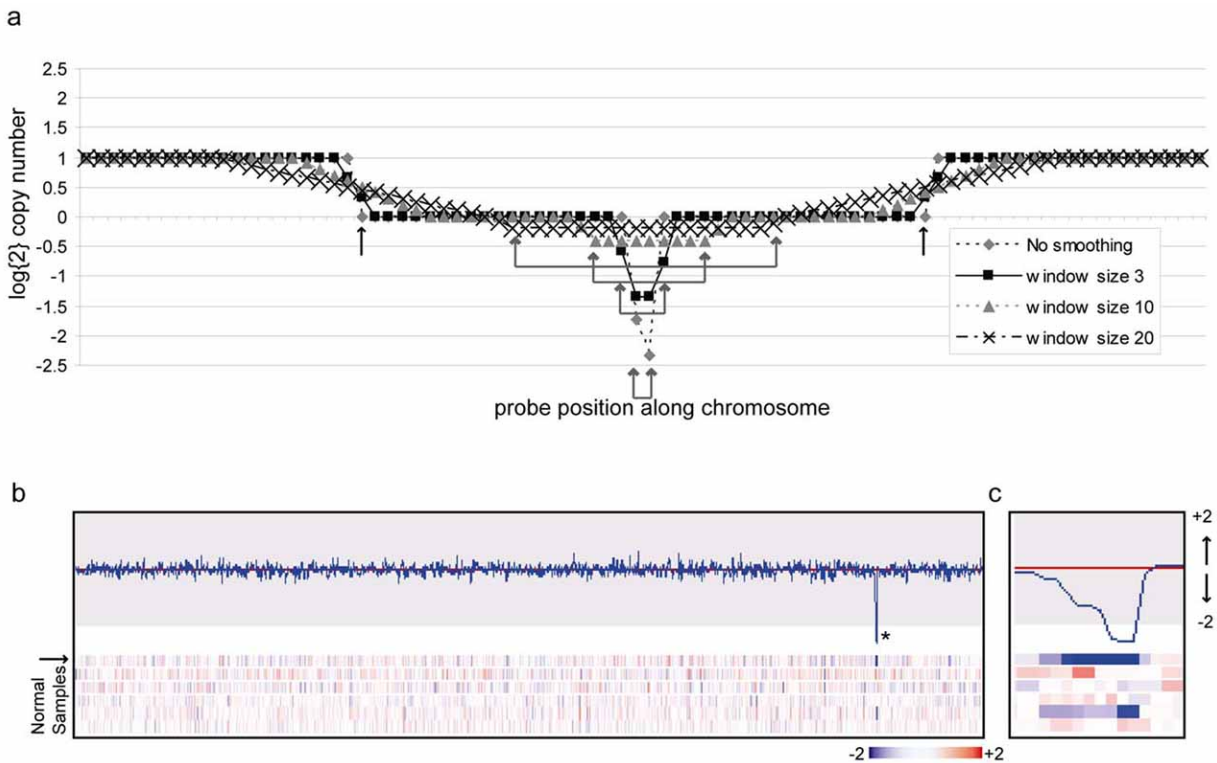


Fig. (3). Detecting focal losses by copy number analysis. (a) The effect of median smoothing of probe copy number is demonstrated for a theoretical locus targeted by a focal homozygous deletion comprised of overlapping large (boundaries indicated with black arrows) and focal (paired gray arrows) allelic losses. As the probe window size for smoothing is increased, the focal deletion becomes obscured by the signal from surrounding probes. (b) Heatmap and graphical display of log₂ copy number for normal tissues. Genomic DNA from six normal samples (peripheral blood mononuclear cells) was hybridized to Affymetrix Human Mapping 250K Sty Arrays and analyzed with dCHIP software (<http://www.hsph.harvard.edu/~cli/complab/dchip/>). Shown is a portion of chromosome 19 demonstrating a focal loss spanning several probes (asterisk) in one sample (indicated by black arrow), consistent with a germline copy number polymorphism. (c) higher resolution view of region indicated by asterisk.

CNVs present a unique challenge to accurately identifying tumor suppressor loci because a germline deletion detected in a tumor by copy number analysis may be incorrectly considered to be of somatic origin. Furthermore, depending on the background rate of generation of genomic losses, a pre-existing heterozygous copy number loss could be converted to a homozygous deletion through somatic loss of the other allele (yet this still may represent a passenger mutation). Homozygous deletions overlapping pre-existing heterozygous CNVs have been observed, although the frequency at which this occurs has not been systematically determined [26].

Normalizing each tumor sample to matched normal tissue can help to establish the germline or somatic origin of each tumor alteration, but is impractical for the reasons discussed above. Therefore, most studies manage the potentially confounding influence of common CNVs by eliminating from consideration any altered probes that map to known CNVs, or by excluding entirely genes overlapping with frequent CNVs, based on the growing database of described genomic variants [55]. Nevertheless, many candidate tumor suppressor genes proposed on the basis of copy number analyses are also marked by frequent germline variation (Table 2); this may complicate determining the true contribution to tumor formation (see discussion of individual candidate genes below).

Table 2. Frequency of Germline Genomic Variants for *Bona fide* and Putative Tumor Suppressor Genes

| Candidate | Cytoband | CNV | INDEL |
|-----------------|----------|-----|-------|
| <i>MSH2</i> | 2p21 | 0 | 1* |
| <i>TGFBR2</i> | 3p24.1 | 0 | 1* |
| <i>CDKN2A/B</i> | 9p21.3 | 1 | 0 |
| <i>PTEN</i> | 10q23.31 | 0 | 0 |
| <i>WT1</i> | 11p13 | 1 | 0 |
| <i>RBI</i> | 13q14.2 | 0 | 2* |
| <i>TP53</i> | 17p13.1 | 0 | 0 |
| <i>NF1</i> | 17q11.2 | 0 | 0 |
| <i>SMAD4</i> | 18q21.1 | 0 | 0 |
| <i>NF2</i> | 22q12.1 | 0 | 0 |
| <i>WTX</i> | Xq11.1 | 0 | 0 |
| <i>FHIT</i> | 3p14.2 | 16 | 4 |
| <i>PARK2</i> | 6q26 | 35 | 10 |
| <i>CSMD1</i> | 8p23.2 | 60 | 31 |
| <i>PTPRD</i> | 9p23 | 30 | 19 |
| <i>WWOX</i> | 16q23.1 | 18 | 3 |

CNV: copy number; INDEL: insertion-deletion.

*intronic.

Source: Database of Genomic Variants [55].

Estimating the Background Rate of Aberration

Early attempts to define altered genomic regions potentially harboring novel cancer genes prioritized regions with

recurrent alteration—those detected in more than one sample [23,40,60]. However, depending on underlying mutation rates and the number of samples analyzed, even recurrent alterations could represent byproducts of tumor genome instability rather than true driver aberrations. Therefore, as with determining the underlying rate of point mutations in tumor genome re-sequencing studies, more recent copy number analyses have attempted to estimate the background rate of mutation as the basis for determining statistical significance. Estimating the rate of point mutations is most often achieved by determining the rate of silent mutations within coding sequences and/or point mutations in non-gene containing regions of the genome (these are presumed to be phenotypically neutral and therefore not acted upon by selection) [61-65]. Furthermore, the wealth of recently generated tumor resequencing data has permitted the influence of nucleotide context on the rates of mutation to be incorporated into models of the background mutation rate. By contrast, the phenotypic consequences of alteration location, size, amplitude and sequence are less well understood. Two recent studies described genome features of human cancer cell lines associated with a higher rate of homozygous deletions (which also varied depending on the tissue of origin) [66,67]. This could suggest that many recurrent homozygous deletions detected in human cancer genomes lie within fragile sites prone to genomic losses rather than marking true cancer genes, thus complicating the identification of *bona fide* tumor suppressor genes on the basis of genome-wide deletion analysis.

Given this uncertainty, the most common strategy is to approximate the background rate (by binomial distribution or permutation testing) with the overall rate of alterations, assuming random distribution across the genome [38,51,53,68,69]. Most approaches are inherently conservative, e.g. they overestimate the background rate because they include alterations affecting *bona fide* cancer genes (which are likely true driving alterations rather than nonspecific passengers). Consequently, this approach could fail to detect true driver aberrations that occur at relatively lower frequency.

