

# Cancer Cell Lines as Genetic Models of Their Parent Histology: Analyses Based on Array Comparative Genomic Hybridization

Joel Greshock,<sup>1</sup> Katherine Nathanson,<sup>2</sup> Anne-Marie Martin,<sup>1</sup> Lin Zhang,<sup>2</sup> George Coukos,<sup>2</sup> Barbara L. Weber,<sup>1</sup> and Tal Z. Zaks<sup>1</sup>

<sup>1</sup>Translational Medicine and Genetics, GlaxoSmithKline and <sup>2</sup>Abramson Cancer Center, University of Pennsylvania, Philadelphia, Pennsylvania

## Abstract

**Tumor-derived cell lines are used as *in vitro* cancer models, but their ability to accurately reflect the phenotype and genotype of the parental histology remains questionable, given the prevalence of documented cell line-specific cytogenetic changes. We have addressed the issue of whether copy number alterations seen in tumor-derived cell lines reflect those observed in studies of fresh tissue by carrying out a meta-analysis of array-based comparative genomic hybridization data that considers both copy number alteration frequencies and the occurrence of cancer gene amplifications and homozygous deletions. Pairwise correlation comparisons between the data sets of seven diagnosis-specific matched tumor and cell line groups indicate that the trends in aberration frequencies are highly correlated between tumors and cell line sets of matched cancer histology relative to unmatched pairings. Despite their similarities, cell lines showed uniformly higher locus-specific alteration frequencies ( $P = 0.004$ ) and several recurring cell line-specific alterations emerged. These include the previously documented losses of 13q and 9p and gains of 20q, as well as additional undescribed cell line-specific gains of 5p, 7p, and 17q and losses of 18q and 4q. These results indicate that, on average, cell lines preserve *in vitro* the genetic aberrations that are unique to the parent histology from which they were derived while acquiring additional locus-specific alterations. These data may enable a more predictive understanding of individual cell lines as *in vitro* models of cancer biology and therapy. [Cancer Res 2007;67(8):3594–600]**

## Introduction

Cell lines derived from human cancers have been crucial to building our understanding of the molecular pathophysiology of cancer and its treatment. Of equal importance, they form an *in vitro* model system for rational drug discovery and development because they are easy to maintain and manipulate *in vitro* and in animal xenograft models. Given their prominence as models, understanding in what ways and to what degrees cell lines grown under artificial conditions reflect their parental *in vivo* genetic architecture is of fundamental importance to cancer biology.

Since they were first established, cell lines have been used to assay for drug sensitivity. Monolayers are generally poor at

reflecting the *in vivo* sensitivity of the parent histology to classic chemotherapeutics (1), although a better correlation can be found with three-dimensional culture methods (2). In more recent years, novel drugs have been developed against specific oncogenic aberrations that are vital to the transformed phenotype. These include trastuzumab and lapatinib, which both target the *ERBB2* subchromosomal amplification in breast cancer; imatinib, which targets *BCR-ABL* translocation in chronic myelogenous leukemia as well as activating mutations in *KIT*; and gefitinib, which is preferentially active against tumors harboring mutant *EGFR*. In each of these cases, *in vitro* drug sensitivity closely parallels the *in vivo* situation. In other words, this is dependent on the presence or absence of the relevant genetic aberration. Moreover, genetically determined *in vivo* secondary drug resistance arising from additional mutations [e.g., secondary *KIT* mutations associated with acquired imatinib resistance (3) or *EGFR* mutations with acquired gefitinib resistance (4, 5)] can be accurately modeled in cell lines. Thus, in these cases, cell lines that share the relevant genotype of the parent tumor accurately model the drug sensitivity phenotype.

With the advent of genome-wide technologies to assay nucleic acids and proteins, a more complete molecular phenotype of cancer has been emerging. Transcriptional profiling, for example, has led to a new understanding of distinct classes of breast cancer (as well as other histologies) that have significant prognostic and therapeutic implications (reviewed in ref. 6). While there have been some studies showing a transcriptional profiling similarity between cell lines and their histology of origin (most notably using the well-characterized NCI-60 panel; ref. 7), the overall concordance is not particularly high and these cell lines seem to lose the tissue-specific up-regulation of genes (8). This difference is not surprising given that transcriptional regulation of many genes is an immediate function of the cellular environment, which is markedly different *in vitro* than *in vivo*.

The study of subchromosomal copy number alterations, as in the case of *ERBB2* amplification, has revealed mechanisms for tumorigenesis and progression. Array-based comparative genomic hybridization (aCGH) technology in the study of cancers has greatly expanded in recent years, furthered by assays with widespread availability (9, 10) and increasing resolution (11). aCGH-based studies of cancers have primarily used survey style approaches in which tissues recognized as being members of a histologically homogeneous panel are interrogated for recurring copy number alterations. However, a comprehensive knowledge of copy number alterations within and across histologies is far from complete. Although cell lines are commonly used in large-scale aCGH studies, they are normally separated from fresh-frozen tissue panels in subsequent analyses due to their suspected discordance with tumor population (12). This is likely a result of a number of factors. Most notably, the cell line immortalization process has

**Note:** Supplementary data for this article are available at Cancer Research Online (<http://cancerres.aacrjournals.org/>).

**Requests for reprints:** Tal Z. Zaks, GlaxoSmithKline, 1250 South Collegeville Road, UP 4430, Collegeville, PA 19426. Phone: 610-917-5124; Fax: 610-917-4830; E-mail: tal.z.zaks@gsk.com.

©2007 American Association for Cancer Research.  
doi:10.1158/0008-5472.CAN-06-3674

been implicated as a source of cytogenetic changes (13, 14). In addition, multiple growth passages, to which commercially available cell lines are routinely subjected, have been shown to be associated with random genomic instability (15). Finally, past studies have noted differences in gene expression patterns between cell lines and their fresh-frozen tissue counterparts (16–18). Despite these observations, more recent analyses of genetic aberrations from larger panels of cell lines and tumors indicate a close concordance of genetic changes within individual histologies (e.g., breast cancer; ref. 19).

