High-Resolution ROMA CGH and FISH Analysis of Aneuploid and Diploid Breast Tumors

J. HICKS,* L. MUTHUSWAMY,* A. KRAUSNITZ,* N. NAVI,* M. RIGGS,* V. GRUBER,* D. ESPOSITO,* J. ALEXANDER,* J. TROGE,* M. WIGLER,* S. MANER,* P. LUNDIN,* and A. ZETTERBERG†

*Cold Spring Harbor Laboratory, ‡Watson School of Biological Sciences, Cold Spring Harbor, New York 11724; †Karolinska Institutet, Department of Oncology-Pathology, Cancer Center Karolinska, 171 76 Stockholm, Sweden

Combining representational oligonucleotide microarray analysis (ROMA) of tumor DNA with fluorescence in situ hybridization (FISH) of individual tumor cells provides the opportunity to detect and validate a wide range of amplifications, deletions, and rearrangements directly in frozen tumor samples. We have used these combined techniques to examine 101 aneuploid and diploid breast tumors for which long-term follow-up and detailed clinical information were available. We have determined that ROMA provides accurate and sensitive detection of duplications, amplifications, and deletions and yields defined boundaries for these events with a resolution of <50 kbp in most cases. We find that diploid tumors exhibit fewer rearrangements on average than aneuploids, but rearrangements occur at the same locations in both types. Diploid tumors reflect at least three consistent patterns of rearrangement. The reproducibility and frequency of these events, especially in very early stage tumors, provides insight into the earliest chromosomal events in breast cancer. We have also identified correlations between certain sets of rearrangement events and clinically relevant parameters such as long-term survival. These correlations may enable novel prognostic indicators for breast and other cancers as more samples are analyzed.

Altered chromosomes in chromosome organization and structure are a hallmark of many human cancers (Balmain et al. 2003; DePinho and Polyak 2004), reflecting the evolution of the tumor and its ability to proliferate and spread within the host. Breast tumors in particular exhibit a wide range of karyotypic changes including duplication or loss of multiple chromosome arms or entire chromosomes, along with a variety of segmental deletions and amplifications.

The first global studies capable of resolving deletions and amplifications combined comparative genomic hybridization (CGH) and cytogenetics (A. Kallioniemi et al. 1992a,b; O.P. Kallioniemi et al. 1992), and this approach has been applied to breast tumors (Kallioniemi et al. 1994; Reid et al. 1997; Tirkkonen et al. 1998). Subsequently, microarray methods employing CGH have increased resolution and reproducibility, and have improved throughput (Reid et al. 1995; Pollack et al. 2002; Albertson et al. 2003; Lage et al. 2003). These published microarray studies have largely validated the results of cytogenetic CGH, but have had not had sufficient resolution to significantly improve our knowledge of the role of genetic events in the etiology of disease, nor assist in the treatment of the patient. On the other hand, knowledge of specific genetic events, like amplification of c-ErbB2, as studied by fluorescence in situ hybridization (FISH) or quantitative PCR, has been clinically useful (van de Vijver et al. 1997; Slamon et al. 1998; Minard et al. 2001).

Representational oligonucleotide microarray analysis (ROMA) provides an extra measure of resolution in genomic analysis that might be useful in clinical evaluation, as well as delineating loci important in disease evolution.

We have therefore begun a long-term genomic study on a clinically defined set of cancer patients that will combine FISH analysis of specific sites with an ultrahigh resolution microarray CGH technique called ROMA (Lucito et al. 2003) capable of detecting chromosomal events at a resolution approaching 35 kbp. This study is intended to determine whether a detailed knowledge of the events observable in various tumor stage and patient outcomes can elucidate the progression of chromosomal events in breast cancer and provide a means for more accurately directing therapy on the basis of a genomic biopsy.

Both FISH and ROMA can reproducibly detect deletions, duplications, and higher-order amplifications in tissue samples, yet the two techniques have specific differences with valuable and complementary features. Interphase FISH has the advantage of revealing the absolute copy number of a specific genomic sequence or locus complementary to the hybridization probe in each cell examined. Therefore, FISH can distinguish tumor cells with aberrant copy numbers distributed among normal cells in a tumor or biopsy sample. It can likewise detect the presence of subpopulations or subclones of cells within a tumor sample that exhibit different copy numbers for a given probe. The disadvantage of FISH is that the technique depends on some foreknowledge of loci likely to be of interest and examined and is limited to only a few different probes for each experiment, usually fewer than ten. It is therefore highly advantageous to couple FISH with a technology that will survey the entire genome for copy number alterations at the highest possible resolution.

ROMA CGH (Lucito et al. 2003) has the advantage of “seeing” the complete genome in each experiment at a resolution that depends on the number of unique features arrayed on the chip. The microarray chip used in this study has nearly 85,000 features spaced at roughly <50-kbp intervals throughout the genome. Like all microarray-based methods, the copy number that is reported re...
flects an average of all cells in the sample. The presence of normal cells in a tumor sample or biopsy will therefore proportionally depress the signal resulting from a rearrangement associated with tumor cells. In addition, although our FISH results confirm all ROMA signals in nearly all tumor cells, some fraction of tumor cells in a sample may not be identical with respect to amplification or deletion at each locus. It is thus possible that tumor heterogeneity may contribute some loss of signal.

