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

J. Hicks,* L. Muthuswamy,* A. Krasnitz,* N. Navin,*[†] M. Riggs,*
V. Grubor,* 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, provide 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.

Alterations 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; Ried et al. 1997; Tirkkonen et al. 1998). Subsequently, microarray methods employing CGH have increased resolution and reproducibility, and have improved throughput (Ried et al. 1995; Pollack et al. 2002; Albertson 2003; Lage et al. 2003). These published microarray studies have largely validated the results of cytogenetic CGH, but have 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. 1987; Slamon et al. 1989; Menard 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-confirm 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 contain-

ing 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 genomic 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 25 μl of hybridization solution (50% formamide, 5x 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/0.2x SSC, 30 seconds in 0.2x SSC, and 30 seconds in 0.05x 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 deviant 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 population, 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 residual 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 deviance 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 SpectrumGreen™ (Vysis Inc., Downers Grove, Illinois). Denaturation of probe and target DNA was performed at 90°C for 5 minutes, followed by hybridization in a humidity chamber at 47°C overnight. The coverglasses were then removed and the slides were washed in 2x SSC for 10 minutes 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; Vectashield) as a counterstain for the nuclei. Evaluation of signals was carried out in an epifluorescence microscope. Selected cells were photographed in a Zeiss Axioplan 2 microscope equipped with Axio Cam MRM CCD camera and Axio Vision software.

Probe design for FISH. Hybridization probes for FISH were constructed by one of two methods. For the interdigitation analysis, probes were created from BACs 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 identified through the PROBER algorithm designed in this laboratory. Genomic sequences of 100 kb containing target amplifications were tiled with 50 probes (800–1400 bp) selected with PROBER Probe Design Software created in our laboratory. PROBER uses a distributed annotated sequence retrieval request (Dowell et al. 2001) to request a genomic sequence and the Mer-Engine (Healy et al. 2003) to mask the sequence for repeats. Mer lengths of 18 that occur more than twice in the human genome (UCSC 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 algorithm 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 Eppendorf (Cat. 0032002.447) from EBV-immortalized cell line DNA (Chp-Skn-1) DNA (100 ng) with 55°C annealing, 72°C extension, 2-minute extension time, and 23 cycles. Probes were purified with Qiagen PCR purification columns (Cat. 28104) and combined into a single probe cocktail (10–25 μg total probes) for dye labeling and metaphase/interphase FISH.

Measurement of DNA content. The ploidy of each tumor was determined by measurement of DNA content using Feulgen photocytometry (Forsslund and Zetterberg 1990; Forsslund 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 major peak at 2c DNA content with a smaller peak of G_2 phase replicating cells that corresponds to the mitotic index. 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 undergoing 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 Institute. 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 available. The clinical characteristics of this sample set are shown in Table 1.

Each of the tumors in this study was initially categorized 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 compared 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 manufactured to our specifications by NimbleGen, Inc. (Reykjavik, 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

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

Karolinska Inst. Sweden	Total	Node (pos/neg)	Median age at diagnosis	Grade I/II/III	Size (mm) <20/>20	PR* (+/−)	ER* (+/−)	ERBB2[+] amp/norm
Diploid (Survival >7 yr)	60	28/31	52	8/11/33	19/41	41/9	43/7	3/57
Diploid (Survival <7 yr)	39	14/25	57	3/12/16	11/25	20/13	24/8	9/30
Aneuploid	41	28/13	49	0/2/22	21/20	14/19	25/10	15/26

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.