Considering the Degree of Copy Number Change

Some approaches incorporate alteration amplitude (how many copies of a gene are gained or lost) into the determination of significance (e.g. together with the frequency of the observed alteration). The general approach is to set very low thresholds for initial scoring of gains and losses, followed by weighting of each alteration by the actual amplitude. Such weighting may be carried out using a continuous scale, or with discrete classes (e.g. single-copy gain, single-copy loss, amplification, or homozygous deletion) [38,50,51]. For amplifications, this may be useful; thus, the dynamic range of gains extends from single copy to more than 30, and, in general, the degree of amplification correlates with overexpression at the protein level (and thus the degree of activation of the targeted gene). However, for deletions, there are only two amplitudes: loss of one or both copies (what is the real meaning of commonly observed noninteger copy number values, and how should they be weighted?) Therefore, depending on the relative contribution of alteration frequency versus amplitude to the overall score, the true importance of

particular genomic losses could potentially be under- or over-estimated.

Finding the Individual Gene(s) Targeted by a Deletion

Although validating candidate genes targeted by very focal deletions is relatively straightforward, prioritizing many candidate genes encompassed by a single large deletion is more challenging. Focusing on the minimal common region (MCR) targeted by overlapping deletions can narrow the number of candidate genes, as can parallel gene expression analysis to identify the subset of genes within a deleted region that are expressed at lower levels [50]. In addition, choosing the subset of probes within an MCR with the greatest (or most significant) amplitude of copy number aberration has been used to further narrow the candidate list [51]. However, this requires that the probes overlying the targeted gene are also the most aberrant, an assumption that has been challenged by examining the characteristics of probes marking *bona fide* tumor suppressor genes [38]. A particularly novel approach to determining whether a large alteration may harbor more than one target is to remove (“peel-off”) the most aberrant (significant) subset of contiguous probes, and to recalculate the significance of the remaining probes and determine whether any other subset of contiguous probes still reaches significance [51,52].

CANDIDATE NOVEL TUMOR SUPPRESSOR GENES IDENTIFIED BY COPY NUMBER ANALYSIS

Having reviewed the major steps involved in narrowing thousands of probe intensities to a manageable number of candidate tumor suppressor genes, we now discuss examples of specific candidate tumor suppressor genes identified on the basis of large-scale array-based copy number analyses. Although the list of candidate genes discussed is by no means comprehensive, our primary goal is illustrate how each candidate was selected and to critically evaluate the data subsequently obtained from deeper genetic and functional validation studies supporting a *bona fide* tumor suppressor function for each candidate.

Novel Glioma Tumor Suppressor Genes: *PARK2* and *PTPRD*

A comprehensive approach to converting probe-level intensity data to candidate cancer regions and genes has been utilized to characterize several cancer genomes, including those from gliomas, lung adenocarcinomas and renal cell carcinomas [51,52,54,70]. The approach, known as Genomic Identification of Significant Targets in Cancer, or GISTIC, illustrates many of the steps of analysis outlined in the first part of this review. First, a series of data preprocessing steps \log_2 -transforms and median centers the probe intensity data (sample by sample), corrects for systematic variation in probe intensities apparent in subsets of arrays (“batch” effects), converts probe intensity to copy number by normalizing each tumor sample only to the subset of (unmatched) normal samples with similar overall probe behavior (“tangent” normalization), and removes duplicate samples and samples contaminated by significant stromal admixture. Next, probe-level copy number data is segmented (with the GLAD segmentation algorithm) and each segment is median

smoothed (e.g. each probe is assigned a new copy number value equal to the median value for all probes contained in the same segment) [42]. Third, probes overlapping polymorphic regions as identified in the Database of Genomic Variants are removed [55]. Fourth, each probe is weighted by the G-score, the product of aberration frequency and average amplitude across all samples, assigned significance by comparing the individual G-score to that obtained with an approximation of the null distribution, and the resulting p-values corrected for multiple comparisons to generate q-values. Probes with q-values <0.25 are considered to be significant. Finally, the candidate gene(s) is determined from the subset of probe(s) with the lowest q-values in each region.

Notable findings from GISTIC analysis with particular relevant to finding new TSGs include:

- (1) The background rate of deletions is higher than the background rate of amplifications. Therefore, deletions must occur at higher frequencies than amplifications to reach statistical significance (based on the G-score statistic described above). The background rate in gliomas is particularly high—every region of the genome is altered in at least one sample.
- (2) Both “broad” (at least a whole chromosome arm) and focal deletions are identified, with broad losses occurring substantially more frequently than focal losses.
- (3) In the glioma studies, the frequency of recurrence of focal alterations decreases with increasing size of the alteration, up until a size equal to ~90% of a chromosome arm, at which point the trend reverses. This could have important implications for the particular model used to estimate the background rate of alteration, since that rate may vary depending on whether the alteration is large or small.
- (4) Several *bona fide* glioma and lung adenocarcinoma tumor suppressor genes were identified, including *CDKN2A/B*, *PTEN*, *RBI* and *CHD5* (glioma) and *CDKN2A/B*, *TP53*, *STK11*, *PTEN* and *RBI* (lung cancer).
- (5) Several deleted regions are not associated with known cancer related genes. Although the majority of these encompass many genes, the peak regions corresponding to several are limited to one or a few candidates.
- (6) The GISTIC method may be remarkably robust for a variety of studies and platforms. In the initial publication, when the data from two previous glioma studies (comprising different samples and array platforms) was reanalyzed with GISTIC, the candidate regions identified by each study were remarkably concordant, even though there was little overlap in the candidate regions when the original algorithms and prioritization schemes were utilized [71,72]. The combined analysis of samples from all three studies permitted adjustment of the peak regions in the previously identified intervals, and allowed eight aberrations (four gains and four losses) that fell above the initial q-value cutoff with the initial analysis to gain significance in the combined data set. Despite this, some regions in the initial study (e.g. encompassing *CDH5*)

do not reach significance in the follow up study and vice versa (e.g. *NF1*, *CDKN2C*).

- (7) Resequencing of selected candidates in additional samples uncovered somatic mutations of *PTPRD* (see below), *CDKN2A*, *PTEN*, *RBI* and *NF1*. Several of these are *bona fide* inactivating mutations (e.g. nonsense, frameshift, splice site), thus providing additional genetic evidence for somatic inactivation in human cancers.

Among the genes nominated as a novel candidate tumor suppressor gene by the glioma copy number analysis is *PARK2*. The gene encodes Parkin, an E3-ubiquitin ligase, mutations in which (including deletions) were shown by classical linkage analysis to account for the most common form of inherited juvenile Parkinson's disease [73]. Loss-of-heterozygosity on the long arm of chromosome 6, centered on 6q25-27, has been frequently described in human tumors, and *PARK2* (6q26) has previously been suggested to be a candidate tumor suppressor gene on the basis of heterozygous (and rare homozygous) deletions in diverse cancers [74-76]. However, the somatic origin of these mutations had not been demonstrated, and tumor re-sequencing initially failed to reveal inactivating point mutations.

The short arm of chromosome 6 was shown in both GISTIC glioma studies to harbor a highly significant recurrent region of loss. Although the number of contiguous probes significantly scoring as lost in these studies spans a large portion of 6q (essentially all of 6q scores as significant), the peak region defined by the set of contiguous probes with the smallest q-value contained only 3 genes: *PARK2*, *PACRG* and *QKI*. Based on these findings, Verriah *et al.* subsequently re-examined individual tumor sample copy number data from the TCGA set of 216 gliomas (as well as 198 colon cancer samples) [77]. These authors discovered loss of at least one allele of *PARK2* in ~25% of each tumor type, with intragenic homozygous deletions within *PARK2* in 2.3% gliomas (5 cases) and 6.1% colon carcinomas (6 cases). In addition, sequencing of 242 tumors identified 12 cases with point mutations (11 heterozygous), including a splice site mutation in a primary glioma and a nonsense mutation in a glioma cell line. Finally, the authors performed extensive functional validation in human cancer cell lines, demonstrating suppression of *in vitro* growth and colony formation and *in vivo* growth of cell line xenografts upon re-expression of wild-type but not several mutant *PARK2* alleles, and only in cells with endogenous loss of *PARK2* (cells expressing wild-type *PARK2* were not affected). Parkin had previously been shown to function as an E3 ubiquitin ligase, and the authors demonstrated wild-type but not cancer-specific mutant Parkin was able to direct the proteasomal-mediated degradation of Cyclin E, and that *PARK2* knockdown led to the an increased fraction of cells in the S and G2-M phases of the cell cycle.