The first phase of this breast cancer survey project is being carried out on frozen tumor tissue collected from 140 breast cancer patients at the Karolinska Institute, Stockholm, Sweden. These tumors represent a wide range of size, clinical stage, and outcome, and all samples carry extensive clinical information. In this paper, we present an outline of our combined ROMA/FISH analysis of a subset of these tumors.

Each of the tumors in this study was initially categorized as aneuploid or diploid based on flow cytometry and was then examined by two-color FISH to determine copy number of 12 critical loci known to be frequently amplified in breast tumors. The amplification profiles obtained by FISH were then compared with profiles obtained by ROMA carried out on DNA isolated from the tumor blocks. ROMA data confirmed all of the events identified by FISH in each sample but, as expected, also revealed many more copy number alterations at additional loci, including deletions as well as amplifications. We then produced hybridization probes for a subset of these loci and carried out FISH on cells from the tumor blocks in order to cross-verify the ROMA results.

These results confirm that ROMA profiles proportionally reflect the copy number of each microarray feature as measured by two-color FISH, and that ROMA can be used to identify the boundaries of deletions, duplications, and amplifications. By compiling data from a large number of samples, we have begun to identify specific types of overall genomic patterns in breast cancer and to relate them to clinical status and eventual patient outcome. The goal of these studies is to identify useful prognostic and therapeutic markers that will eventually help direct therapy in a clinical setting.

MATERIALS AND METHODS

Patient samples. A total of 140 frozen tumor specimens was selected from archives at the Cancer Center of the Karolinska Institute, Stockholm, Sweden. Samples in this particular data set were selected to represent several distinct diagnostic categories in order to populate groups for comparison by FISH and ROMA.

Clinical parameters. Status of the estrogen and progesterone receptors (ER, PR) was determined by ligand binding with a threshold value of >0.05 fg/μg protein for classification as receptor positive.

ROMA DNA microarray analysis. ROMA was performed on a high-density oligonucleotide array containing approximately 85,000 features, manufactured by NimbleGen (Reykjavik, Iceland). Hybridization conditions and statistical analysis have been described previously (Lucito et al. 2003).

Sample preparation, microarray hybridization, and image analysis. The preparation of genome representations, labeling, and hybridization were performed as described previously (Lucito et al. 2003). Briefly, the complexity of the samples was reduced by making BglII genomic representations, consisting of small (200–1200 bp) fragments amplified by adapter-mediated PCR of genomic DNA. For each experiment, two different samples were prepared in parallel. DNA samples (10 μg) were then labeled differentially with Cy5-dCTP or Cy3-dCTP using Amersham-Pharmacia Megaprime Labeling Kit and hybridized in comparison to each other. Each experiment was hybridized in duplicate, where in one replicate, the Cy5 and Cy3 dyes were swapped (i.e., color reversal). Hybridizations consisted of 22 μl of hybridization solution (50% formamide, 5× SSC, and 0.1% SDS) and 10 μl of labeled DNA. Samples were denatured in an MJ Research Tetrad at 95°C for 5 minutes, and then preannealed at 37°C for 30 minutes. This solution was then applied to the microarray and hybridized under a coverslip at 42°C for 14–16 hours. After hybridization, slides were washed 1 minute in 0.2× SDS/0.2× SSC, 30 seconds in 0.2× SSC, and 30 seconds in 0.05× SSC. Slides were dried by centrifugation and scanned immediately. An Axon GenePix 4000B scanner was used, setting the pixel size to 5 μm. GenePix Pro 4.0 software was used for quantitation of intensity for the arrays.

Data processing. Array data were imported into S-PLUS for further analysis. Measured intensities without background subtraction were used to calculate ratios. Data were normalized using an intensity-based lowess curve-fitting algorithm similar to that described in Yang et al. (2002). Log ratio values obtained from color-reversal experiments were averaged and displayed as presented in the figures.

Segmentation algorithm. Segmentation views the probe ratio distribution as an ordered series of probe log ratios, placed in genome order, and breaks it into intervals each with a mean and a standard deviation. At the end of this process, the probe data, in genome order, are divided into segments (long and certain intervals), each segment and feature with its own mean and standard deviation, and each feature associated with a likelihood that the feature is not the result of chance clustering of probes with devout ratios. The ratio data are processed in three phases. In the first phase, we iteratively segment the log ratio data by minimizing variance, then test the segment boundaries, and move them slightly if needed, by setting a very stringent Kolmogorov-Smirnov (K-S) p-value statistic for each segment relative to its neighboring segment (p = 10–5). No segment smaller than six probes in length is considered. In the second phase, we compute the “residual string” of segmented log ratio data, adjusting the mean and standard deviation of each segment so that the residual string has a mean of 0 and a standard deviation of 1.
“Outliers” are defined based on deviance within the popu-
lation, and features are defined as clusters of outliers (at
least two). In the third phase, the features are assigned
likelihood. We determine a “deviance measure” for each
feature that reflects its deviance from the remainder of
the data string. We then, in effect, either randomize or model
randomization of the residual string (i.e., look at the resid-
ual data in a randomized order) many times, and collect
deviance measures of all features generated by purely
random processes. After binning the features by their
length and their deviance measure, we can determine the
likelihood that a given feature with a given length and de-
viance measure would have been generated by random
processes if the probe data were noise.