We undertook the current analysis to evaluate the degree to which cell lines display an accurate genome-wide model of the DNA copy number aberrations found in human cancers. Specifically we asked: (a) Do common aberrations occur in the same genomic regions in cell lines as their primary tumor counterparts? (b) Are sporadic aberrations resulting from cell line immortalization and growth passages recurrent and predictable, which would facilitate the ability to use cell line panels to model *in vivo* tumors? To answer these questions, we compiled and compared 19 aCGH data sets from 7 different cancer types with the purpose of estimating the similarity between cell line–tumor groups at the data set level. Specifically, this required focusing on trends in genome-wide gain and loss frequencies between data sets stratified by histology and DNA origin (i.e., tumor or cell line). We report that, when taken as a group, (a) copy number aberrations in cell lines from a given histology reflect their cell of origin and (b) specific genomic regions are prone to more frequent gain or loss in cell lines as compared with those seen *in vivo*. Finally, we begin to identify the patterns of common alterations observed in both tumor and cell line populations that can serve as cancer type delineators.

## Materials and Methods

**Data collection.** Data were obtained either from our groups at the University of Pennsylvania or, where available, from published reports (Table 1). At the University of Pennsylvania, aCGH was conducted on a panel of fresh, frozen, and paraffin-embedded tumors ( $n = 206$ ) and cell lines ( $n = 109$ ) from four cancer types. All samples were assayed twice on an array that was constructed with 4,135 bacterial artificial chromosome (BAC) clones spaced at ~1-Mb intervals across the entire human genome. Clones were mapped to the human genome build 34 (June 2003) using BAC end sequences (69%), a sequence tag site (28%), or a full clone sequence (3%). Array details and hybridization protocols are described in detail in ref. 10. For quality control purposes, low intensity and variable spots were removed from the set before averaging the Cy3/Cy5 ratios for all replicates. aCGH data were obtained from public data resources for fresh tumors of five cancer types ( $n = 445$ ) as well as cancer cell line data for three cancer types ( $n = 112$ ). For analyses, all samples were organized into separate aCGH data sets that were stratified by (a) cancer type, (b) fresh tumor– versus cell line–derived DNA, and (c) data source (University of Pennsylvania generated data or public data). The resulting final data set consisted of 872 distinct samples from 19 separate data sets representing 7 different cancer types (Table 1). Notably, cell lines in common between two separate lung cancer data sets (20, 21) were removed in one set such that each cell line was represented in just one set.

**Data processing.** Data were processed with the objective of accurately describing copy number alteration in each sample while structuring data to allow comparisons between samples from different platforms with distinct probe sets. Several steps were used to accomplish this. First, all samples were normalized under the assumption that the mean copy number of every sample is diploid (mean log<sub>2</sub> copy number ratio = 0). Copy number break points were estimated with a circular binary segmentation algorithm (22) implemented by the *DNACopy* package in the R programming language.

For genome-wide copy number alteration frequency analysis, segmentation output was assigned to 1-Mb bins across the entire genome, where the log<sub>2</sub>-based metric assigned by the segmentation algorithm represented the relative copy number status for each bin in each sample. All raw and processed data are available in MIAME compliant format (23).<sup>3</sup> This procedure attained a common data format for all data sets with no probe dependence and allowed direct comparisons between data that were drawn from different assays with different probes. Categorization of low-level copy number gains ( $\leq 5$  copies) and heterozygous losses for each bin was made by applying thresholds of  $>0.25$  and  $<-0.25$ , respectively. For published studies, it was confirmed that our calculated genome-wide aberration frequencies reflected those measured by the study from which the data were obtained. Similarly, high-level amplifications ( $>5$  copies) and homozygous deletions were identified with threshold scores of 0.81 and  $-1.0$ , respectively. Due to sex mismatching between test and reference in some samples, chromosomes X and Y were excluded from all analyses.

**Data analysis.** Estimations of the fraction of the genome experiencing either DNA copy number gain or loss in each sample were made by compiling the total number of segments classified as gained or lost in each sample based on the threshold scores. Aberration rates across each data set were collated by determining the frequency of change for each 1-Mb bin. The gain and loss frequencies assigned to each bin were used to measure similarities in aberration trends between data sets. Specifically, using Pearson distance as a metric, unsupervised hierarchical clustering was done where distance scores were calculated based on genome-wide aberration trends of each data set. Tumor- and cell line-specific aberrations were calculated by subtracting the mean cell line aberration rate of each bin from that of the tumors of a given histology. As a means of determining the specific alterations driving the similarities between data sets, all bins estimated to have a gain or loss frequency of  $>25\%$  in any data set were mapped to a cytogenetic band and subjected to further clustering analysis. To analyze known cancer-related alterations, a set of 323 cancer genes (24) were mapped and assigned a copy number status based on the previously mentioned log<sub>2</sub> thresholds.

## Results

**Regional trends in gain and loss frequencies seem to be histology specific.** Hierarchical clustering based on 1-Mb bin gain and loss frequencies indicates the tendency for data sets to cluster with matched histologies. Considering the clustering results of all 19 data sets, the lone data sets whose nearest neighbor is not of matched histology are the sarcoma tumor and cell line groups, a single breast tumor data set, and a lung tumor data set (Fig. 1). The tendency for both tumor and cell line data sets of matched histologies to cluster together is highly statistically significant ( $P < 0.0001$ ) considering a permutation-based randomized class assignment ( $n = 1,000$ ). Four independent colon cancer data sets clustered together as did three melanoma data sets; data from these histologic groups were derived from different platforms, negating the possibility of platform-specific clustering. Additionally, two distinct sets of lung cancer cell lines data sets derived from different platforms (BAC-based arrays versus “SNP Chip” arrays) clustered together (data from refs. 20 and 21, respectively).

The occurrence of patterns of frequent alterations seems to drive the similarities between data sets and recapitulate what is seen across the entire genome (Fig. 2; Supplementary Table S1). For example, gains of 19p13.3 are common to both tumors and cell lines from breast, lung, and ovarian cancers, as well as in melanomas, while being virtually absent in colon and pancreatic

<sup>3</sup>Available at: <http://acgh.afcri.upenn.edu/tvcl>.