PTPRD, encoding a receptor tyrosine phosphatase, is a second candidate tumor suppressor gene identified by GISTIC analysis. *PTPRD* was previously found to be deleted in both small cell and non-small cell lung cancer cell lines [78,79]. However, its proximity to *CDKN2A/B*, the high rate of homozygous deletions affecting 9p in lung cancer (~30%) and the identification of frequent germline copy number po-

lymorphisms also raised the possibility that *PTPRD* is part of a fragile site prone to recombination [67,78] and is not a true tumor suppressor gene. Initial GISTIC analysis of 371 primary lung adenocarcinomas identified 5 tumors with focal deletions affecting the 5' untranslated region of *PTPRD* [52]. Sequencing of all *PTPRD* coding exons in 188 tumors revealed somatic, heterozygous missense mutations in 11 samples (6%). Notably, sequencing of two other candidate genes encompassed by significant deletions (*AUTS2* and *PDE4D*) for a total of 34 coding exons failed to uncover any mutations in these genes. Although *PTPRD* was not uniquely identified by the GISTIC analysis of gliomas, it was scored as one of the most significant candidate genes in the same data set using a different algorithm (RAE) [38]. Following up on these results, a subsequent re-analysis of the TCGA glioma data set revealed frequent (~40%) hemizygous or homozygous losses, most encompassing nearly all of chromosome 9p. The majority of these (two-thirds) demonstrated loss of either *PTPRD* or *CDKN2A/B*. Although *PTPRD* loss was accompanied by co-deletion of *CDKN2A/B* in all but one case, for one-third of co-deleted samples, a region of normal copy number separates the *PTPRD* and *CDKN2A/B* deletions (suggesting the possibility that each deletion was selected independently). Furthermore, two samples demonstrated intragenic homozygous deletions within the *PTPRD* gene that are predicted to remove coding exons. Additional evidence for somatic inactivation of *PTPRD* was provided both by sequencing analysis, which uncovered eight somatic mutations, including a single nonsense mutation, and by methylation-specific PCR and bisulfite sequencing, which uncovered specific methylation of the *PTPRD* promoter in ~one-third of primary gliomas analyzed (as well as a significant fraction of breast and colon cancers). Finally, introduction of wild-type *PTPRD* (but not cancer-specific mutants) inhibited, and shRNA knockdown promoted, cell growth *in vitro* and in xenografts, possibly *via* effects on the phosphorylation-dependent activation of STAT3.

For both *PARK2* and *PTPRD*, the discovery of focal intragenic microdeletions, somatic point mutations (and for *PTPRD*, promoter methylation) and tumor suppressor activity in functional assays provide compelling evidence for their specific inactivation during gliomagenesis. However, two features of both of these genes complicate their classification as TSGs with a highly significant role in gliomagenesis. First, more than 100 germline genomic variants overlapping *PARK2* and *PTPRD* have been identified, including 46 copy number polymorphisms (e.g. more than 1kB in size) in *PARK2* and 40 in *PTPRD* (the remaining consist of small insertions and deletions from 0.1 to 1kB in size) [55]. For *PTPRD*, three losses overlap coding exons, one with a frequency of 3% (in the population examined). For *PARK2*, at least four losses eliminate coding exons, one at a frequency of ~5%. Although the initial genomic studies took steps to minimize the impact of germline polymorphisms on the selection of candidate regions (for the lung adenocarcinoma study in particular, matched normal samples were used to determine copy number for the majority of tumor samples), it is surprising that inactivating germline deletions occurring at a significant frequency in the general population are not associated with a highly penetrant predisposition to cancer (for example, compare this frequency to the much lower

overall population frequency of germline inactivating mutations in *bona fide* tumor suppressor genes responsible for rare familial cancer predisposition syndromes). In addition, very few germline polymorphisms have been catalogued for established TSGs (Table 2). This may be particularly relevant for *PARK2*, since most deletions are hemizygous, and the authors hypothesize that haploinsufficiency of *PARK2* is sufficient to promote tumorigenesis. Second, although there is a precedent for heterozygous missense mutations leading to inactivation of protein function (e.g. the *TP53* gene), generally this requires dominant-negative activity for the mutant protein or epigenetic loss of expression of the wild-type allele. In both studies, overexpressed cancer-specific mutant Parkin or PTPRD proteins lacked the growth inhibitory effects of the wild-type protein in cells with loss of the endogenous gene; however, there was no demonstration of potential dominant-negative activity by testing the effect of the mutants on cells wild-type for each protein. Furthermore, in tumors or cells with hemizygous mutations (deletions or point mutations), expression from the wild-type allele was not examined.

The First X Chromosome Tumor Suppressor Genes: *WTX* and *UTX*

The discovery of *WTX* illustrates the power of genome-wide copy number analysis when focused on a limited set of highly related tumors with a low background rate of genomic alterations [19]. Wilms tumor is the most common pediatric kidney cancer. Germline heterozygous genetic inactivation of the zinc finger transcription factor *WT1* accounts for two forms of familial Wilms tumor (the second allele is inactivated in the tumors), and somatic biallelic inactivation has been described in 5-10% of sporadic Wilms tumors [12,15,19]. In addition, activating mutations in the *CTNBB1* gene and epigenetic dysregulation of the IGF axis have been described in rare cases [80]. However, in the majority of cases, no specific genetic abnormalities had been identified.

To search for additional Wilms tumor genes, investigators performed oligonucleotide array-based CGH on a collection of 51 primary Wilms tumor specimens. In marked contrast to most adult cancers, the array CGH profile in Wilms tumor was remarkably stable, with very few gains, single copy losses or homozygous deletions. As a result, focal homozygous deletions at Xq11.1 spanning just 1-3 probes were able to be clearly identified in 5/26 tumors, with the minimal common area of deletion implicating a previously uncharacterized gene, which was renamed *WTX*. Sequencing of the coding region in additional tumors uncovered six truncating mutations; no mutations were present in matched normal tissue or 269 unmatched normal DNA samples. Notably, in Wilms tumors from female patients, inactivating mutations exclusively target the active X chromosome through a single hit mechanism, with expression of the intact allele eliminated by X inactivation. Furthermore, expression of *WTX* is restricted to the renal stem cell population that is the presumed cell of origin of Wilms tumors. Although cell lines with endogenous loss of *WTX* are not available, the authors showed that overexpression in immortalized kidney cells suppressed colony formation through induction of apoptosis.

By comparison to the relatively focused analysis of a limited number of tumors of a single type, analysis of 1,390

cancer cell lines and primary tumors by a combination of array-based copy number analyses and candidate gene re-sequencing uncovered 39 inactivating mutations in the histone H3 lysine 27 demethylase gene *UTX*, also located on the X chromosome [81]. The mutations comprised 16 homozygous deletions (13 intragenic microdeletions) and 23 point mutations (nine nonsense, 12 frameshift and two splice site mutations). In contrast to *WTX*, mutations in which so far appear limited to Wilms tumors, *UTX* mutations were discovered in multiple tumor types. Furthermore, *UTX* has been shown to escape X inactivation in females. Instead, a Knudsen two-hit model of inactivation is likely, on the basis of specific loss of the second *UTX* allele in the majority of affected cell lines derived from cancers in females, and concurrent loss of the *UTX* Y chromosome paralogue *UTY* in cell lines derived from cancers in males. Finally, re-expression of wild-type *UTX* significantly prolonged the *in vitro* doubling time of two *UTX* mutant cell lines but not a wild-type cell line. Although no correlation was observed between *UTX* status and global H3K27 trimethylation, re-introduction of *UTX* resulted in decreased H3K27 trimethylation of differentially expressed genes.

The combined genetic and functional data provide very compelling evidence for a *bona fide* tumor suppressor function for both *WTX* and *UTX*. In contrast to *PTPRD* and *PARK2*, highly frequent inactivating frameshifts and nonsense mutations were uncovered; no germline polymorphisms overlapping either *WTX* or *UTX* have previously been described; and no instances of inactivating mutations or deletions were identified in hundreds of normal DNA samples analyzed in each study.