Fluorescence in situ hybridization. FISH analysis was
performed using interphase cells, and probes were either
prepared from bacterial artificial chromosomes (BACs) or
amplified from specific genomic regions by PCR. Based
on the human genome sequence, primers (1–2 kb in length)
were designed from the repeat-masked sequence of each
copy number polymorphism (CNP) interval, and limited to
an interval no larger than 100 kb. For each probe, a total
of 20–25 different fragments were amplified, then pooled,
and purified by ethanol precipitation. Probe DNA was then
labeled by nick translation with SpectrumOrange™ or
Denaturation of probe and target DNA was performed at
90°C for 5 minutes, followed by hybridization in a humid-
ity chamber at 47°C overnight. The coverglasses were then
removed and the slides were washed in 2× SSC for 10 min-
utes at 72°C, and slides were dehydrated in graded alcohol.
The slides were mounted with anti-fade mounting medium
containing DAPI (4′, 6-diamino-2-phenylindole; Vecta-
tashield) as a counterstain for the nuclei. Evaluation of sig-
nals was carried out in an epifluorescence microscope. Se-
lected cells were photographed in a Zeiss Axiosoplan 2
microscope equipped with Axis Cam MRM CCD camera
and AxiosVision software.

Probe design for FISH. Hybridization probes for FISH
were constructed by one of two methods. For the interdig-
titation analysis, probes were either prepared from BAC’s selected
using the University of California, Santa Cruz, genome browser.
For the determination of copy number in the deletions and amplifications of the aneuploid tumors, probes were made by PCR amplification of primers iden-
tified through the PROBER algorithm designed in this
laboratory. Genomic sequences of 100 kb containing tar-
get amplifications were tiled with 50 probes (800–1400
bp) selected with PROBER Probe Design Software cre-
dated in our laboratory. PROBER uses a distributed anno-
tated sequence retrieval request (Dowell et al. 2001) to re-
quest a genomic sequence and the Mit-Engine (Healy et
al. 2003) to mask the sequence for repeats. Mer lengths of
18 that occur more than twice in the human genome
(UCSF Goldenpath Apr. 10, 2004) with a geometric mean
greater than 2 were masked with (N). Probes were selected
from the remaining unmasked regions according to an al-
gorithm to be published elsewhere.

Oligonucleotide primers were ordered in 96-well plates
from Sigma Genosys and resuspended to 25 μM. Probes
were amplified with the PCR Mastermix kit from Eppen-
dorf (Cat. 10032002 447) from EBV-immortalized cell
line DNA (CpG-Skn-1) DNA (100 ng) with 55°C anneal-
ing, 72°C extension, 2-minute extension time, and 23 cy-
cles. Probes were purified with Qiagen PCR purification
columns (Cat. 28104) and combined into a single probe
ticktail (10–25 ng total probes) for dye labeling and
metaphase/interphase FISH.

Measurement of DNA content. The ploidy of each tu-
mor was determined by measurement of DNA content us-
ing Feulgen photometry (Forselland and Zetterberg
1990; Forselland et al. 1996). The optical densities of the
nuclei in a sample were measured, and a DNA index was
calculated and displayed as a histogram (Kronenwett et
al. 2004). Normal cells and diploid tumors display a ma-
jor peak at 2c DNA content with a smaller peak of G2
phase replicating cells that corresponds to the mitotic in-
dex. Highly aneuploid tumors display broad peaks that
often center on 4c copy number but may include cells from
2c to 6c or above.

Patient consent and institutional review board (IRB)
approvals. KI samples were collected from patients un-
dergoing radical mastectomy at the Karolinska Institute
between 1984 and 1991. Patient consent for research use
was specified under clinical research approvals from the
IRB of the Karolinska Hospital, Stockholm, Sweden.
Work at Cold Spring Harbor Laboratory was carried out
under approval by the CSHL IRB on October 17, 2005 for
a project entitled “Quantitative determination of gene
amplification in breast tumors.”

RESULTS
A subset of 140 frozen tumor specimens was selected from
archives at the Cancer Center of the Karolinska Insti-
tute. Samples in this particular data set were selected to
represent several distinct diagnostic categories in order to
populate groups for comparison by FISH and ROMA.
Most important, these samples are from patients for
whom complete clinical data have been kept and for
whom long-term outcome data (15–18 years) are avail-
able. The clinical characteristics of this sample set are
shown in Table 1.