et al. 2004). 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.
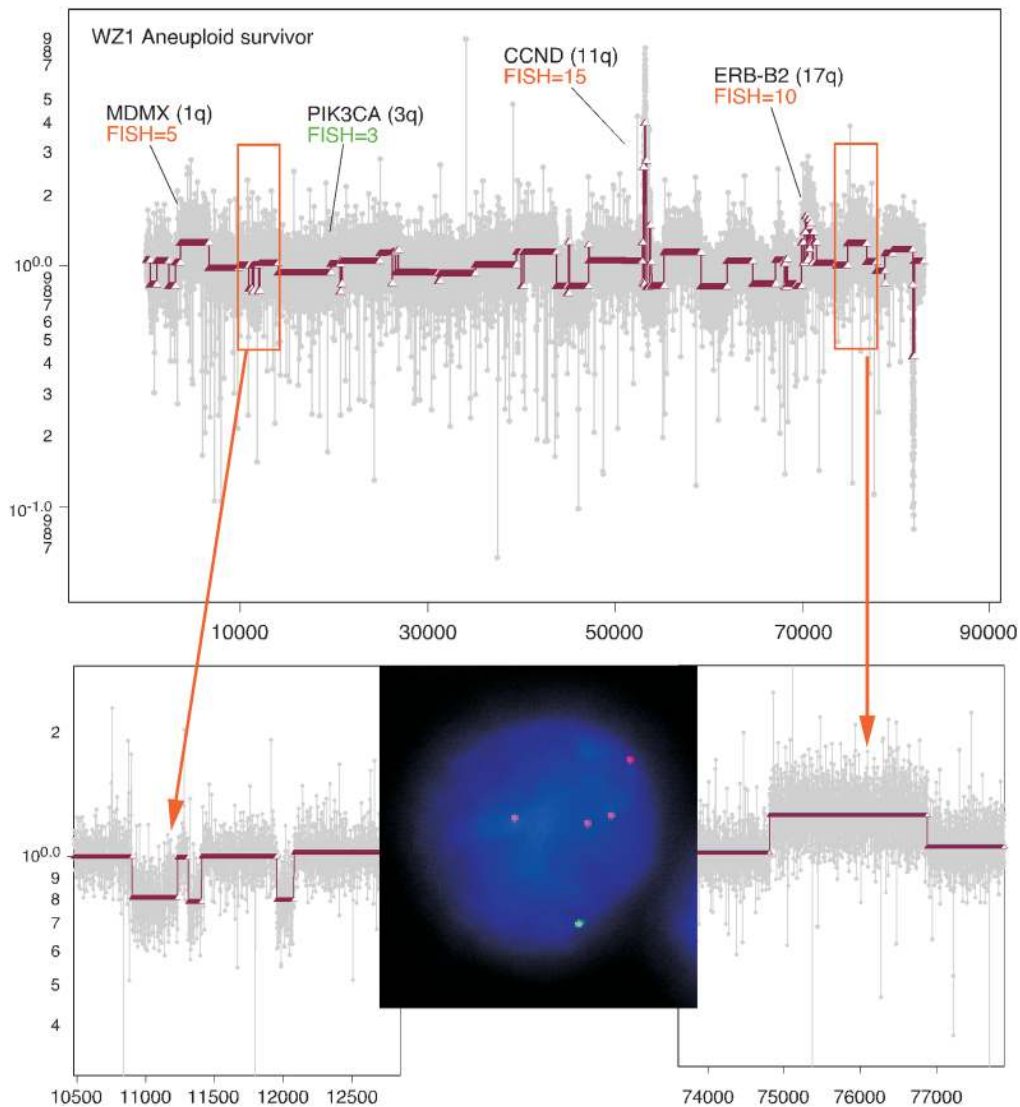
Figure 1. Comparison of copy number as assayed by ROMA and FISH. Tumor WZ1 is aneuploid with an average genome copy number of 3n by FACS analysis. The results of FISH probes for various loci are indicated in the top graph. The bottom panels show enlarged views of small deletions and duplications picked to demonstrate the correspondence between FISH and ROMA. The photograph 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 (Ried et al. 1997; Pollack et al. 2002; Albertson 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 PROBER 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 (not shown here) provide confidence 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. Furthermore, we can use ROMA to define a mathematical parameter that reflects the degree to which a population of tumor cells differs from a normal euploid genome.

We note that the segmented mean value for the X chromosome in a typical diploid female/diploid male experiment ranges from 1.3 to 1.5. We have established a theoretical peak broad mean value for the X chromosome at 1.65. This is significantly higher than values reported for an expected 2:1 ratio in non-representational microarray CGH methods (Pollack et al. 2002), but still less than the expected value of 2. This ratio, which averages about 1.45 in our experiments, sets a rough benchmark for other events, particularly duplications or deletions of chromosome arms or segments. Most other broad events, particularly in diploids, show amplitudes less than that of the X as would be expected since tumor samples generally contain a certain fraction of normal cells. Additionally, because all chromosomal events may not have occurred at the same time in the development of the tumor, the segmentation value of later events would have a characteristic fractional representation in the ROMA profile. Using FISH to confirm copy numbers, we have determined that whereas ROMA values underestimate copy number, they are very sensitive to the existence of events and can accurately detect events with a deviation from the baseline segmentation of as little as ± 0.1.