Acute Lymphoblastic Leukemia (ALL) and *PAX5*

As with *WTX*, the discovery of inactivating mutations in *PAX5* and other B-lymphocyte lineage determination genes in pediatric ALL illustrates the power of copy number analysis for uncovering critical tumor suppressor genes, particularly when focused on a single cancer type with a relatively low background rate of copy number alterations [82,83]. In contrast to Wilms tumor, tumor initiating genetic lesions in many pediatric ALLs have been defined, including chromosomal translocations (e.g. *TEL-AML1*, *E2A-PBX1*, *BCR-ABL*, *E2A-PBX1*) and rearrangements involving *MLL* [84]. However, none of these lesions yields a full leukemic phenotype in animal models, suggesting a requirement for additional cooperating alterations. To search for such genomic alterations, Mullighan *et al.* interrogated the purified leukemic blasts from 242 cases of pediatric ALL using high-resolution SNP arrays. Copy number was determined in a novel fashion, by normalizing each probe to the intensity of probes derived from known diploid regions within the same sample's genome (as determined by cytogenetics) prior to segmentation. Segments comprised of less than 3 SNPs were filtered out, and the remaining segments that contained an average log₂ copy number ratio ≥ 0.2 or ≤ -0.2 were scored as gains or losses. Germline copy number variants were eliminated for 228 samples by using the available paired remission sample as the copy number reference, and for the 14 unmatched samples by eliminating probes that appeared polymorphic in the original pool of reference samples.

Notably, the mean number of copy number alterations per tumor is only 6.46, with amplifications outnumbering deletions ~2:1 (this is the opposite of what was observed for gliomas). Among 54 recurrently deleted regions were previously identified *bona fide* tumor suppressor genes in leukemias, including *CDKN2A*, *ATM*, and *RBI*. However, most strikingly were 91 deletions in 7 regulators of normal B-lymphocyte development, limited to B-lymphocyte-derived ALL cases (n=192). Among these were 57 tumors with deletions of *PAX5* (53 single copy, 3 biallelic, 1 internal amplification, 23 intragenic), resulting in loss of expression of the mutated allele or expression of internally deleted or truncated proteins lacking critical functional domains. Importantly, since the majority of *PAX5* mutations were heterozygous, the authors tested the functional effects of a panel of cancer-specific mutant *PAX5* target genes in the presence of the wild-type protein. In all cases, the mutant proteins inhibited transcriptional activation of transfected reporter constructs and known endogenous *PAX5* target genes. Using a similar analysis, the same authors recently demonstrated recurrent intragenic deletions in the B-lineage developmental regulator *IKZF1* in more than 80% of *BCR-ABL* positive ALL cases [83]. The high frequency of genetic inactivation of several genes comprising the same pathway (e.g. B lineage determination), the demonstration of dominant-negative activity, the lack of previously described germline variants and the use of matched normal reference samples to confirm the somatic origin of the observed mutations provide strong evidence for a highly frequent and penetrant tumor suppressor role for these genes.

CONCLUSIONS

In no other field of biology or medicine has the potential value of genome-wide approaches been as thoroughly exploited as human cancer genetics. An unprecedented body of data describing human tumor copy number aberrations has been accumulated in recent years, leading to the identification of many recurrent regions of gains and losses that may harbor critical cancer genes. These studies have been carried out on a variety of array platforms, using a variety of algorithms for identifying genomic regions, different thresholds for scoring gains and losses and various criteria for prioritizing candidate genes and determining statistical significance. As discussed, given the particular challenges associated with detecting genomic losses (compared to gains), accurate detection of single-copy and homozygous deletions may depend on the specific assumptions and methodologies of the individual approaches. Therefore, knowledge of how each method specifically manages the detection of losses can be helpful in choosing individual candidates for further validation from large lists of genes.

When evaluating the evidence supporting a tumor suppressor role for a candidate gene identified through array-based copy number analysis, several critical questions need to be considered. Which array platform was used? What was the density of probes? What was the nature of the reference samples used to determine copy number? If segmentation was used, how were very small segments considered? What types of thresholds were used to select losses? Were germline copy number variants considered? Was the background rate of losses estimated (and how)? Were individual candi-

dates genes validated with deeper analysis and functional assays? Do the mutations (deletions, point mutations) disrupt both alleles? If not, is the wild-type allele expressed?

The sheer volume of cancer genomic data will continue to increase as newer technologies, such as whole cancer genome re-sequencing and genome-wide analysis of epigenetic DNA modifications, are refined to permit their application to large numbers of samples. This overview of array-based copy number analysis hopefully provides a framework for evaluating the large number of candidate tumor suppressor genes likely to emerge from these technologies in the near future.

ACKNOWLEDGEMENTS

The authors wish to thank Miguel N. Rivera for critical reading of the manuscript. S.M. Rothenberg is supported by the MGH Tosteson Postdoctoral Fellowship.

REFERENCES

- [1] Bishop, J.M.; Baker, B.; Fujita, D.; McCombe, P.; Sheiness, D.; Smith, K.; Spector, D.H.; Stehelin, D.; Varmus, H.E. Genesis of a virus-transforming gene. *Natl Cancer Inst. Monogr.*, **1978**, (48), 219-223.
- [2] Druker, B.J.; Talpaz, M.; Resta, D.J.; Peng, B.; Buchdunger, E.; Ford, J.M.; Lydon, N.B.; Kantarjian, H.; Capdeville, R.; Ohno-Jones, S.; Sawyers, C.L. Efficacy and safety of a specific inhibitor of the BCR-ABL tyrosine kinase in chronic myeloid leukemia. *N. Engl. J. Med.*, **2001**, *344*, 1031-1037.
- [3] Lynch, T.J.; Bell, D.W.; Sordella, R.; Gurubhagavatula, S.; Okimoto, R.A.; Brannigan, B.W.; Harris, P.L.; Haserlat, S.M.; Supko, J.G.; Haluska, F.G.; Louis, D.N.; Christiani, D.C.; Settleman, J.; Haber, D.A. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.*, **2004**, *350*, 2129-2139.
- [4] Paez, J.G.; Janne, P.A.; Lee, J.C.; Tracy, S.; Greulich, H.; Gabriel, S.; Herman, P.; Kaye, F.J.; Lindeman, N.; Boggon, T.J.; Naoki, K.; Sasaki, H.; Fujii, Y.; Eck, M.J.; Sellers, W.R.; Johnson, B.E.; Meyerson, M. EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science*, **2004**, *304*, 1497-1500.
- [5] Pao, W.; Miller, V.; Zakowski, M.; Doherty, J.; Politi, K.; Sarkaria, I.; Singh, B.; Heelan, R.; Rusch, V.; Fulton, L.; Mardis, E.; Kupfer, D.; Wilson, R.; Kris, M.; Varmus, H. EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 13306-13311.
- [6] Slamon, D.J.; Leyland-Jones, B.; Shak, S.; Fuchs, H.; Paton, V.; Bajamonde, A.; Fleming, T.; Eiermann, W.; Wolter, J.; Pegram, M.; Baselga, J.; Norton, L. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.*, **2001**, *344*, 783-792.
- [7] Vogel, C.; Cobleigh, M.A.; Tripathy, D.; Gutheil, J.C.; Harris, L.N.; Fehrenbacher, L.; Slamon, D.J.; Murphy, M.; Novotny, W.F.; Burchmore, M.; Shak, S.; Stewart, S.J. First-line, single-agent Herceptin(trastuzumab) in metastatic breast cancer: a preliminary report. *Eur. J. Cancer*, **2001**, *37*(Suppl 1), S25-29.
- [8] Lisitsyn, N.; Wigler, M. Cloning the differences between two complex genomes. *Science*, **1993**, *259*, 946-951.
- [9] Nelson, S.F.; McCusker, J.H.; Sander, M.A.; Kee, Y.; Modrich, P.; Brown, P.O. Genomic mismatch scanning: a new approach to genetic linkage mapping. *Nat. Genet.*, **1993**, *4*, 11-18.
- [10] Pinkel, D.; Seagraves, 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.; Albertson, D.G. High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays. *Nat. Genet.*, **1998**, *20*, 207-211.
- [11] Pollack, J.R.; Perou, C.M.; Alizadeh, A.A.; Eisen, M.B.; Pergamenschikov, A.; Williams, C.F.; Jeffrey, S.S.; Botstein, D.; Brown, P.O. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.*, **1999**, *23*, 41-46.