**Table 1.** A collation of 19 distinct data sets were used to measure trends in copy number aberrations between cancer types

| Site/diagnosis   | Sample type | <i>n</i> | Source              | Probe type | No. probes |
|------------------|-------------|----------|---------------------|------------|------------|
| Breast           | Tumors      | 47       | Penn database       | BAC        | 4,135      |
|                  | Tumors      | 43       | Pollack et al. (36) | cDNA       | 5,686      |
|                  | Cell lines  | 22       | Penn database       | BAC        | 4,135      |
| Ovary            | Tumors      | 76       | Penn database       | BAC        | 4,135      |
|                  | Cell lines  | 27       | Penn database       | BAC        | 4,135      |
| Sarcoma          | Tumors      | 63       | Penn database       | BAC        | 4,135      |
|                  | Cell lines  | 18       | Penn database       | BAC        | 4,135      |
| Melanoma         | Tumors      | 20       | Penn database       | BAC        | 4,135      |
|                  | Tumors      | 126      | Curtin et al. (28)  | BAC        | 2,463      |
|                  | Cell lines  | 42       | Penn database       | BAC        | 4,135      |
| Colon            | Tumors      | 125      | Nakao et al. (26)   | BAC        | 2,463      |
|                  | Tumors      | 37       | Douglas et al. (25) | BAC        | 3,500      |
|                  | Tumors      | 50       | Mehta et al. (46)   | BAC        | 3,500      |
|                  | Cell lines  | 48       | Douglas et al. (25) | BAC        | 3,500      |
| Pancreas         | Tumors      | 13       | Aguirre et al. (12) | cDNA       | 14,120     |
|                  | Cell lines  | 24       | Aguirre et al. (12) | cDNA       | 14,120     |
| Lung             | Tumors      | 51       | Zhao et al. (21)    | SNP        | 115,593    |
|                  | Cell lines  | 20       | Zhao et al. (21)    | SNP        | 115,593    |
|                  | Cell lines  | 20       | Garnis et al. (20)  | BAC        | 32,000     |
| Tumor totals     |             | 651      |                     |            |            |
| Cell line totals |             | 221      |                     |            |            |
| Grand totals     |             | 872      |                     |            |            |

tumors and cell lines. Similarly, gains of 13q21.33 are relatively common in colon tumors and cell lines whereas this aberration is rare in all other histologies. It seems that the most aberrations drive similarity between breast tumors and cell lines, whereas few are seen in colon and pancreatic cancers. Lung and ovarian cancers and melanomas seem to be intermediate in this respect.

**Pairwise analysis of histology-matched tumor and cell line pairs shows uniformly higher rates of locus-specific alteration rates in cell lines ( $P = 0.004$ ; data seen in Supplementary Table S2).** To identify cell line and tumor-specific alterations, the mean gain or loss frequency was calculated for each 1-Mb bin of every histology tumor group and subtracting from that of the matched cell line group. Large values represent 1-Mb regions frequently gained or lost in cell lines but not in tumors or vice versa. For genome-wide copy number gains, breast and lung had the largest difference between cell lines and tumors ( $\mu = 13.7\%$ ;  $\sigma = 14.4\%$  and  $\mu = 9.1\%$ ;  $\sigma = 11.6\%$ , respectively) whereas sarcoma and ovary showed the smallest difference between cell lines and tumors across the entire genome ( $\mu = 0.6\%$ ;  $\sigma = 7.5\%$  and  $\mu = 2.1\%$ ;  $\sigma = 8.5\%$ , respectively). Specific regions that differed between tumors and cell lines were identified in each cancer type by querying for those with a mean difference in frequency that was 2 SD larger than the mean difference for that cancer type. This set of regions reflect those that are the most different between tumor and cell line data sets of matched histology. Several DNA copy number alterations seem to consistently occur at disproportionately higher frequencies in cell lines in at least three cancer groups (Supplementary Table S3). These include gains of large genomic regions such as 20q12-13.33 (~24 Mb) and 17q23.2-24.3 (~11 Mb), as well as more localized gains such as 5q35.1-35.3 (~5 Mb) and 11q13.2-13.4 (~6 Mb). Additionally, more frequent losses of 18q12.2-23 and 9p23-21.3 are seen in cell lines than in tumors (Supplementary Table S3).

Genomic rearrangements, including high level DNA amplifications and homozygous deletions of known disease-related genes, are often defining features in cancer tissues. Consistent with genome-wide data, the frequency of copy number gains of cancer genes ( $n = 323$ ) is more prevalent in cell lines than tumors ( $P < 0.0001$ ,  $t$  test). Similarly, cell lines exhibit more frequent high level amplifications (log 2 ratio,  $>0.81$ ) of cancer-related genes than tumors ( $P < 0.0001$ , paired  $t$  test). Total homozygous losses follow a similar trend by having higher overall rates of occurrence in cell lines ( $P = 0.0139$ , paired  $t$  test). The disproportionate increased amplification and homozygous deletion occurrence of several genes seem to be consistent with divergent frequencies of gain and loss between tumors and cell lines. For example, breast cell lines show a 27% (6 of 22) amplification rate of the *SSI18LI* locus (20q13.33), where only ~1% (1 of 90) of tumors are amplified at this locus (Table 2). This gene falls in a region that shows significantly higher overall gain frequencies in breast, melanoma, colon, and lung cell lines than their respective tumors. Similarly, the frequency of homozygous loss (log 2 ratio,  $<-1.0$ ) of *CDKN2A* (9p21.3) in cell lines seems to be higher in lung and melanoma cell lines than in tumors [7 of 40 (17.5%) cell lines versus 3 of 51 (5.9%) tumors and 5 of 42 (11.9%) cell lines versus 1 of 145 (0.8%) tumors, respectively]. These two histologies are two of four (breast and sarcoma are the others) that showed overall higher loss frequencies of this locus in cell lines. Interestingly, pancreatic samples also had a recurring loss of *CDKN2A* in cell lines whereas there were no occurrences of this aberration in tumors [3 of 24 cell lines (12.5%) versus 0 of 13 tumors (0%)]. Pancreatic tumors and cell lines showed no difference in overall loss frequency at this locus. Several other cancer-related genes also showed differences in amplification frequencies between data groups. Most notably, higher rates of cell line amplification of