Aneuploids Versus Diploids

Because of the complexity of data accumulated in CGH experiments, it is usually necessary to process multiple experiments together and to analyze the aggregate by statistical methods. The drawback of such methods is that they obscure the potential for identifying unique patterns and phenotypes among individual tumors. We therefore present in Figure 2A a representative set of ROMA profiles for tumors to demonstrate the variety of forms that samples in this study can take.

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 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 amplifications 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, but the frequency 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
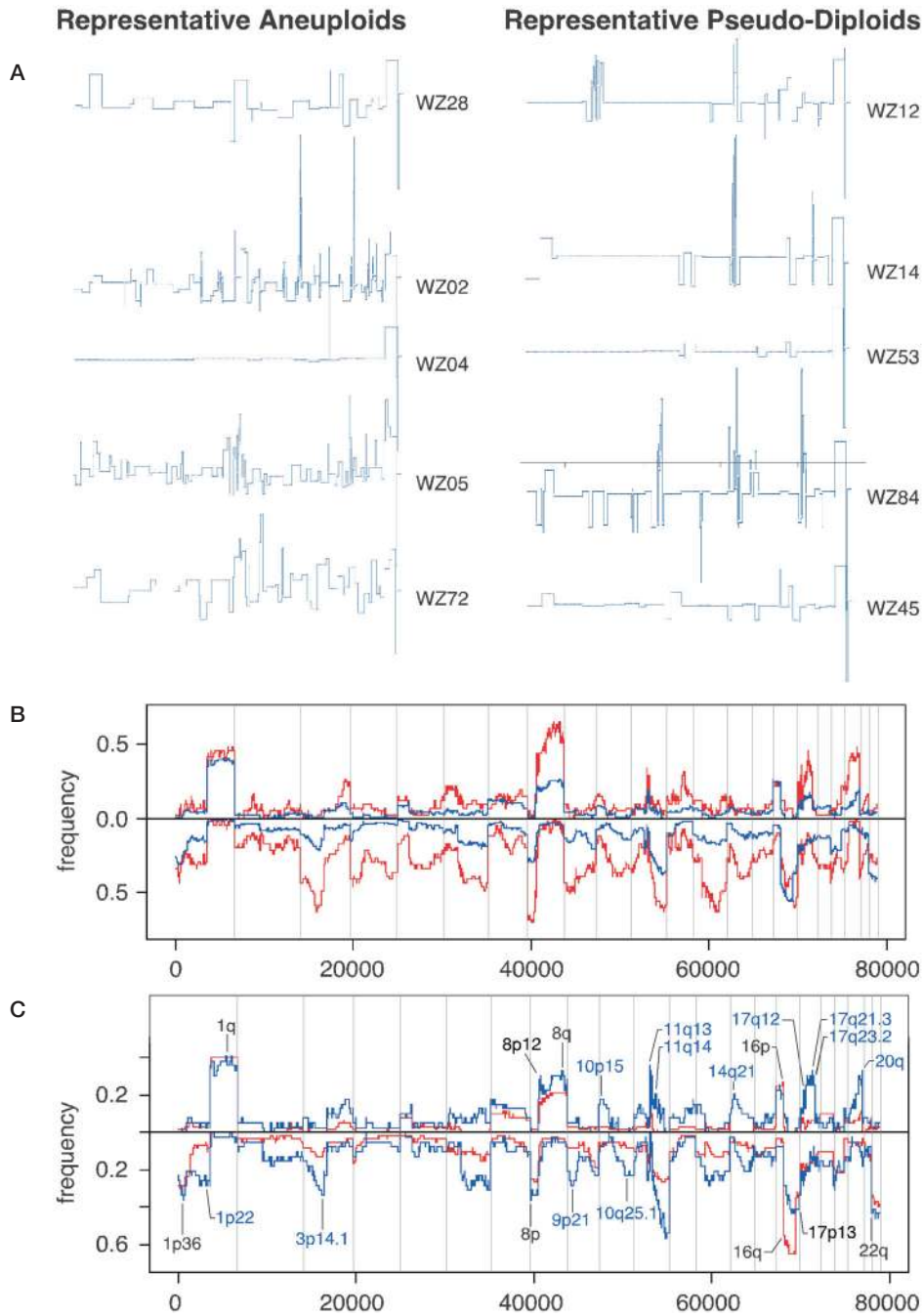
Figure 2. Examples of aneuploid and pseudo-diploid tumors. (*A*) Representative ROMA profiles showing that aneuploid tumors in general exhibit an overall greater frequency of chromosome rearrangements than do pseudo-diploid tumors. (*B*) Comparative frequency plots of amplification (up) and deletion (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*).