- [12] Call, K.M.; Glaser, T.; Ito, C.Y.; Buckler, A.J.; Pelletier, J.; Haber, D.A.; Rose, E.A.; Kral, A.; Yeager, H.; Lewis, W.H.; *et al.* Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell*, **1990**, *60*, 509-520.
- [13] Dryja, T.P.; Rapaport, J.M.; Joyce, J.M.; Petersen, R.A. Molecular detection of deletions involving band q14 of chromosome 13 in retinoblastomas. *Proc. Natl. Acad. Sci. USA*, **1986**, *83*, 7391-7394.
- [14] Groden, J.; Thliveris, A.; Samowitz, W.; Carlson, M.; Gelbert, L.; Albertsen, H.; Joslyn, G.; Stevens, J.; Spirio, L.; Robertson, M.; Sargeant, L.; Krapcho, K.; Wolff, E.; Burt, R.; Hughes, J.P.; Warrington, J.; McPherson, J.; Wasmuth, J.; Le Paslier, D.; Abderrahim, H.; Cohen, D.; Leppert, M.; White, R. Identification and characterization of the familial adenomatous polyposis coli gene. *Cell*, **1991**, *66*, 589-600.
- [15] Haber, D.A.; Buckler, A.J.; Glaser, T.; Call, K.M.; Pelletier, J.; Sohn, R.L.; Douglass, E.C.; Housman, D.E. An internal deletion within an 11p13 zinc finger gene contributes to the development of Wilms' tumor. *Cell*, **1990**, *61*, 1257-1269.
- [16] Joslyn, G.; Carlson, M.; Thliveris, A.; Albertsen, H.; Gelbert, L.; Samowitz, W.; Groden, J.; Stevens, J.; Spirio, L.; Robertson, M.; Sargeant, L.; Krapcho, K.; Wolff, E.; Burt, R.; Hughes, J.P.; Warrington, J.; McPherson, J.; Wasmuth, J.; Le Paslier, D.; Abderrahim, H.; Cohen, D.; Leppert, M.; White, R. Identification of deletion mutations and three new genes at the familial polyposis locus. *Cell*, **1991**, *66*, 601-613.
- [17] Kamb, A.; Gruis, N.A.; Weaver-Feldhaus, J.; Liu, Q.; Harshman, K.; Tavtigian, S.V.; Stockert, E.; Day, R.S.; 3rd, Johnson, B.E.; Skolnick, M.H. A cell cycle regulator potentially involved in genesis of many tumor types. *Science*, **1994**, *264*, 436-440.
- [18] Li, J.; Yen, C.; Liaw, D.; Podsypanina, K.; Bose, S.; Wang, S.I.; Puc, J.; Miliareis, C.; Rodgers, L.; McCombie, R.; Bigner, S.H.; Giovannella, B.C.; Ittmann, M.; Tycko, B.; Hibshoosh, H.; Wigler, M.H.; Parsons, R. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science*, **1997**, *275*, 1943-1947.
- [19] Rivera, M.N.; Kim, W.J.; Wells, J.; Driscoll, D.R.; Brannigan, B.W.; Han, M.; Kim, J.C.; Feinberg, A.P.; Gerald, W.L.; Vargas, S.O.; Chin, L.; Iafate, A.J.; Bell, D.W.; Haber, D.A. An X chromosome gene, WTX, is commonly inactivated in Wilms tumor. *Science*, **2007**, *315*, 642-645.
- [20] Steck, P.A.; Pershouse, M.A.; Jasser, S.A.; Yung, W.K.; Lin, H.; Ligon, A.H.; Langford, L.A.; Baumgard, M.L.; Hattier, T.; Davis, T.; Frye, C.; Hu, R.; Swedlund, B.; Teng, D.H.; Tavtigian, S.V. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nat. Genet.*, **1997**, *15*, 356-362.
- [21] Thiagalingam, S.; Lengauer, C.; Leach, F.S.; Schutte, M.; Hahn, S.A.; Overhauser, J.; Willson, J.K.; Markowitz, S.; Hamilton, S.R.; Kern, S.E.; Kinzler, K.W.; Vogelstein, B. Evaluation of candidate tumour suppressor genes on chromosome 18 in colorectal cancers. *Nat. Genet.*, **1996**, *13*, 343-346.
- [22] Weaver-Feldhaus, J.; Gruis, N.A.; Neuhausen, S.; Le Paslier, D.; Stockert, E.; Skolnick, M.H.; Kamb, A. Localization of a putative tumor suppressor gene by using homozygous deletions in melanomas. *Proc. Natl. Acad. Sci. USA*, **1994**, *91*, 7563-7567.
- [23] Zhao, X.; Li, C.; Paez, J.G.; Chin, K.; Janne, P.A.; Chen, T.H.; Girard, L.; Minna, J.; Christiani, D.; Leo, C.; Gray, J.W.; Sellers, W.R.; Meyerson, M. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. *Cancer Res.*, **2004**, *64*, 3060-3071.
- [24] Schutte, M.; da Costa, L.T.; Hahn, S.A.; Moskaluk, C.; Hoque, A.T.; Rozenblum, E.; Weinstein, C.L.; Bittner, M.; Meltzer, P.S.; Trent, J.M.; Yeo, C.J.; Hruban, R.H.; Kern, S.E. Identification by representational difference analysis of a homozygous deletion in pancreatic carcinoma that lies within the BRCA2 region. *Proc. Natl. Acad. Sci. USA*, **1995**, *92*, 5950-5954.
- [25] Yajnik, V.; Paulding, C.; Sordella, R.; McClatchey, A.I.; Saito, M.; Wahrer, D.C.; Reynolds, P.; Bell, D.W.; Lake, R.; van den Heuvel, S.; Settleman, J.; Haber, D.A. DOCK4, a GTPase activator, is disrupted during tumorigenesis. *Cell*, **2003**, *112*, 673-684.
- [26] Lucito, R.; Healy, J.; Alexander, J.; Reiner, A.; Esposito, D.; Chi, M.; Rodgers, L.; Brady, A.; Sebat, J.; Troge, J.; West, J.A.; Rostan, S.; Nguyen, K.C.; Powers, S.; Ye, K.Q.; Olshen, A.; Venkatraman, E.; Norton, L.; Wigler, M. Representational oligonucleotide microarray analysis: a high-resolution method to detect copy number variation. *Genome Res.*, **2003**, *13*, 2291-2305.
- [27] Kallioniemi, A.; Kallioniemi, O.P.; Sudar, D.; Rutovitz, D.; Gray, J.W.; Waldman, F.; Pinkel, D. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science*, **1992**, *258*, 818-821.
- [28] 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.; Ried, T. Multicolor spectral karyotyping of human chromosomes. *Science*, **1996**, *273*, 494-497.
- [29] Lockhart, D.J.; Dong, H.; Byrne, M.C.; Follettie, M.T.; Gallo, M.V.; Chee, M.S.; Mittmann, M.; Wang, C.; Kobayashi, M.; Horton, H.; Brown, E.L. Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.*, **1996**, *14*, 1675-1680.
- [30] Schena, M.; Shalon, D.; Davis, R.W.; Brown, P.O. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science*, **1995**, *270*, 467-470.
- [31] Geschwind, D.H.; Gregg, J.; Boone, K.; Karrim, J.; Pawlikowska-Haddal, A.; Rao, E.; Ellison, J.; Ciccodicola, A.; D'Urso, M.; Woods, R.; Rappold, G.A.; Swerdloff, R.; Nelson, S.F. Klinefelter's syndrome as a model of anomalous cerebral laterality: testing gene dosage in the X chromosome pseudoautosomal region using a DNA microarray. *Dev. Genet.*, **1998**, *23*, 215-229.
- [32] Solinas-Toldo, S.; Lampel, S.; Stilgenbauer, S.; Nickolenko, J.; Benner, A.; Dohner, H.; Cremer, T.; Lichter, P. Matrix-based comparative genomic hybridization: biochips to screen for genomic imbalances. *Genes Chromosomes Cancer*, **1997**, *20*, 399-407.
- [33] Lindblad-Toh, K.; Tanenbaum, D.M.; Daly, M.J.; Winchester, E.; Lui, W.O.; Villapakkam, A.; Stanton, S.E.; Larsson, C.; Hudson, T.J.; Johnson, B.E.; Lander, E.S.; Meyerson, M. Loss-of-heterozygosity analysis of small-cell lung carcinomas using single-nucleotide polymorphism arrays. *Nat. Biotechnol.*, **2000**, *18*, 1001-1005.
- [34] Wang, T.L.; Maierhofer, C.; Speicher, M.R.; Lengauer, C.; Vogelstein, B.; Kinzler, K.W.; Velculescu, V.E. Digital karyotyping. *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 16156-16161.
- [35] Nakayama, K.; Nakayama, N.; Davidson, B.; Katabuchi, H.; Kurman, R.J.; Velculescu, V.E.; Shih Ie, M.; Wang, T.L. Homozygous deletion of MKK4 in ovarian serous carcinoma. *Cancer Biol. Ther.*, **2006**, *5*, 630-634.
- [36] Chiang, D.Y.; Getz, G.; Jaffe, D.B.; O'Kelly, M.J.; Zhao, X.; Carter, S.L.; Russ, C.; Nusbaum, C.; Meyerson, M.; Lander, E.S. High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat. Methods*, **2009**, *6*, 99-103.
- [37] Li, C.; Wong, W.H. Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 31-36.
- [38] Taylor, B.S.; Barretina, J.; Socci, N.D.; Decarolis, P.; Ladanyi, M.; Meyerson, M.; Singer, S.; Sander, C. Functional copy-number alterations in cancer. *PLoS One*, **2008**, *3*, e3179.
- [39] Autio, R.; Hautaniemi, S.; Kauraniemi, P.; Yli-Harja, O.; Astola, J.; Wolf, M.; Kallioniemi, A. CGH-Plotter: MATLAB toolbox for CGH-data analysis. *Bioinformatics*, **2003**, *19*, 1714-1715.
- [40] Heidenblad, M.; Schoenmakers, E.F.; Jonson, T.; Gorunova, L.; Veltman, J.A.; van Kessel, A.G.; Hoglund, M. Genome-wide array-based comparative genomic hybridization reveals multiple amplification targets and novel homozygous deletions in pancreatic carcinoma cell lines. *Cancer Res.*, **2004**, *64*, 3052-3059.
- [41] Hodgson, G.; Hager, J.H.; Volik, S.; Hariono, S.; Wernick, M.; Moore, D.; Nowak, N.; Albertson, D.G.; Pinkel, D.; Collins, C.; Hanahan, D.; Gray, J.W. Genome scanning with array CGH delineates regional alterations in mouse islet carcinomas. *Nat. Genet.*, **2001**, *29*, 459-464.
- [42] Hupe, P.; Stransky, N.; Thiery, J.P.; Radvanyi, F.; Barillot, E. Analysis of array CGH data: from signal ratio to gain and loss of DNA regions. *Bioinformatics*, **2004**, *20*, 3413-3422.
- [43] Linn, S.C.; West, R.B.; Pollack, J.R.; Zhu, S.; Hernandez-Boussard, T.; Nielsen, T.O.; Rubin, B.P.; Patel, R.; Goldblum, J.R.; Siegmund, D.; Botstein, D.; Brown, P.O.; Gilks, C.B.; van de Rijn, M. Gene expression patterns and gene copy number changes in dermatofibrosarcoma protuberans. *Am. J. Pathol.*, **2003**, *163*, 2383-2395.
- [44] Olshen, A.B.; Venkatraman, E.S.; Lucito, R.; Wigler, M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics*, **2004**, *5*, 557-572.
- [45] Pollack, J.R.; Sorlie, T.; Perou, C.M.; Rees, C.A.; Jeffrey, S.S.; Lonning, P.E.; Tibshirani, R.; Botstein, D.; Borresen-Dale, A.L.;