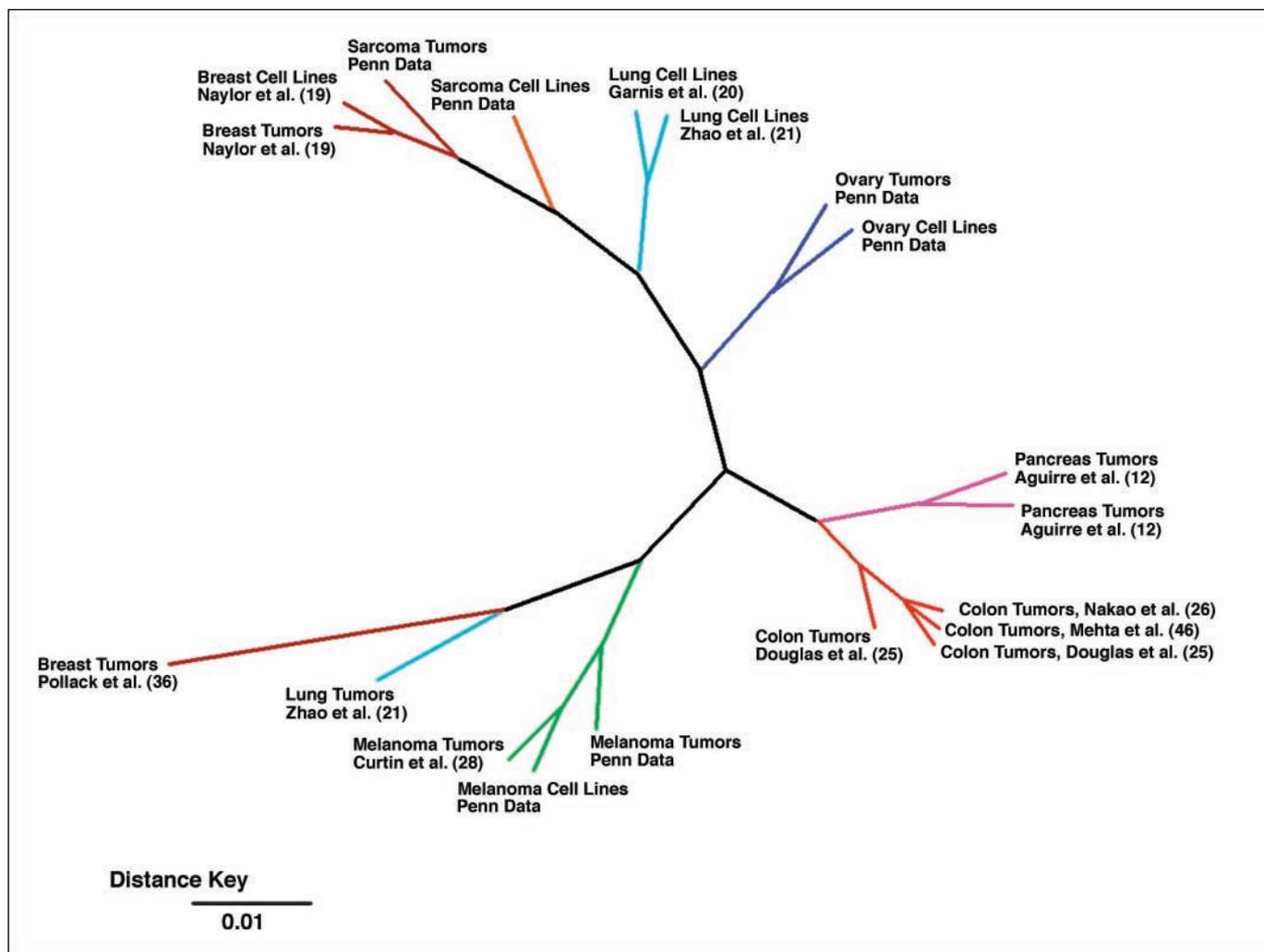
the *MYC* locus (8q24.21) appear in breast, ovary, lung, and colon cancers than in their respective tumor sets. A full list of cancer-related gene DNA copy number gains and losses can be seen in Supplementary Tables S4 and S5, respectively.

## Discussion

In this study, copy number alteration data representing 7 tumor types and 19 different data sets derived from 7 different copy number alteration detection platforms showed strong correlation between regional gain and loss frequencies within cancer types. Although cell lines showed higher overall aberration rates than tumors, these histology-wise pairings showed clear tendencies for similar DNA copy number alterations. Several recurring deviations are apparent. These include previously described cell line-specific gains of 20q12-13.33 and losses of 13q22 (13, 14) as well as newly identified cell line-specific gains of 5q35.1-35.3 and losses of 18q12.2-23. A homozygous loss of the tumor suppressor locus *CDKN2A* and amplification of the *MYC* oncogene seem to be more frequent in cell lines in several histologies, indicating that the dysregulation of these genes may be acquired as part of cell immortalization or their occurrence is selected when tumors are chosen for transformation.

The general concordance of results from independent surveys of recurring copy number aberrations in cancers shows that traditional histology-based groupings are proper first tier stratifiers. For example, two studies of colon cancers yield very similar trends in genome aberration frequencies despite their distinct panel of tissues and array platform (25, 26). Similarly, two separate surveys of non-small-cell lung cancer cell lines provided independent validation of important, previously described cytogenetic changes (20, 21).