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.
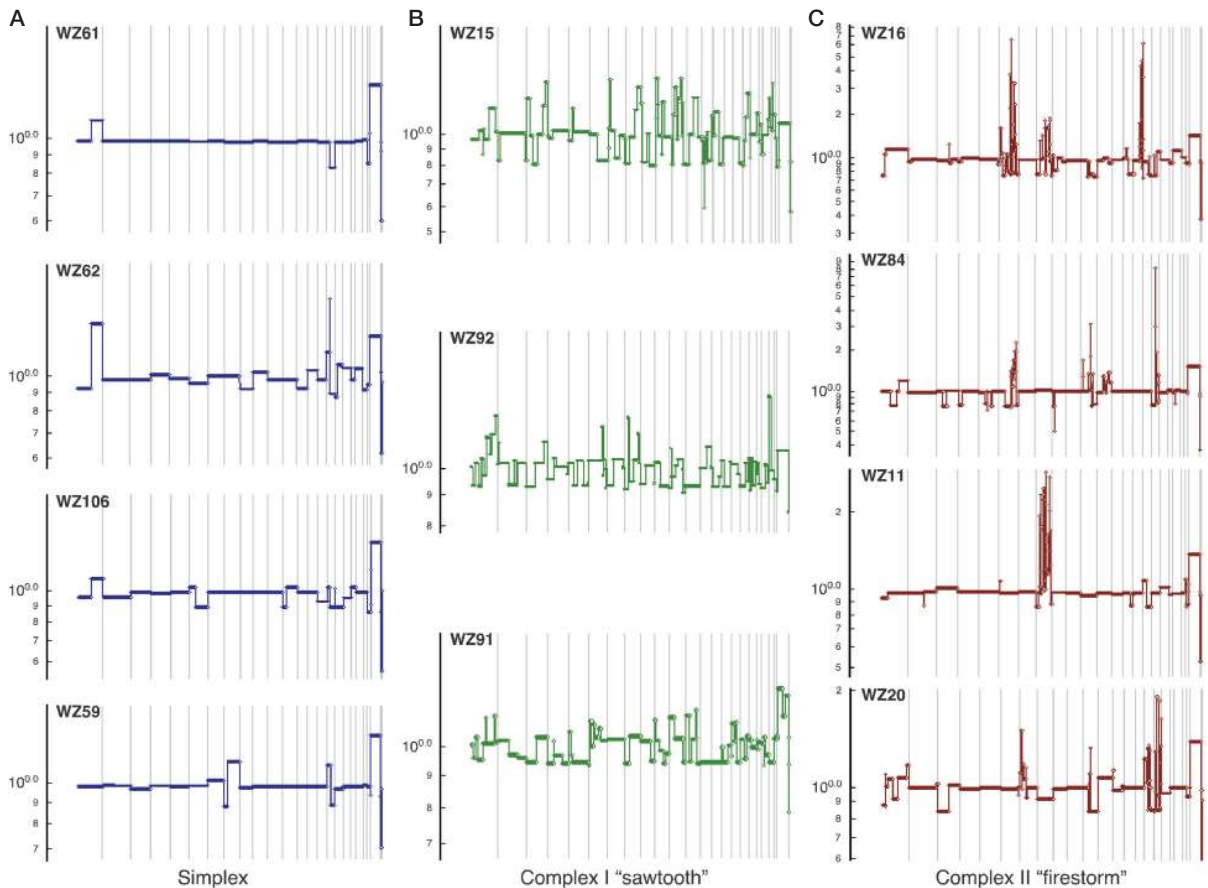
Figure 3. Major types of tumor genomic profiles. Segmentation profiles for individual tumors representing each category: (*A*) simplex; (*B*) complex type I or "sawtooth"; (*C*) complex type II or "firestorm." Scored events consist of a minimum of six consecutive probes in the same state. Y-axis displays the geometric mean value of two experiments on log scale. Note that the scale of the amplifications in panel *C* is compressed relative to panels *A* and *B* due to the high levels of amplification in firestorms. Chromosomes 1–22 plus X and Y are displayed in order from left to right according to probe position.