- Brown, P.O. 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*, **2002**, *99*, 12963-12968.
- [46] Price, T.S.; Regan, R.; Mott, R.; Hedman, A.; Honey, B.; Daniels, R.J.; Smith, L.; Greenfield, A.; Tiganescu, A.; Buckle, V.; Ventress, N.; Ayyub, H.; Salhan, A.; Pedraza-Diaz, S.; Broxholme, J.; Ragoussis, J.; Higgs, D.R.; Flint, J.; Knight, S.J. SW-ARRAY: a dynamic programming solution for the identification of copy-number changes in genomic DNA using array comparative genome hybridization data. *Nucleic Acids Res.*, **2005**, *33*, 3455-3464.
- [47] Snijders, A.M.; Fridlyand, J.; Mans, D.A.; Segev, R.; Jain, A.N.; Pinkel, D.; Albertson, D.G. Shaping of tumor and drug-resistant genomes by instability and selection. *Oncogene*, **2003**, *22*, 4370-4379.
- [48] Wang, P.; Kim, Y.; Pollack, J.; Narasimhan, B.; Tibshirani, R. A method for calling gains and losses in array CGH data. *Biostatistics*, **2005**, *6*, 45-58.
- [49] Weiss, M.M.; Snijders, A.M.; Kuipers, E.J.; Ylstra, B.; Pinkel, D.; Meuwissen, S.G.; van Diest, P.J.; Albertson, D.G.; Meijer, G.A. Determination of amplicon boundaries at 20q13.2 in tissue samples of human gastric adenocarcinomas by high-resolution microarray comparative genomic hybridization. *J. Pathol.*, **2003**, *200*, 320-326.
- [50] Aguirre, A.J.; Brennan, C.; Bailey, G.; Sinha, R.; Feng, B.; Leo, C.; Zhang, Y.; Zhang, J.; Gans, J.D.; Bardeesy, N.; Cauwels, C.; Cordon-Cardo, C.; Redston, M.S.; DePinho, R.A.; Chin, L. High-resolution characterization of the pancreatic adenocarcinoma genome. *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 9067-9072.
- [51] Beroukhi, R.; Getz, G.; Nghiemphu, L.; Barretina, J.; Hsueh, T.; Linhart, D.; Vivanco, I.; Lee, J.C.; Huang, J.H.; Alexander, S.; Du, J.; Kau, T.; Thomas, R.K.; Shah, K.; Soto, H.; Perner, S.; Prensner, J.; Debiasi, R.M.; Demichelis, F.; Hatton, C.; Rubin, M.A.; Garraway, L.A.; Nelson, S.F.; Liao, L.; Mischel, P.S.; Cloughesy, T.F.; Meyerson, M.; Golub, T.A.; Lander, E.S.; Mellinger, I.K.; Sellers, W.R. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl. Acad. Sci. USA*, **2007**, *104*, 20007-20012.
- [52] Weir, B.A.; Woo, M.S.; Getz, G.; Perner, S.; Ding, L.; Beroukhi, R.; Lin, W.M.; Province, M.A.; Kravaja, A.; Johnson, L.A.; Shah, K.; Sato, M.; Thomas, R.K.; Barletta, J.A.; Borecki, I.B.; Broderick, S.; Chang, A.C.; Chiang, D.Y.; Chiriac, L.R.; Cho, J.; Fujii, Y.; Gazdar, A.F.; Giordano, T.; Greulich, H.; Hanna, M.; Johnson, B.E.; Kris, M.G.; Lash, A.; Lin, L.; Lindeman, N.; Mardis, E.R.; McPherson, J.D.; Minna, J.D.; Morgan, M.B.; Nadel, M.; Orringer, M.B.; Osborne, J.R.; Ozenberger, B.; Ramos, A.H.; Robinson, J.; Roth, J.A.; Rusch, V.; Sasaki, H.; Shepherd, F.; Sougnez, C.; Spitz, M.R.; Tsao, M.S.; Twomey, D.; Verhaak, R.G.; Weinstock, G.M.; Wheeler, D.A.; Winckler, W.; Yoshizawa, A.; Yu, S.; Zakowski, M.F.; Zhang, Q.; Beer, D.G.; Wistuba, I.I.; Watson, M.A.; Garraway, L.A.; Ladanyi, M.; Travis, W.D.; Pao, W.; Rubin, M.A.; Gabriel, S.B.; Gibbs, R.A.; Varmus, H.E.; Wilson, R.K.; Lander, E.S.; Meyerson, M. Characterizing the cancer genome in lung adenocarcinoma. *Nature*, **2007**, *450*, 893-898.
- [53] Rouveire, C.; Stransky, N.; Hupe, P.; Rosa, P.L.; Viara, E.; Barillot, E.; Radvanyi, F. Computation of recurrent minimal genomic alterations from array-CGH data. *Bioinformatics*, **2006**, *22*, 849-856.
- [54] The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*, **2008**, *455*, 1061-1068.
- [55] Iafraite, A.J.; Feuk, L.; Rivera, M.N.; Listewnik, M.L.; Donahoe, P.K.; Qi, Y.; Scherer, S.W.; Lee, C. Detection of large-scale variation in the human genome. *Nat. Genet.*, **2004**, *36*(9), 949-51.
- [56] McCarroll, S.A.; Kuruville, F.G.; Korn, J.M.; Cawley, S.; Nemesh, J.; Wysoker, A.; Shaper, M.H.; de Bakker, P.I.; Maller, J.B.; Kirby, A.; Elliott, A.L.; Parkin, M.; Hubbell, E.; Webster, T.; Mei, R.; Veitch, J.; Collins, P.J.; Handsaker, R.; Lincoln, S.; Nizzari, M.; Blume, J.; Jones, K.W.; Rava, R.; Daly, M.J.; Gabriel, S.B.; Altshuler, D. Integrated detection and population-genetic analysis of SNPs and copy number variation. *Nat. Genet.*, **2008**, *40*, 1166-1174.
- [57] International Schizophrenia Consortium. Rare chromosomal deletions and duplications increase risk of schizophrenia. *Nature*, **2008**, *455*, 237-241.
- [58] McCarroll, S.A.; Huett, A.; Kuballa, P.; Chileski, S.D.; Landry, A.; Goyette, P.; Zody, M.C.; Hall, J.L.; Brant, S.R.; Cho, J.H.; Duerr, R.H.; Silverberg, M.S.; Taylor, K.D.; Rioux, J.D.; Altshuler, D.; Daly, M.J.; Xavier, R.J. Deletion polymorphism upstream of IRGM associated with altered IRGM expression and Crohn's disease. *Nat. Genet.*, **2008**, *40*(9), 1107-12.