Each of the tumors in this study was initially categor-
ized as aneuploid or diploid based on flow cytometry
(see Materials and Methods) and then examined by two-
color FISH to determine copy number of several loci
known to be frequently amplified in breast tumors. The
amplification profiles obtained by FISH were then com-
pared with profiles obtained by ROMA carried out on
DNA isolated from the frozen tumor blocks. ROMA was
run by using 85K BglII Version 4 chip design manufac-
tured to our specifications by NimbleGen, Inc. (Reyk-
javik, Iceland) which displays 82,972 separate features,
each consisting of single-stranded DNA, 60 bases in
length, as described previously (Lucito et al. 2003; Sebat
After hybridization and fluorescent scanning, the data consist of ratios calculated by taking the geometric mean of normalized hybridization data from two separate color-reversed chips, each comparing a tumor sample to the laboratory standard male fibroblast cell line. Typical results are shown for sample WZ1 in Figure 1.

Table 1. Distribution of Patients and Clinical Parameters in the Swedish and Norwegian Data Sets

<table>
<thead>
<tr>
<th>Karolinska Inst.</th>
<th>Node</th>
<th>Median age at diagnosis</th>
<th>Grade</th>
<th>Size (mm)</th>
<th>ER*</th>
<th>PR*</th>
<th>ERBB2 +</th>
<th>Tumor</th>
<th>Survival &gt;7 yr</th>
<th>Survival &lt;7 yr</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sweden Total</td>
<td>60</td>
<td>28/31</td>
<td>52</td>
<td>8/11/33</td>
<td>41</td>
<td>43</td>
<td>41/9</td>
<td>39</td>
<td>14/25</td>
<td>21/20</td>
</tr>
<tr>
<td>Diploid</td>
<td>39</td>
<td>14/25</td>
<td>57</td>
<td>3/12/16</td>
<td>19</td>
<td>43</td>
<td>20/13</td>
<td>24</td>
<td>8</td>
<td>24/13</td>
</tr>
<tr>
<td>(Survival &gt;7 yr)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diploid</td>
<td>41</td>
<td>28/13</td>
<td>49</td>
<td>0/2/22</td>
<td>11</td>
<td>20</td>
<td>24/8</td>
<td>41</td>
<td>14/19</td>
<td>25/10</td>
</tr>
</tbody>
</table>

Numbers will not add up exactly because of partial information on certain individual cases. *Progesterone (PR) and estrogen (ER) receptors measured by ligand binding; (pos) >0.5 fg/μg protein. †ERBB2 amplification scored by ROMA as segmented ratio greater than 0.1 above baseline.

Figure 1. Comparison of copy number as assayed by ROMA and FISH. Tumor WZ1 is aneuploid with an average genome copy number of 3n by FACS analysis. The results of FISH probes for various loci are indicated in the top graph. The bottom panels show enlarged views of small deletions and duplications picked to demonstrate the correspondence between FISH and ROMA. The photograp shows a two-color FISH experiment using probes for the deletion and duplication, respectively, depicting loss and gain, respectively, of the two probes relative to the nominal genome copy number. PIK3CA on chromosome 3q yields a value of 1.0 by ROMA and 3 copies by FISH. MDMX on 1q yields a copy number of 5 by FISH, consistent with a near doubling of the copy number of the entire 1q arm as shown by ROMA.
Figure 1 depicts the typical ROMA profile used for all of the breast cancer samples presented in this study, with genomes arranged in chromosome order from left to right. The figure shows the normalized data, known as the “geomean ratio” (Lucito et al. 2003), for each probe, in gray. These “raw” geometric mean ratio data must be further refined in order to reliably identify specific amplifications, duplications, and deletions and to determine their amplitudes and, most importantly, their boundaries. This refinement is achieved through a series of statistical methods that comprise the Bridge 5 segmentation algorithm, described in Sebat et al. (2004) and in Materials and Methods. Segmentation provides a consistent and reliable method for interpretation of data by associating each data feature with a likelihood measure that the feature is not the result of the chance clustering of random noise in probe ratios. The geomean ratio data in Figure 1 are overlaid with the results of the segmentation algorithm in red. The expected ratio differences for the X and Y chromosomes for female versus male DNA are clearly visible. It is clear from the profile of WZ1 in Figure 1 that there are at least two major classes of events: large segmental deletions and duplications of one or two copies of chromosome arms and narrow, high-copy-number amplifications, both of which have been observed previously by other CGH microarray methods (Raed et al. 1997; Pollack et al. 2002; Albertson et al. 2003; Lage et al. 2003). The values predicted by ROMA and the observed values measured by FISH are shown above representative loci. As shown for this one example, ROMA data were consistent with all of the amplifications identified by FISH in each sample but also revealed copy number alterations at additional loci, including deletions as well as amplifications. We then produced hybridization probes for a subset of these loci, and carried out FISH on cells from the tumor blocks in order to confirm the ROMA results.