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 amplifications. 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 recombination/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 WZ04 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 amplicons in

these firestorms are often separated by segments that do not appear to be amplified, yielding a pattern of interdigitated amplification as shown for chromosome 8 in tumor WZ11 (shown graphically in Fig. 3C). In cases where this pattern of amplification is observed, the interdigitated amplification pattern is limited to one or a few chromosome arms, and amplicons are clearly not randomly distributed throughout the genome (Figs. 2A and 3B, C). We infer from this observation that the phenomenon is a result of aberrant replication/recombination events that occur on a particular chromosome arm rather than a general induction of amplification functions operating throughout the genome.

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 colocalized 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 colocalized in a single cluster. Figure 4B shows two examples using one pair of probes corresponding to MYC and CKS1 and another pair carrying FGFR1/BAG4 on the p arm of chromosome 8 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 chromosome 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 CCND (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 locus 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 amplitude to the value of a hemizygous loss. Regions of hemizygous loss would be expected to yield ratios of 0.5, but 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
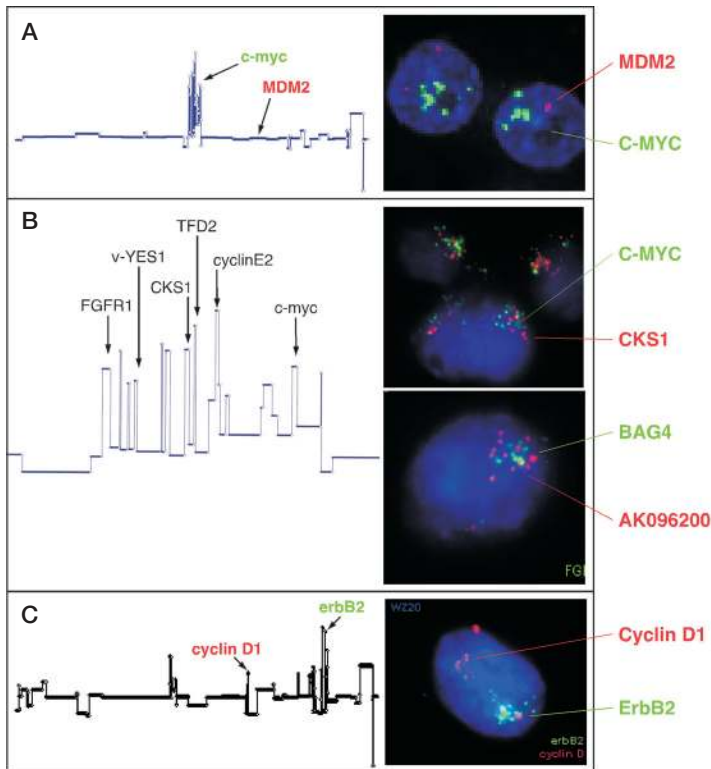
Figure 4. FISH analysis of multiply amplified regions. Photographs show two-color FISH images of loci labeled in the ROMA profiles. (*A*) Tumor WZ11 showing a firestorm of amplification on chromosome 8 and cluster of spots compared to single-copy MDM2 on chromosome 12. (*B*) Enlarged view of chromosome 8 showing location of amplicons 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.

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 aggres-

siveness 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 amplicons 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.