- [59] Stefansson, H.; Rujescu, D.; Cichon, S.; Pietilainen, O.P.; Ingason, A.; Steinberg, S.; Fossdal, R.; Sigurdsson, E.; Sigmundsson, T.; Buizer-Voskamp, J.E.; Hansen, T.; Jakobsen, K.D.; Muglia, P.; Francks, C.; Matthews, P.M.; Gylfason, A.; Halldorsson, B.V.; Gudbjartsson, D.; Thorgeirsson, T.E.; Sigurdsson, A.; Jonasdottir, A.; Bjornsson, A.; Mattiasdottir, S.; Blondal, T.; Haraldsson, M.; Magnusdottir, B.B.; Giegling, I.; Moller, H.J.; Hartmann, A.; Shianna, K.V.; Ge, D.; Need, A.C.; Crombie, C.; Fraser, G.; Walker, N.; Lonnqvist, J.; Suvisaari, J.; Tuulio-Henriksson, A.; Paunio, T.; Touloupoulou, T.; Bramon, E.; Di Forti, M.; Murray, R.; Ruggeri, M.; Vassos, E.; Tosato, S.; Walshe, M.; Li, T.; Vasilescu, C.; Muhleisen, T.W.; Wang, A.G.; Ullum, H.; Djurovic, S.; Melle, I.; Olesen, J.; Kiemene, L.A.; Franke, B.; Sabatti, C.; Freimer, N.B.; Gulcher, J.R.; Thorsteinsdottir, U.; Kong, A.; Andreasson, O.A.; Ophoff, R.A.; Georgi, A.; Rietschel, M.; Werge, T.; Petursson, H.; Goldstein, D.B.; Nothen, M.M.; Peltonen, L.; Collier, D.A.; St Clair, D.; Stefansson, K. Large recurrent microdeletions associated with schizophrenia. *Nature*, **2008**, *455*, 232-236.
- [60] Bashyam, M.D.; Bair, R.; Kim, Y.H.; Wang, P.; Hernandez-Boussard, T.; Karikari, C.A.; Tibshirani, R.; Maitra, A.; Pollack, J.R. Array-based comparative genomic hybridization identifies localized DNA amplifications and homozygous deletions in pancreatic cancer. *Neoplasia*, **2005**, *7*, 556-562.
- [61] Rubin, A.F.; Green, P. Comment on "The consensus coding sequences of human breast and colorectal cancers". *Science*, **2007**, *317*, 1500.
- [62] Greenman, C.; Stephens, P.; Smith, R.; Dalgleish, G.L.; Hunter, C.; Bignell, G.; Davies, H.; Teague, J.; Butler, A.; Stevens, C.; Edkins, S.; O'Meara, S.; Vastrik, I.; Schmidt, E.E.; Avis, T.; Barthorpe, S.; Bhamra, G.; Buck, G.; Choudhury, B.; Clements, J.; Cole, J.; Dicks, E.; Forbes, S.; Gray, K.; Halliday, K.; Harrison, R.; Hills, K.; Hinton, J.; Jenkinson, A.; Jones, D.; Menzies, A.; Mironenko, T.; Perry, J.; Raine, K.; Richardson, D.; Shepherd, R.; Small, A.; Tofts, C.; Varian, J.; Webb, T.; West, S.; Widaa, S.; Yates, A.; Cahill, D.P.; Louis, D.N.; Goldstraw, P.; Nicholson, A.G.; Brasseur, F.; Looijenga, L.; Weber, B.L.; Chiew, Y.E.; DeFazio, A.; Greaves, M.F.; Green, A.R.; Campbell, P.; Birney, E.; Easton, D.F.; Chenevix-Trench, G.; Tan, M.H.; Khoo, S.K.; Teh, B.T.; Yuen, S.T.; Leung, S.Y.; Wooster, R.; Futreal, P.A.; Stratton, M.R. Patterns of somatic mutation in human cancer genomes. *Nature*, **2007**, *446*, 153-158.
- [63] Getz, G.; Hofling, H.; Mesirov, J.P.; Golub, T.R.; Meyerson, M.; Tibshirani, R.; Lander, E.S. Comment on "The consensus coding sequences of human breast and colorectal cancers". *Science*, **2007**, *317*, 1500.
- [64] Forrest, W.F.; Cavet, G. Comment on "The consensus coding sequences of human breast and colorectal cancers". *Science*, **2007**, *317*, 1500; author reply 1500.
- [65] Sjoblom, T.; Jones, S.; Wood, L.D.; Parsons, D.W.; Lin, J.; Barber, T.D.; Mandelker, D.; Leary, R.J.; Ptak, J.; Silliman, N.; Szabo, S.; Buckhaults, P.; Farrell, C.; Meeh, P.; Markowitz, S.D.; Willis, J.; Dawson, D.; Willson, J.K.; Gazdar, A.F.; Hartigan, J.; Wu, L.; Liu, C.; Parmigiani, G.; Park, B.H.; Bachman, K.E.; Papadopoulos, N.; Vogelstein, B.; Kinzler, K.W.; Velculescu, V.E. The consensus coding sequences of human breast and colorectal cancers. *Science*, **2006**, *314*, 268-274.
- [66] Bignell, G.R.; Greenman, C.D.; Davies, H.; Butler, A.P.; Edkins, S.; Andrews, J.M.; Buck, G.; Chen, L.; Beare, D.; Latimer, C.; Widaa, S.; Hinton, J.; Fahey, C.; Fu, B.; Swamy, S.; Dalgleish, G.L.; Teh, B.T.; Deloukas, P.; Yang, F.; Campbell, P.J.; Futreal, P.A.; Stratton, M.R. Signatures of mutation and selection in the cancer genome. *Nature*, **2010**, *463*, 893-898.
- [67] Cox, C.; Bignell, G.; Greenman, C.; Stabenau, A.; Warren, W.; Stephens, P.; Davies, H.; Watt, S.; Teague, J.; Edkins, S.; Birney, E.; Easton, D.F.; Wooster, R.; Futreal, P.A.; Stratton, M.R. A survey of homozygous deletions in human cancer genomes. *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 4542-4547.
- [68] Diskin, S.J.; Eck, T.; Greshock, J.; Mosse, Y.P.; Naylor, T.; Stoeckert, C.J., Jr.; Weber, B.L.; Maris, J.M.; Grant, G.R. STAC: A method for testing the significance of DNA copy number aberrations across multiple array-CGH experiments. *Genome Res.*, **2006**, *16*, 1149-1158.