Recent studies have used DNA copy number alterations for the molecular pathology of cancers. For example, common copy number gains of 1q as distinct to breast cancers have been described from a pool of tumor types, whereas 13q gains are largely unique to colon cancers (27). Similar results were observed here, as these alterations were components driving tumor/cell line relationships (Fig. 2). Although few alterations were common (>25%) in only a single cancer type, gains of 8q24.21 (encompassing the *MYC* locus), gains of 20q13.31, and losses of 18q21.1 seem to occur frequently in this panel of cancers. Conversely, whereas gains of 14q13.1 are relatively common in lung and pancreatic cancers, they are largely absent in the others. Jointly, these and previous results (27) suggest that each histology broadly bears a range of unique DNA aberrations. Copy number alterations have also been used to discern several cancer



**Figure 1.** Hierarchical clustering of all data sets using Pearson correlation scores and average cluster weighting shows strong relationships between histology-matched data sets.

subtypes. For example, more frequent gains of 11q and 17q have been seen in acral and mucosal melanomas compared with those originating from the skin (28). We would expect that cell line models should similarly represent genotypic subtypes as they do overall histology. For example, it has been shown that colon tumors exhibiting microsatellite instability (MSI+) harbor a distinct set of alterations than microsatellite stable tumors (MSI-; ref. 26). Concordantly, when all colon cancers are stratified by MSI status and tumor and cell line origin, those showing the two MSI+ sample sets segregate from MSI- cancers when subjected to the genome-wide clustering described previously (Supplementary Fig. S1).

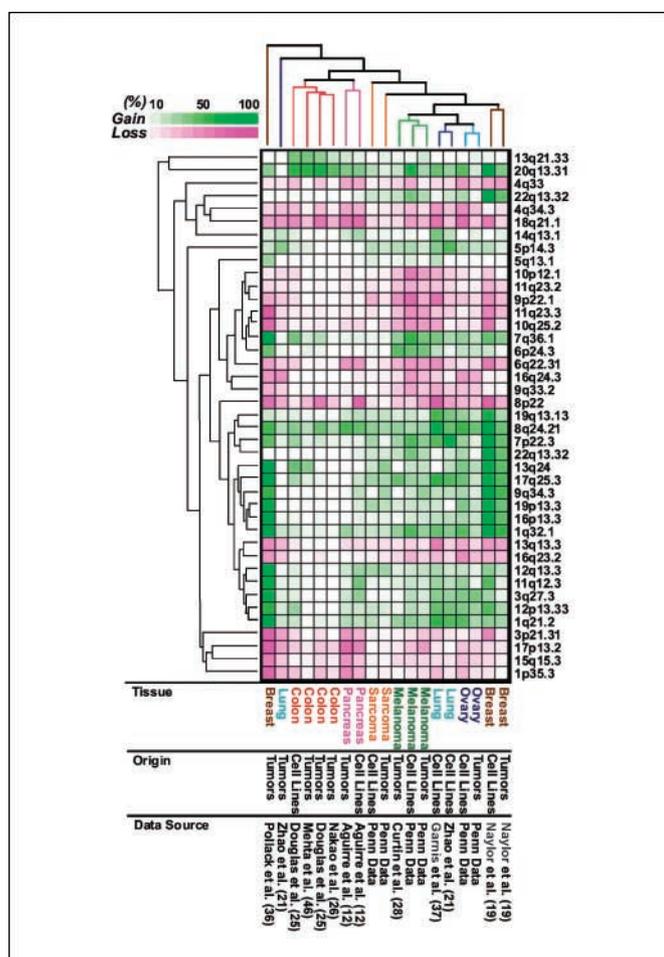
Collectively, these results help quantify cancer cell lines as accurate, reflective models for investigating *in vitro* genomic alterations in human cancers. Cancer cell lines exist as appealing models for studying DNA copy number and, by extension, therapeutic response prediction for several reasons. First, cancer cell lines provide a more homogeneous cell population where cell-to-cell variation in copy number is thought to be reduced. Tumor heterogeneity in primary lesions can limit the ability to accurately describe copy number alterations due to the infusion of normal cells, causing a diluted signal (20), which is consistent with the observation that alterations occur at higher frequencies in cell lines than tumors. Most importantly, tumors cannot be analyzed for copy number alterations *in vivo*, whereas the use of cell lines for aCGH analysis opens the possibility of time course analysis (29) and drug treatments (30, 31). In parental cell lines and their derivatives, drug response can be engineered and studied in relationship to basal and evolved genomic changes (e.g., ref. 32). Our analyses substantiate the translation of these observations to primary tissues by suggesting that relatively large-scale copy number genetic aberrations seen in cell lines *in vitro* accurately reflect their parent histology. Specifically, these results support the notion that cell lines can serve as relevant *in vitro* models for developing specific therapies and imply that, by understanding the exact genetic determinants of a phenotype in a cell line, it is

possible to accurately target similar genotypes in a patient population. In the case of well-known genetic aberrations, using cell lines as *in vitro* models for biomarker discovery and validation is already routinely done (i.e., modeling the effectiveness of kinase inhibitors on cells that have the relevant target amplification or deletion). More generally, the use of cell lines with defined genetic aberrations allows inferences to be made not only on which histologies but also the specific patients that may respond to a given therapy. This is an observation that could be deduced from previous meta-analysis of gene expression data (33) but has not yet been observed as a trend in microarray-based DNA copy number data. In addition, these analyses validate the nature of published CGH study designs, where the calculation of the most common alterations yields biomarkers that are most likely relevant to the oncogenic phenotype (34). By comparing data across a wide spectrum of cell line histologies, loci that are more commonly aberrant in cell lines than in tumors (perhaps as a result of artificial selection pressure) can be identified. Cell lines harboring these recurrent aberrations may be less faithful models of the primary tumor. Further, understanding which recurrent loci are more associated with cell lines will allow them to be accounted for. For example, models aimed at predicting phenotypes that use one or more of these discordant loci may be suspect.