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

Microarry 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
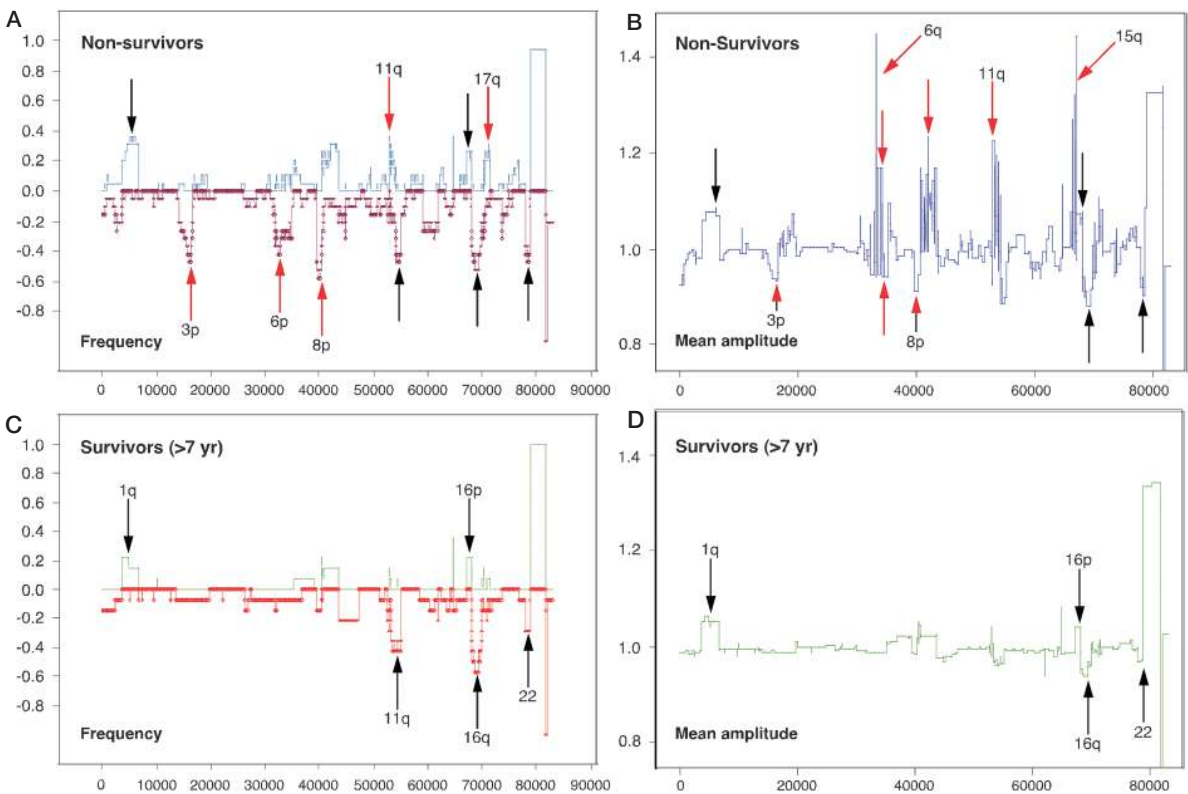
Figure 5. Comparison of 35 Grade III diploid tumors from eventual survivors vs. non-survivors. Survival longer or shorter than 7 years as shown in Fig. 2C. (*A* and *B*) Frequency plots; (*C* and *D*) mean amplitude plots. Black arrows indicate events common to both classes. Red arrows indicate events enriched in the non-survivor class.

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 Nimble-Gen 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 (Difilippantonio 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. These lesions, all of which have been reported elsewhere at various times in different contexts (Kallioniemi et al. 1994; Ried et al. 1995; Tirkkonen et al. 1998; Pollack et al. 2002; Nessling 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