The three small panels in Figure 1 are an example of the probes made specifically for this tumor using the PROBES software (Materials and Methods) to regions that had undergone less obvious events. The image shows a two-color FISH result for probes made to the two regions of deletion and duplication identified in the flanking panels. The result clearly shows that this tumor, with a genomic equivalent of 3c, has lost at least two copies of the chromosome 2 locus and gained one copy of the chromosome 20 locus. Similar results from 10 different tumors, both of which have been observed previously by other CGH microarray methods (Raed et al. 1997; Pollack et al. 2002; Albertson et al. 2003; Lage et al. 2003). The values predicted by ROMA and the observed values measured by FISH are shown above representative loci. As shown for this one example, ROMA data were consistent with all of the amplifications identified by FISH in each sample but also revealed copy number alterations at additional loci, including deletions as well as amplifications. We then produced hybridization probes for a subset of these loci, and carried out FISH on cells from the tumor blocks in order to confirm the ROMA results.

As in Figure 1, breast cancer profiles provide a rough internal calibration for copy number based on having 2:1 copy number for X and complete lack (equivalent to a homozygous loss) of the Y. One important point to note is that this expectation has limitations because ROMA measures the average copy number of cells in tumors, and some tumor cells have lost one or more of their X chromosomes. Furthermore, the presence of a variable number of normal cells in any tumor cells complicates the estimates of copy number based purely on ROMA. It is clear from inspection that diploids, in general, exhibit fewer events than aneuploids, and with the exception of the certain clustered amplicons described below, the events are most often gains or losses of whole chromosome arms. Aneuploids average 42 events, whereas diploids average 16, and it is only logical to assume that aneuploids, having multiple copies of most chromosomes, have more degrees of freedom to gain or lose copies without deleterious effects on proliferation that might be caused by wholesale gene imbalances, as would be the case in diploids. Yet, on a case-by-case basis, diploid tumors can exhibit the same pathogenic potential for proliferation and for local and distant metastasis as aneuploids. In fact, the locations of the events for diploids and aneuploids are comparable, as shown in Figure 2B, and the number of these events in aneuploids is higher, as expected.

The combination of fewer overall events coupled with the frequent narrow, high-copy-number amplicons makes it particularly advantageous to focus on diploid tumors for CGH analysis in general. In particular, exercises in...
novel oncogene and tumor suppressor discovery may be facilitated by the lower frequency of observable events in diploids. It is likely that diploids may exhibit less background “chatter” from unselected events that might occur randomly in the more permissive aneuploid environment, thus reducing the number of events and loci that must be screened. Likewise, the apparent restriction on gain or loss in diploids leads to the generation of smaller, more discrete events, particularly amplifications that can point directly to oncogenes. The insights gained from the increased resolution of ROMA combined with FISH for both of these aspects of CGH are described below.

Figure 2. Examples of aneuploid and pseudo-diploid tumors. (A) Representative ROMA profiles showing that aneuploid tumors in general exhibit an overall greater frequency of chromosome rearrangements than do pseudo-diploid tumors. (B) Comparative frequency plots of amplification (up) and deletion (down) in various data sets. Frequency calculated on normalized, segmented ROMA profiles using a minimum of six consecutive probes identifying a segment with a minimum mean of 0.1 above (amplification) or below (deletion) baseline. Frequencies are plotted only for chromosomes 1–22. (C) The Swedish diploid subset (blue) is compared to the total Swedish aneuploid subset (red). Comparative frequency plots of Swedish diploid subset >7-year survivors (red) and ≤7-year non-survivors (blue).
Patterns of Diploid Genome Profiles

Visual inspection of segmented diploid profiles suggests those with rearrangements comprise three basic profile types. The first profile pattern (Fig. 3A), which we call “simplex,” has broad segments of duplication and deletion, usually comprising entire chromosomes or chromosome arms, with occasional isolated narrow peaks of amplification. This type represents 60% of the diploid tumors in this sample. In the second type, “sawtooth” (Fig. 3B), the cancer cells have many sub-arm-length segments of amplification and deletions, often alternating, more or less affecting all the chromosomes. Little of the genome remains at normal copy number in this type, which makes up less than 5% of this selected data set. Sawtooth patterns apparently result from a genome-wide loss of mitotic segregation control that eventually becomes clonal.

The third pattern (Fig. 3C) resembles the simplex type except that the cancers contain at least one localized region of clustered peaks of amplification, each cluster confined to a single chromosome arm, which we call “firestorms.” In contrast to the sawtooth pattern, the clusters of amplifications in these tumors are clearly due to repeated rearrangement events that result from a structural change, such as telomere loss, that affects the stability of that arm alone. We cannot distinguish all profiles with this system, but the fundamental difference in the patterns may represent genomic lesions resulting from different mechanisms, and more than one mechanism may be operant to varying degrees within any given cancer.

A fourth type is the “flat” profile, cancer cells in which we observe no clear amplifications or deletions other than CNPs (Sebat et al. 2004) and single probe events, as discussed above, and the difference in the sex chromosomes. These profiles may represent either a sample with few tumor cells relative to the surrounding stroma, or a cancer that has no genomic rearrangements. Flat profiles such as W204 in Figure 2A represent less than 10% of the samples we have analyzed.