There are several apparent limitations of these analyses. Most significantly, cell line panels may not provide true representation of the range of phenotypes of the parent histology. For example, DNA amplification of the *N-Myc* locus in human neuroblastomas, a biomarker for a malignant phenotype, occurs at significantly higher rates in cell lines than tumors (35). The overrepresentation of such genetic changes may reflect a bias in those tumors selected for or those capable of undergoing transformation. Furthermore, although all data sets in this study are derived from sub-megabase resolution microarrays, platform variation can confound analyses. Probe density (i.e., genome-wide spacing), noise levels, and

**Table 2.** Several cancer-related genes showed higher amplification and deletion frequencies in cell lines than tumors of their respective histology

|                      | Gene           | Chromosome | Mb    | Histology | Cell lines, n (%) | Tumor, n (%)    |
|----------------------|----------------|------------|-------|-----------|-------------------|-----------------|
| Amplifications       | <i>MYC</i>     | 8          | 128.7 | Breast    | 5 of 22 (22.7)    | 4 of 90 (4.4)   |
|                      |                |            |       | Ovary     | 6 of 27 (22.2)    | 7 of 76 (9.2)   |
|                      |                |            |       | Lung      | 6 of 40 (15)      | 0 of 51 (0)     |
|                      |                |            |       | Colon     | 8 of 37 (21.6)    | 11 of 223 (4.9) |
|                      | <i>RECQL4</i>  | 8          | 145.7 | Breast    | 15 of 22 (68.1)   | 10 of 90 (11.1) |
|                      |                |            |       | Ovary     | 8 of 27 (29.6)    | 8 of 76 (10.5)  |
|                      | <i>SS18L1</i>  | 20         | 61.4  | Breast    | 6 of 22 (27.3)    | 1 of 90 (1.1)   |
|                      | <i>CCND1</i>   | 11         | 69.2  | Lung      | 5 of 40 (12.5)    | 0 of 51 (0)     |
|                      | <i>FANCA</i>   | 16         | 89.6  | Breast    | 7 of 22 (31.8)    | 0 of 90 (0)     |
|                      | <i>CBFA2T3</i> | 16         | 88.7  | Breast    | 7 of 22 (31.8)    | 0 of 90 (0)     |
|                      | <i>ARNT</i>    | 1          | 148.0 | Lung      | 7 of 40 (17.5)    | 0 of 51 (0)     |
|                      | <i>AF1Q</i>    | 1          | 148.2 | Lung      | 7 of 40 (17.5)    | 0 of 51 (0)     |
|                      | <i>EXT1</i>    | 8          | 118.8 | Ovary     | 6 of 27 (22.2)    | 5 of 76 (6.6)   |
|                      | <i>ERCC5</i>   | 13         | 101.2 | Colon     | 6 of 37 (16.2)    | 5 of 223 (2.2)  |
|                      | <i>BRCA2</i>   | 13         | 30.7  | Colon     | 8 of 37 (21.6)    | 5 of 223 (2.2)  |
| Homozygous deletions | <i>CDKN2A</i>  | 9          | 22.0  | Lung      | 7 of 40 (17.5)    | 3 of 51 (5.9)   |
|                      |                |            |       | Melanoma  | 5 of 42 (11.9)    | 1 of 145 (0.7)  |
|                      |                |            |       | Pancreas  | 3 of 24 (12.5)    | 0 of 13 (0)     |
|                      | <i>MLL3</i>    | 9          | 20.3  | Melanoma  | 4 of 42 (9.5)     | 1 of 145 (0.7)  |
|                      |                |            |       | Pancreas  | 3 of 24 (12.5)    | 0 of 13 (0)     |



**Figure 2.** Selected genomic regions ( $n = 41$ ) with common copy number alterations show a concordance with genome-wide observations. Regional patterns of gain and loss to these loci drive the similarities between data sets.

mechanical factors (e.g., normalization process) can vary between platforms and laboratories. Each of these variables is capable of affecting the accurate description of copy number aberrations. Finally, CGH analyses reflect an average of genomic aberrations that have occurred in populations of cells. Due to the confounding dilution effect caused noncancerous cell populations, all primary tissues in this study (University of Pennsylvania and published data) were subject to macrodissection under light microscopy to maximize the percent tumor. Resulting tumor cell proportions were variable for these studies, ranging from a minimum of >70% (21) to 50% (36). Although it has been shown that aCGH assays can tolerate up to 50% infusion of normal cells (37), less pure tumors are likely to be prone to increased rates of false negatives. Further, heterogeneous clonal populations may be more common in tumors than in cancer cell lines and can lead to an uninformative profile when assayed by CGH. Ultimately, this possibility can confound the translation from cell line model to tumor genome and provide insight into why aberrations appear in uniformly lower frequencies in tumor data sets. In the future, this may be remedied by techniques that effectively evaluate the genome of single cells (38). Several data sets showed discordance between tumor and cell line populations, including lung tumors and sarcomas. Whereas tumor heterogeneity and

normal cell infusions can account for nonuniform frequencies of alterations, the pattern differences noted in both lung and sarcoma tumors from their cell line models could be the result of the diversity of pathologies within these respective populations. For example, sarcomas represent a host distinct molecular subtypes (reviewed in ref. 39) that are likely to be subject to a unique set of DNA copy number alterations. Tumor and cell line populations were represented disproportionately in several prominent sarcoma subtypes, such as leiomyosarcomas (5% of cell lines, 19% of tumors) or teratomas (0% of cell lines, 15% of tumors). Finally, few studies have been devoted to documenting specific changes to DNA copy number profiles associated with the cell line transformation process. These meta-analyses focus on trends across panels of cancers. Although they do not address directly whether individual cell lines maintain the copy number profile of the parental tumor, they do rediscover findings of previous studies. For example, meta-analysis has suggested disproportionately high rates of loss of 9p21.3, a specific acquired alteration that is described in the immortalization of normal epithelium (40). Similarly, the 20q instability associated with cell lines was also noted in several cancers (14, 41). This suggests that querying large sample sets is also a viable means of discriminating cell line-specific copy number instability.