Common Origins in Low Grade Breast Tumors
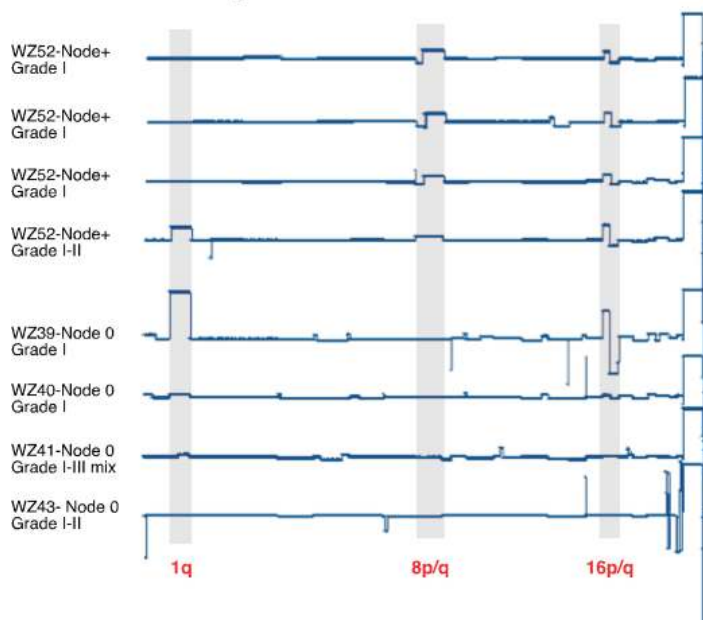
Figure 6. Comparison of Grade I/II diploid tumors by ROMA. A total of 10 low-grade tumors were included in the data set. The two samples not shown exhibited no detectable events. Regions of common chromosomal rearrangements are shaded. All of the shaded areas are among the most common sites of rearrangement in all breast tumors collectively shown in Fig. 5.

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.

REFERENCES

Albertson D.G. 2003. Profiling breast cancer by array CGH. *Breast Cancer Res. Treat.* **78:** 289.

Balmain A., Gray J., and Ponder B. 2003. The genetics and genomics of cancer. *Nat. Genet.* (suppl.) **33:** 238.

Coquelle A., Pipiras E., Toledo F., Buttin G., and Debatisse M. 1997. Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell* **89:** 215.

DePinho R.A. and Polyak K. 2004. Cancer chromosomes in crisis. *Nat. Genet.* **36:** 932.

Difilippantonio M.J., Petersen S., Chen H.T., Johnson R., Jasin M., Kanaar R., Ried T., and Nussenzweig A. 2002. Evidence for replicative repair of DNA double-strand breaks leading to oncogenic translocation and gene amplification. *J. Exp. Med.* **196:** 469.

Dowell R.D., Jokerst R.M., Day A., Eddy S.R., and Stein L. 2001. The distributed annotation system. *BMC Bioinformatics* **2:** 7.

Forsslund G. and Zetterberg A. 1990. Ploidy level determinations in high-grade and low-grade malignant variants of prostatic carcinoma. *Cancer Res.* **50:** 4281.

Forsslund G., Nilsson B., and Zetterberg A. 1996. Near tetraploid prostate carcinoma. Methodologic and prognostic aspects. *Cancer* **78:** 1748.

Gisselsson D., Pettersson L., Hoglund M., Heidenblad M., Gorunova L., Wiegant J., Mertens F., Dal Cin P., Mitelman F., and Mandahl N. 2000. Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proc. Natl. Acad. Sci.* **97:** 5357.

Healy J., Thomas E.E., Schwartz J.T., and Wigler M. 2003. Annotating large genomes with exact word matches. *Genome Res.* **13:** 2306.

Kallioniemi A., Kallioniemi O.P., Sudar D., Rutovitz D., Gray J.W., Waldman F., and Pinkel D. 1992a. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. *Science* **258:** 818.

Kallioniemi A., Kallioniemi O.P., Waldman F.M., Chen L.C., Yu L.C., Fung Y.K., Smith H.S., Pinkel D., and Gray J.W. 1992b. Detection of retinoblastoma gene copy number in metaphase chromosomes and interphase nuclei by fluorescence in situ hybridization. *Cytogenet. Cell Genet.* **60:** 190.

Kallioniemi A., Kallioniemi O.P., Piper J., Tanner M., Stokke T., Chen L., Smith H.S., Pinkel D., Gray J.W., and Waldman F.M. 1994. Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. Natl. Acad. Sci.* **91:** 2156.

Kallioniemi O.P., Kallioniemi A., Kurisu W., Thor A., Chen L.C., Smith H.S., Waldman F.M., Pinkel D., and Gray J.W. 1992. ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. *Proc. Natl. Acad. Sci.* **89:** 5321.

Kronenwett U., Huwendiek S., Ostring C., Portwood N., Roblick U.J., Pawitan Y., Alaiya A., Sennerstam R., Zetterberg A., and Auer G. 2004. Improved grading of breast adenocarcinomas based on genomic instability. *Cancer Res.* **64:** 904.