Characterization of Firestorm Instability

In ROMA profiles, firestorms display dramatic multiple segmental amplifications grouped on one arm, or occasionally, on both arms. The individual ampicons in
Firestorms have been observed at least once on most chromosomes in the tumors we have analyzed, but certain arms undergo this process more frequently. In particular, chromosomes 6, 8, 11, 17, and 20 are often affected, with 11q and 17q being the most frequently subject to these dramatic rearrangements. Notably, within the latter, the loci containing cyclin D1 on 11q and ERBB2 on 17q are most frequently amplified and may “drive” the selection of the events. Chromosomes 6, 8, and 20 have comparable frequency of firestorms, but the “drivers” for these events are less obvious.

The prediction that the amplification events were taking place on a chromosome arm was tested by a series of FISH experiments. We selected BACs or made primer-based probes from each narrow amplicon and each of the “spacer” regions in between. Two-color FISH experiments were performed on touch preparations made from a section of tumor samples WZ11 (presented here) and others to be published elsewhere. The results of the FISH experiments showed complete correspondence with the ROMA profile shown in Figure 4B. Probes from each amplicon yielded 8–15 spots in the FISH exposures, whereas probes for the intervening regions showed only the 2 spots expected for a diploid genome. Moreover, as shown previously for the aneuploid amplicons in WZ1, the spots corresponding to amplicons were clustered, suggesting that they co-localized on a single chromosome arm rather than being distributed throughout the genome as is the case for supernumerary or double minute chromosomes that are sometimes observed in cell culture. More notable, however, was the observation that when cells were exposed to probes from two different amplified peaks from the same firestorm in a two-color FISH experiment, the resulting sets of spots were co-localized in a single cluster. Figure 4B shows two examples using one pair of probes corresponding to MYC and CKS1 and another pair carrying FGFRI/BAG4 on the p arm of chromosome 7 and an unknown locus AK096200 on the 8q arm. These results suggest that, at least for the firestorm in WZ11, all of the amplified DNA regions are being carried on the same region of a single chromosome, as would be expected if the chromosome had entered into break-fusion-bridge (BFB) (McClintock 1938, 1941; Coquelle et al. 1997; Gisselsson et al. 2000) or break-induced replication (BIR) (Difilippantonio et al. 2002) models that have been invoked to explain chromosome instability in cancer cell lines, and by inference, in tumors themselves.

We have also been able to test the localization of the amplicons from two different multiply amplified chromosomes arms occurring in the same tumor sample. A chromosome localization model would predict that the spots from amplicons on different chromosomes would cluster separately from each other. This is what was observed in two-color FISH experiments using probes for ERBB2 on 17p and CCND1 (cyclin D1) on 11q in three tumor samples where both genes had been previously shown to be amplified by both FISH and ROMA. An example of this result is shown in Figure 4C using cells from sample WZ20 where earlier FISH experiments had shown ERBB2 to be present in more than 15 copies per cell and cyclin D1 to be present in 6 copies per cell. Two separate clusters are clearly visible, one containing only the red spots corresponding to cyclin D1 and the large cluster of green spots corresponding to ERBB2. Similar results were obtained using samples WZ1 (Fig. 1), WZ2 (Fig. 2A), and WZ17.

Prognostic Potential of Chromosome Rearrangement Patterns

One of the fundamental targets of this initial study is the comparison of whole-genome ROMA profiles from different clinical groups to evaluate the potential for ROMA as a prognostic tool. In this heuristic example, we analyzed all of the diploid samples in this collection by comparing subsets of patients grouped according to tumor grade, tumor size, node condition, and outcome (7-year survival). Due to the small numbers in this preliminary analysis, the samples were not sorted according to postoperative treatment. Two graphical methods for visualizing the aggregate data sets were frequency plots and mean amplitude plots. The frequency plot method reflects the fraction of samples in the subset for which each data point rises above (amplification) or below (deletion) a threshold value determined by the noise level in the experiments. The frequency plot method gives frequency of amplification or deletion of a given region, but it does not provide any indication of the degree of amplification, a factor that may often correlate with importance of a given focus in breast cancer.

The mean amplitude method sums the mean segmentation values for each probe over multiple experiments and divides by the total number of experiments. The rationale behind the mean amplitude plot is to provide an indication of the potential at any site for high-level amplification, while maintaining the ability to visualize deletions, which are generally limited in negative amplification or deletion of a given region, but it does not operationally yield an intermediate value approaching 0.75 at most. Amplification, on the other hand, can yield very strong peaks (comparative ratios of sample to control approaching 5.0) reflecting up to 30 copies of a given locus in the tumor as measured by FISH. Based on the ubiquity of amplification in breast tumors, it is logical to assume that copy number is related to phenotype.
in some way and, therefore, that peak height must be considered in comparative studies. The mean amplitude method takes into account both frequency and amplitude of a given locus, but peak height clearly can be driven by high values in a small fraction of the experiments.

The mean amplitude method also yields more information than a simple frequency plot when comparing deletions. Hemizygous deletions would be expected to give similar values on a cell-by-cell basis, but a deletion that has only recently appeared in a tumor will be less well represented in the ROMA profile than one that occurred earlier and is carried by a larger percentage of tumor cells. Therefore, the mean amplitude of the deletions shown in Figure 4C may be less than the frequency, where each event gets a unit value.