Although genome-wide RNA profiles of cancer cell lines have often been used in recent years to model drug response and other characteristics *in vitro* (42), it may be advantageous to conduct these investigations with DNA-based profiles or with combined DNA/RNA profiles. First, although careful *in vitro* modeling of oncogenic activation can dissect RNA profiles that correlate with drug response and may correlate with *in vivo* tumor profiles (43), the relationship between cell lines used for *in vitro* studies and their fresh-frozen tissue counterparts is often questionable (16–18). In addition, it is unknown whether successful transcriptional meta-profiles (33) could be independently calculated by exclusively using cell lines. Moreover, the above-mentioned caveats of analyzing potentially heterogeneous cell populations still apply and may be further confounded by the obvious effect of *in vitro* culture conditions on the transcriptome. When a specific genetic aberration is known, its relevance to modeling drug activity is much more obvious [e.g., *BRAF* mutations and MEK inhibition (44), *MET* amplifications and response to MET inhibitors (45), and the above-mentioned studies of *KIT* and *ERB* inhibitors]. Moreover, the measurement of discrete genetic aberrations is easier translated into clinical usefulness.

In summary, we have shown that aCGH-based surveys of tumor cell lines preferentially cluster with their parent histology when considering data set with wide trends in copy number aberrations. This finding supports their usefulness as faithful *in vitro* models of *in vivo* tumors across a wide range of solid tumors. Further, this observation enables an analysis of the discrete aberrations that may distinguish specific tumor types. Future work will be needed to elucidate these loci to help understand the histologic specificity of underlying oncogenic aberrations as well as to determine how they interact with known aberrations (i.e., oncogenic mutations).

## Acknowledgments

Received 10/3/2006; revised 1/31/2007; accepted 2/13/2007.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

## References

1. Weisenthal LM, Kern DH. Prediction of drug resistance in cancer chemotherapy: the Kern and DiSC assays. *Oncology (Huntingt)* 1991;5:93-103; discussion 4, 11-4, 17-8.
2. Vescio RA, Connors KM, Kubota T, Hoffman RM. Correlation of histology and drug response of human tumors grown in native-state three-dimensional histoculture and in nude mice. *Proc Natl Acad Sci U S A* 1991; 88:5163-6.
3. Bertucci F, Goncalves A, Monges G, et al. Acquired resistance to imatinib and secondary KIT exon 13 mutation in gastrointestinal stromal tumour. *Oncol Rep* 2006;16:97-101.
4. Kobayashi S, Boggon TJ, Dayaram T, et al. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2005;352:786-92.
5. Pao W, Miller VA, Politi KA, et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med* 2005;2:e73.
6. Tinker AV, Boussioutas A, Bowtell DD. The challenges of gene expression microarrays for the study of human cancer. *Cancer Cell* 2006;9:333-9.
7. Ross DT, Scherf U, Eisen MB, et al. Systematic variation in gene expression patterns in human cancer cell lines. *Nat Genet* 2000;24:227-35.
8. Sandberg R, Ernberg I. Assessment of tumor characteristic gene expression in cell lines using a tissue similarity index (TSI). *Proc Natl Acad Sci U S A* 2005; 102:2052-7.
9. Bignell GR, Huang J, Greshock J, et al. High-resolution analysis of DNA copy number using oligonucleotide microarrays. *Genome Res* 2004;14:287-95.
10. Greshock J, Naylor TL, Margolin A, et al. 1-Mb resolution array-based comparative genomic hybridization using a BAC clone set optimized for cancer gene analysis. *Genome Res* 2004;14:179-87.
11. Ishkanian AS, Malloff CA, Watson SK, et al. A tiling resolution DNA microarray with complete coverage of the human genome. *Nat Genet* 2004;36:299-303.
12. Aguirre AJ, Brennan C, Bailey G, et al. High-resolution characterization of the pancreatic adenocarcinoma genome. *Proc Natl Acad Sci U S A* 2004;101:9067-72.
13. Jin Y, Zhang H, Tsao SW, et al. Cytogenetic and molecular genetic characterization of immortalized human ovarian surface epithelial cell lines: consistent loss of chromosome 13 and amplification of chromosome 20. *Gynecol Oncol* 2004;92:183-91.
14. Ratsch SB, Gao Q, Srinivasan S, Wazer DE, Band V. Multiple genetic changes are required for efficient immortalization of different subtypes of normal human mammary epithelial cells. *Radiat Res* 2001;155: 143-50.
15. Meisner LF, Wu SQ, Christian BJ, Reznikoff CA. Cytogenetic instability with balanced chromosome changes in an SV40 transformed human uroepithelial cell line. *Cancer Res* 1988;48:3215-20.
16. Kees UR, Rudduck C, Ford J, et al. Two malignant peripheral primitive neuroepithelial tumor cell lines established from consecutive samples of one patient: characterization and cytogenetic analysis. *Genes Chromosomes Cancer* 1992;4:195-204.
17. Lee H, Maihle NJ. Isolation and characterization of four alternate c-erbB3 transcripts expressed in ovarian carcinoma-derived cell lines and normal human tissues. *Oncogene* 1998;16:3243-52.
18. Mackay AR, Ballin M, Pelina MD, et al. Effect of phorbol ester and cytokines on matrix metalloproteinase and tissue inhibitor of metalloproteinase expression in tumor and normal cell lines. *Invasion Metastasis* 1992;12:168-84.
19. Naylor TL, Greshock J, Wang Y, et al. High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. *Breast Cancer Res* 2005;7:R1186-98.
20. Garnis C, Lockwood WW, Vucic E, et al. High resolution analysis of non-small cell lung cancer cell lines by whole genome tiling path array CGH. *Int J Cancer* 2006;118:1556-64.
21. 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* 2005;65:5561-70.
22. Olshen AB, Venkatraman ES, Lucito R, Wigler M. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 2004;5: 557-72.
23. Brazma A, Hingamp P, Quackenbush J, et al. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat Genet* 2001;29:365-71.
24. Futreal PA, Coin L, Marshall M, et al. A census of human cancer genes. *Nat Rev Cancer* 2004;4:177-83.
25. Douglas EJ, Fiegler H, Rowan A, et al. Array comparative genomic hybridization analysis of colorectal cancer cell lines and primary carcinomas. *Cancer Res* 2004;64:4817-25.
26. Nakao K, Mehta KR, Fridlyand J, et al. High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis* 2004;25:1345-57.
27. Jong K, Marchiori E, van der Vaart A, et al. Cross-platform array comparative genomic hybridization meta-analysis separates hematopoietic and mesenchymal from epithelial tumors. *Oncogene* 2007;26: 1499-506.
28. Curtin JA, Fridlyand J, Kageshita T, et al. Distinct sets of genetic alterations in melanoma. *N Engl J Med* 2005;353: 2135-47.
29. Singh B, Kim SH, Carew JF, et al. Genome-wide screening for radiation response factors in head and neck cancer. *Laryngoscope* 2000;110:1251-6.
30. Hattinger CM, Reverter-Branchat G, Remondini D, et al. Genomic imbalances associated with methotrexate resistance in human osteosarcoma cell lines detected by comparative genomic hybridization-based techniques. *Eur J Cell Biol* 2003;82:483-93.
31. O'Toole SA, Sheppard BL, Sheils O, O'Leary JJ, Spengler B, Christoffel V. Analysis of DNA in endometrial cancer cells treated with phyto-estrogenic compounds using comparative genomic hybridisation microarrays. *Planta Med* 2005;71:435-9.
32. de Angelis PM, Fjell B, Kravik KL, et al. Molecular characterizations of derivatives of HCT116 colorectal cancer cells that are resistant to the chemotherapeutic agent 5-fluorouracil. *Int J Oncol* 2004;24:1279-88.
33. Rhodes DR, Yu J, Shanker K, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc Natl Acad Sci U S A* 2004;101:9309-14.
34. Fridlyand J, Snijders AM, Ylstra B, et al. Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* 2006;6:96.
35. Schwab M, Ellison J, Busch M, Rosenau W, Varmus HE, Bishop JM. Enhanced expression of the human gene N-myc consequent to amplification of DNA may contribute to malignant progression of neuroblastoma. *Proc Natl Acad Sci U S A* 1984;81:4940-4.
36. 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 U S A* 2002;99:12963-8.
37. Garnis C, Coe BP, Lam SL, MacAulay C, Lam, WL. High-resolution array CGH increases heterogeneity tolerance in the analysis of clinical samples. *Genomics* 2005;85:790-3.
38. Chin K, de Solorzano CO, Knowles D, et al. *In situ* analyses of genome instability in breast cancer. *Nat Genet* 2004;36:984-8.
39. Nielsen TO. Microarray analysis of sarcomas. *Adv Anat Pathol* 2006;13:166-73.
40. Li HM, Man C, Jin Y, et al. Molecular and cytogenetic changes involved in the immortalization of nasopharyngeal epithelial cells by telomerase. *Int J Cancer* 2006; 119:1567-76.
41. Savelieva E, Belair CD, Newton MA, et al. 20q gain associates with immortalization:20q13.2 amplification correlates with genome instability in human papillomavirus 16 E7 transformed human uroepithelial cells. *Oncogene* 1997;14:551-60.
42. Rickardson L, Fryknas M, Dhar S, et al. Identification of molecular mechanisms for cellular drug resistance by combining drug activity and gene expression profiles. *Br J Cancer* 2005;93:483-92.
43. Bild AH, Yao G, Chang JT, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006;439:353-7.
44. Solit DB, Garraway LA, Pratilas CA, et al. BRAF mutation predicts sensitivity to MEK inhibition. *Nature* 2006;439:358-62.
45. Smolen GA, Sordella R, Muir B, et al. Amplification of MET may identify a subset of cancers with extreme sensitivity to the selective tyrosine kinase inhibitor PHA-665752. *Proc Natl Acad Sci U S A* 2006; 103:2316-21.
46. Mehta KR, Nakao K, Zuraek MB, et al. Fractional genomic alteration detected by array-based comparative genomic hybridization independently predicts survival after hepatic resection for metastatic colorectal cancer. *Clin Cancer Res* 2005;11:1791-7.