Lage J.M., Leamon J.H., Pejovic T., Hamann S., Lacey M., Dillon D., Segraves R., Vossbrinck B., Gonzalez A., Pinkel D., Albertson D.G., Costa J., and Lizardi P.M. 2003. Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Res.* **13:** 294.

Lucito R., Healy J., Alexander J., Reiner A., Esposito D., Chi M., Rodgers L., Brady A., Sebat J., Troge J., West J.A., Rostan S., Nguyen K.C.Q., Powers S., Ye K.Q., Olshen A., Venkatraman E., Norton L., and Wigler M. 2003. Representational oligonucleotide microarray analysis: A high-resolution method to detect genome copy number variation. *Genome Res.* **13:** 2291.

McClintock B. 1938. The production of homozygous deficient tissues with mutant characteristics by means of the aberrant mitotic behavior of ring-shaped chromosomes. *Genetics* **23:** 315.

———. 1941. The stability of broken ends of chromosomes in *Zea mays. Genetics* **26:** 234.

Menard S., Fortis S., Castiglioni F., Agresti R., and Balsari A. 2001. HER2 as a prognostic factor in breast cancer. *Oncology* **61:** 67.

Nessling M., Richter K., Schwaenen C., Roerig P., Wrobel G., Wessendorf S., Fritz B., Bentz M., Sinn H.-P., Radwimmer B., and Lichter P. 2005. Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. *Cancer Res.* **65:** 439.

Pollack J.R., Sorlie T., Perou C.M., Rees C.A., Jeffrey S.S., Lonning P.E., Tibshirani R., Botstein D., Borresen-Dale A.L., and Brown P.O. 2002. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional

program of human breast tumors. *Proc. Natl. Acad. Sci.* **99:** 12963.

Ried T., Liyanage M., du Manoir S., Heselmeyer K., Auer G., Macville M., and Schröck E. 1997. Tumor cytogenetics revisited: Comparative genomic hybridization and spectral karyotyping. *J. Mol. Med.* **75:** 801.

Ried T., Just K.E., Holgreve-Grez H., du Manoir S., Speicher M.R., Schröck E., Latham C., Blegen H., Zetterberg A., Cremer T., and Auer G. 1995. Comparative genomic hybridization of formalin-fixed, paraffin-enbedded breast tumors reveals different patterns of chromosomal gains and losses in fibroadenomas and diploid and aneuploid carcinomas. *Cancer Res.* **55:** 5415.

Sebat J., Lakshmi B., Troge J., Alexander J., Young J., Lundin P., Maner S., Massa H., Walker M., Chi M., Navin N., Lucito R., Healy J., Hicks J., Ye K., Reiner A., Gilliam T.C., Trask B., Patterson N., Zetterberg A., and Wigler M. 2004. Large-scale copy number polymorphism in the human genome. *Science* **305:** 525.

Slamon D.J., Godolphin W., Jones L.A., Holt J.A., Wong S.G., Keith D.E., Levin W.J., Stuart S.G., Udove J., and Ullrich A., et al. 1989. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **244:** 707.

Tanaka H., Tapscott S.J., Trask B.J., and Yao M.-C. 2002. Short inverted repeats initiate gene amplification through the formation of a large DNA palindrome in mammalian cells. *Proc. Natl. Acad. Sci.* **99:** 8772.

Tirkkonen M., Tanner M., Karhu R., Kallioniemi A., Isola J., and Kallioniemi O.P. 1998. Molecular cytogenetics of primary breast cancer by CGH. *Genes Chromosomes Cancer* **21:** 177.

van de Vijver M.J., van de Bersselaar R., Devilee P., Cornelisse C., Peterse J., and Nusse R. 1987. Amplification of the neu (c-erbB-2) oncogene in human mammmary tumors is relatively frequent and is often accompanied by amplification of the linked c-erbA oncogene. *Mol. Cell. Biol.* **7:** 2019.

Yang Y.H., Dudoit S., Luu P., Lin D.M., Peng V., Ngai J., and Speed T.P. 2002. Normalization of cDNA microarray data: A robust composite method addressing single and multiple slide sytematic variation. *Nucleic Acids Res.* **30:** e15.