The data plotted in Figure 5 result from combining segmented data from 19 diploid, Grade III non-survivors (<7 years) and comparing them to 16 long-term diploid survivors matched for tumor size and grade. Clearly, designating a patient as a “survivor” or “non-survivor” at a specific time after surgery is not accurate in terms of the real progression of the disease. However, it is useful for understanding the relationship of disease progression to molecular events.

It is clear from Figure 5, A and B, and Figure 5, C and D, that, on average, tumors from non-survivors have suffered more genomic rearrangement than comparable survivors. This is consistent with accepted models for the relationship between genome instability and aggressiveness in breast cancer. What is perhaps surprising is that both the differences and the similarities between the survivor and non-survivor plots by either plotting method are nonrandom. The black arrows in each panel denote places where the activity as measured by frequency or amplitude is very similar between the two data sets and can be easily seen in individual tumors. These regions include duplications of 16p, deletion of 16q and 11q, and the duplication of 1q, as well as deletion of all of chromosome 22. With the possible exception of the frequency of chromosome 22 deletion, the frequencies of these events are nearly identical between the two data sets. That identity makes the differences, denoted by red arrows, at 8p, chromosome 6 amplification and deletion, 3p deletion, 11q amplification, 15q amplification, and 17q amplification. Although these data sets are too small to draw clear conclusions regarding prognosis, they do point to genomic regions that may well harbor such markers.

Nonetheless, the degree of similarity observed between the two analytical methods is striking. This means that important regions tend to be frequently affected by a high-amplitude genomic event. As described in subsequent sections, these amplifications are often parts of multiple amplification events on the same chromosome arm and are often very narrow. We have observed known oncogenes and tumor suppressors in breast cancer using these two methods. They have also pointed to regions that have not been previously identified as important in breast cancer.

Figure 4. FISH analysis of multiply amplified regions. Photographs show two-color FISH images of loci labeled in the ROMA profiles. (A) Tumor WZ11 showing a firestorm of amplification on chromosome 8 and cluster of spots compared to single-copy MDM2 on chromosome 12. (B) Enlarged view of chromosome 8 showing location of amplifications and putative oncogenes. FISH images show results of probing two separate pairs of amplicons within the same region. (C) Tumor WZ20 where amplicons appear on different chromosomes. FISH image shows that the repeated loci occupy separate regions of the nucleus.
Rearrangements in Low-Grade Tumors

Grade I diploid tumors in which the cells maintain their differentiation are generally considered to be less aggressive and have a very good prognosis irrespective of migration to the lymph nodes. Ten examples of Grade I tumors were examined in the current study, including four in which one or more nodes were affected. All were medium to large tumors between 20 mm and 30 mm in size. Although the number of samples is small, the similarity in ROMA profiles among the eight samples depicted in Figure 6 is dramatic and may provide insight into some of the earliest events leading to invasive breast cancer. Two of the ten Grade I samples yielded no detectable events and were not included in the figure. Six of eight tumors with any detectable events showed a characteristic rearrangement in chromosome 16 along with either a similar rearrangement of the arms of chromosome 8 or a duplication of the q arm of chromosome 1. All three of these events are seen in more highly rearranged breast cancer genomes such as those in Figure 6, and in fact, are among the most common events by frequency in all samples (see Fig. 2B). We believe that these low-grade tumors with little rearrangement in the genome provide an ideal opportunity to study the importance of these frequent events. Moreover, it is tempting to infer that these events are very likely among the earliest events taking place in a large fraction of tumors.

DISCUSSION

Microarray CGH and FISH Are Complementary Methods for Analyzing Genomic Change

The progression of cancer cells from their original normal state to uncontrolled growth, invasion, and metastasis clearly involves multiple genetic changes and may occur through a multiplicity of distinct pathways. Microarray CGH and FISH provide complementary tools for examining those events that involve gene copy number and nonreciprocal chromosome rearrangements. Microarray methods allow examination of the whole genome in one experiment, but by necessity, the data reflect an average of all of the genomes in all of the cells present in the original sample, both normal and cancerous. On the other hand, FISH reveals the exact number of copies of a given locus in each individual nucleus and can therefore detect and quantify the cancer-related events in tumor cells even when they are mixed with a significant fraction of normal cells, as is the case in most biopsy or surgical samples. Interphase FISH can also provide limited but important information concerning the structures of rearranged loci in a tumor cell population, as demonstrated by the “clustering” phenomenon observed in this work that bolsters (but does not prove) our firestorm interpretation. By itself, however, FISH is limited to testing only a few genes in each.
experiment. We have used a combined approach, using ROMA CGH to survey the genome and subsequent FISH to examine individual loci. Among the various microarray CGH methods, ROMA provides the highest resolution and sensitivity through the combination of reduced target complexity and the high density of features available from our proprietary version of the NimbleGen feature array.