# Cancer Research

The Journal of Cancer Research (1916–1930) | The American Journal of Cancer (1931–1940)

## Cancer Cell Lines as Genetic Models of Their Parent Histology: Analyses Based on Array Comparative Genomic Hybridization

Joel Greshock, Katherine Nathanson, Anne-Marie Martin, et al.

*Cancer Res* 2007;67:3594-3600.

|                               |   |
|-------------------------------|---|
| <b>Updated version</b>        | Access the most recent version of this article at:<br><a href="http://cancerres.aacrjournals.org/content/67/8/3594">http://cancerres.aacrjournals.org/content/67/8/3594</a>   |
| <b>Supplementary Material</b> | Access the most recent supplemental material at:<br><a href="http://cancerres.aacrjournals.org/content/suppl/2007/04/11/67.8.3594.DC1">http://cancerres.aacrjournals.org/content/suppl/2007/04/11/67.8.3594.DC1</a> |

|                        |   |
|------------------------|---|
| <b>Cited articles</b>  | This article cites 46 articles, 13 of which you can access for free at:<br><a href="http://cancerres.aacrjournals.org/content/67/8/3594.full#ref-list-1">http://cancerres.aacrjournals.org/content/67/8/3594.full#ref-list-1</a>                |
| <b>Citing articles</b> | This article has been cited by 9 HighWire-hosted articles. Access the articles at:<br><a href="http://cancerres.aacrjournals.org/content/67/8/3594.full#related-urls">http://cancerres.aacrjournals.org/content/67/8/3594.full#related-urls</a> |

|                                   |  |
|-----------------------------------|--|
| <b>E-mail alerts</b>              | <a href="#">Sign up to receive free email-alerts</a> related to this article or journal.   |
| <b>Reprints and Subscriptions</b> | To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at <a href="mailto:pubs@aacr.org">pubs@aacr.org</a> .   |
| <b>Permissions</b>                | To request permission to re-use all or part of this article, use this link<br><a href="http://cancerres.aacrjournals.org/content/67/8/3594">http://cancerres.aacrjournals.org/content/67/8/3594</a> .<br>Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site. |