- [69] Kim, T.M.; Jung, Y.C.; Rhyu, M.G.; Jung, M.H.; Chung, Y.J. GEAR: genomic enrichment analysis of regional DNA copy number changes. *Bioinformatics*, **2008**, *24*, 420-421.
- [70] Beroukhi, R.; Brunet, J.P.; Di Napoli, A.; Mertz, K.D.; Seeley, A.; Pires, M.M.; Linhart, D.; Worrell, R.A.; Moch, H.; Rubin, M.A.; Sellers, W.R.; Meyerson, M.; Linehan, W.M.; Kaelin, W.G.; Jr., Signoretti, S. Patterns of gene expression and copy-number alterations in von-hippel lindau disease-associated and sporadic clear cell carcinoma of the kidney. *Cancer Res.*, **2009**, *69*, 4674-4681.
- [71] Kotliarov, Y.; Steed, M.E.; Christopher, N.; Walling, J.; Su, Q.; Center, A.; Heiss, J.; Rosenblum, M.; Mikkelsen, T.; Zenklusen, J.C.; Fine, H.A. High-resolution global genomic survey of 178 gliomas reveals novel regions of copy number alteration and allelic imbalances. *Cancer Res.*, **2006**, *66*, 9428-9436.
- [72] Maher, E.A.; Brennan, C.; Wen, P.Y.; Durso, L.; Ligon, K.L.; Richardson, A.; Khattry, D.; Feng, B.; Sinha, R.; Louis, D.N.; Quackenbush, J.; Black, P.M.; Chin, L.; DePinho, R.A. Marked genomic differences characterize primary and secondary glioblastoma subtypes and identify two distinct molecular and clinical secondary glioblastoma entities. *Cancer Res.*, **2006**, *66*, 11502-11513.
- [73] Kitada, T.; Asakawa, S.; Hattori, N.; Matsumine, H.; Yamamura, Y.; Minoshima, S.; Yokochi, M.; Mizuno, Y.; Shimizu, N. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature*, **1998**, *392*, 605-608.
- [74] Cesari, R.; Martin, E.S.; Calin, G.A.; Pentimalli, F.; Bichi, R.; McAdams, H.; Trapasso, F.; Drusco, A.; Shimizu, M.; Masciullo, V.; D'Andrilli, G.; Scambia, G.; Picchio, M.C.; Alder, H.; Godwin, A.K.; Croce, C.M. Parkin, a gene implicated in autosomal recessive juvenile parkinsonism, is a candidate tumor suppressor gene on chromosome 6q25-q27. *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 5956-5961.
- [75] Denison, S.R.; Wang, F.; Becker, N.A.; Schule, B.; Kock, N.; Phillips, L.A.; Klein, C.; Smith, D.I. Alterations in the common fragile site gene Parkin in ovarian and other cancers. *Oncogene*, **2003**, *22*, 8370-8378.
- [76] Wang, F.; Denison, S.; Lai, J.P.; Phillips, L.A.; Montoya, D.; Kock, N.; Schule, B.; Klein, C.; Shridhar, V.; Roberts, L.R.; Smith, D.I. Parkin gene alterations in hepatocellular carcinoma. *Genes Chromosomes Cancer*, **2004**, *40*, 85-96.
- [77] Veeriah, S.; Taylor, B.S.; Meng, S.; Fang, F.; Yilmaz, E.; Vivanco, I.; Janakiraman, M.; Schultz, N.; Hanrahan, A.J.; Pao, W.; Ladanyi, M.; Sander, C.; Heguy, A.; Holland, E.C.; Paty, P.B.; Mischel, P.S.; Liau, L.; Cloughesy, T.F.; Mellinghoff, I.K.; Solit, D.B.; Chan, T.A. Somatic mutations of the Parkinson's disease-associated gene PARK2 in glioblastoma and other human malignancies. *Nat. Genet.*, **2010**, *42*, 77-82.
- [78] Sato, M.; Takahashi, K.; Nagayama, K.; Arai, Y.; Ito, N.; Okada, M.; Minna, J.D.; Yokota, J.; Kohno, T. Identification of chromosome arm 9p as the most frequent target of homozygous deletions in lung cancer. *Genes Chromosomes Cancer*, **2005**, *44*, 405-414.
- [79] Zhao, X.; Weir, B.A.; LaFramboise, T.; Lin, M.; Beroukhi, R.; Garraway, L.; Beheshti, J.; Lee, J.C.; Naoki, K.; Richards, W.G.; Sugarbaker, D.; Chen, F.; Rubin, M.A.; Janne, P.A.; Girard, L.; Minna, J.; Christiani, D.; Li, C.; Sellers, W.R.; Meyerson, M. Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. *Cancer Res.*, **2005**, *65*, 5561-5570.
- [80] Rivera, M.N.; Haber, D.A. Wilms' tumour: connecting tumorigenesis and organ development in the kidney. *Nat. Rev. Cancer*, **2005**, *5*, 699-712.
- [81] van Haaften, G.; Dalgliesh, G.L.; Davies, H.; Chen, L.; Bignell, G.; Greenman, C.; Edkins, S.; Hardy, C.; O'Meara, S.; Teague, J.; Butler, A.; Hinton, J.; Latimer, C.; Andrews, J.; Barthorpe, S.; Beare, D.; Buck, G.; Campbell, P.J.; Cole, J.; Forbes, S.; Jia, M.; Jones, D.; Kok, C.Y.; Leroy, C.; Lin, M.L.; McBride, D.J.; Maddison, M.; Maquire, S.; McLay, K.; Menzies, A.; Mironenko, T.; Mulderrig, L.; Mudie, L.; Pleasance, E.; Shepherd, R.; Smith, R.; Stebbings, L.; Stephens, P.; Tang, G.; Tarpey, P.S.; Turner, R.; Turrell, K.; Varian, J.; West, S.; Widaa, S.; Wray, P.; Collins, V.P.; Ichimura, K.; Law, S.; Wong, J.; Yuen, S.T.; Leung, S.Y.; Tonon, G.; DePinho, R.A.; Tai, Y.T.; Anderson, K.C.; Kahnoski, R.J.; Massie, A.; Khoo, S.K.; Teh, B.T.; Stratton, M.R.; Futreal, P.A. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat. Genet.*, **2009**, *41*, 521-523.
- [82] Mullighan, C.G.; Goorha, S.; Radtke, I.; Miller, C.B.; Coustan-Smith, E.; Dalton, J.D.; Girtman, K.; Mathew, S.; Ma, J.; Pounds, S.B.; Su, X.; Pui, C.H.; Relling, M.V.; Evans, W.E.; Shurtleff, S.A.; Downing, J.R. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*, **2007**, *446*, 758-764.
- [83] Mullighan, C.G.; Miller, C.B.; Radtke, I.; Phillips, L.A.; Dalton, J.; Ma, J.; White, D.; Hughes, T.P.; Le Beau, M.M.; Pui, C.H.; Relling, M.V.; Shurtleff, S.A.; Downing, J.R. BCR-ABL1 lymphoblastic leukaemia is characterized by the deletion of Ikaros. *Nature*, **2008**, *453*, 110-114.
- [84] Cheok, M.H.; Evans, W.E. Acute lymphoblastic leukaemia: a model for the pharmacogenomics of cancer therapy. *Nat. Rev. Cancer*, **2006**, *6*, 117-129.
- [85] Rothenberg, S.M.; Mohapatra, G.; Rivera, M.N.; Winokur, D.; Greninger, P.; Nitta, M.; Sadow, P.M.; Sooriyakumar, G.; Brannigan, B.W.; Ulman, M.J.; Perera, R.M.; Wang, R.; Tam, A.; Ma, X.J.; Erlander, M.; Sgroi, D.C.; Rocco, J.W.; Lingen, M.W.; Cohen, E.E.; Louis, D.N.; Settleman, J.; Haber, D.A. A genome-wide screen for microdeletions reveals disruption of polarity complex genes in diverse human cancers. *Cancer Res.*, **2010**, *70*, 2158-2164.