Firestorms and Sawtooth Patterns

The complex genome profiles seen in highly rearranged breast tumors by ROMA CGH appear to represent different paths that cells may take in acquiring the altered gene expression that leads first to tumorigenesis and ultimately to metastasis. We have gone to some lengths to validate these patterns in view of their potential use in both prognosis and oncogene discovery. First, we have shown by interphase FISH studies on firestorm tumors that narrow peaks resolved by ROMA represent separate amplicons and are not simply the result of any noise in the system. Furthermore, we have shown that the multiple amplifications seen by ROMA CGH most often occur in the same cell and therefore represent an accumulation of events in a clonal population. Finally, we have learned that firestorms occur at preferred sites that are correlated with the genomic locations associated with higher risk, based on frequency plots of survivors and non-survivors. Additional work is under way using a combined FISH and ROMA approach to understand the mechanisms that induce global sawtooth patterns of rearrangement and the chromosome-limited rearrangements characteristic of firestorms. Multiple head-to-tail and head-to-head repeating amplicons have been observed in cancer cell lines (Coquelle et al. 1997; Gisselsson et al. 2000). Likewise, the telltale anaphase bridges characteristic of breakage-fusion-bridge cycles (McClintock 1938, 1941) are also frequently seen in cancer mitotic figures, leading to the suggestion that telomere fusion of chromatids is the major mechanism for high levels of amplification as observed in firestorms. Clearly, the process requires some structural characteristic of the recombining chromosome arms. Whether the key to that process resides in telomere loss or in recombination at short inverted repeats (Tanaka et al. 2002) or through a related mechanism, break-induced replication, where segments are copied from internal chromosome breaks (Salippantionio et al. 2002), is as yet an open question. It will be most interesting to determine whether a component of that peculiar cancer-related process can be blocked, thus providing another target for anticancer therapy.

A Possible Pattern to Progression

Another intriguing possibility that stems from studies of genomic rearrangement is the possibility of dissecting the pathways leading from noninvasive to invasive to metastatic cancer by tracking the events that occur in the most highly differentiated (least evolved) breast tumors. Certain specific chromosome arm gains and losses appear to be unexpectedly frequent in those tumors that show less than five total events. Those lesions, all of which have been reported elsewhere at various times in different contexts (Kallioniemi et al. 1994; Ried et al. 1995; Turkkonen et al. 1998, Pollack et al. 2002; Nesling et al. 2005), are duplication of 1q, 8q, and 16p, and deletion of 8p, 16q, and 22q. Not all of the events occur together in the same tumor, and there are not enough data as yet to test whether there is any intrinsic order to the timing of their appearance. We do note, however, that the frequency of these specific changes remains constant when we compare tumors from surviving
patients (or those with few events) (Fig. 6B) with subsets of tumors that have poor survival (and many more total events) (Fig. 6A). One interpretation of these results is that in the early stages of tumor development, cells undergo a subset of these specific gain or loss events as they give rise to proliferating clones. Subsequently, as these clones become less differentiated and gain potential to spread in the host, additional events accumulate. Thus, it is reasonable to speculate that there are early and late genomic events that can be separated according to the degree of progression exhibited by the cancer and that there is likely to be a genetic pathway, albeit a complex one, at work in the evolution of tumors.

This work, along with our previous published results (Lucito et al. 2003), confirms that ROMA profiles proportionally reflect the copy number of each microarray feature as measured by two-color FISH and that ROMA can be used to identify the boundaries of deletions, duplications, and amplifications. By compiling data from a large number of samples, we have begun to identify specific types of overall genomic patterns in breast cancer and relate them to clinical status and eventual patient outcome. The goal of these studies is to identify useful prognostic and therapeutic markers that will eventually help direct therapy in a clinical setting. We are confident that as the number of clinically annotated samples grows, prognostic information regarding clinical outcome as well as information regarding preferred treatment modalities can and will be derived.

ACKNOWLEDGMENTS

This work was supported by grants to M.W. from the National Institutes of Health (5R01-CA78544-07); the Department of the Army (W81XWH-04-1-0477; W81XWH-05-1-0068; W81XWH-04-0905); The Si- mions Foundation; Miracle Foundation; Breast Cancer Research Foundation; Long Islanders Against Breast Cancer; West Islip Breast Cancer Foundation; Long Island Breast Cancer (1 in 9); Elizabeth McFarland Breast Cancer Research Grant; and Breast Cancer Help Inc. M.W. is an American Cancer Society Research Professor. This work was supported by grants to A.Z. from the Swedish Cancer Society (grant number 04Rd-B04- 38XAC) and from the Stockholm Cancer Society (03:171 and 02:144). The authors thank Dr. James Feramisco for careful review and helpful comments on the manuscript.

REFERENCES

High-Resolution ROMA CGH and FISH Analysis of Aneuploid and Diploid Breast Tumors


Cold Spring Harb Symp Quant Biol 2005 70: 51-63
Access the most recent version at doi:10.1101/sqb.2005.70.055

References
This article cites 30 articles, 19 of which can be accessed free at:
http://symposium.cshlp.org/content/70/51.full.html#ref-list-1

License

Email Alerting Service
Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or click here.

To subscribe to Cold Spring Harbor Symposia on Quantitative Biology go to:
http://symposium.cshlp.org/subscriptions

Copyright 2005, Cold Spring Harbor Laboratory Press