

Novel patterns of genome rearrangement and their association with survival in breast cancer

James Hicks,^{1,10} Alexander Krasnitz,¹ B. Lakshmi,¹ Nicholas E. Navin,^{1,2} Michael Riggs,¹ Evan Leib,¹ Diane Esposito,¹ Joan Alexander,¹ Jen Troge,¹ Vladimir Grubor,¹ Seungtae Yoon,¹ Michael Wigler,¹ Kenny Ye,⁹ Anne-Lise Børresen-Dale,^{3,4} Bjørn Naume,⁵ Ellen Schlichting,⁶ Larry Norton,⁷ Torsten Hägerström,⁸ Lambert Skoog,⁸ Gert Auer,⁸ Susanne Månér,⁸ Pär Lundin,⁸ and Anders Zetterberg⁸

¹Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA; ²Watson School of Biological Sciences, Cold Spring Harbor, New York 11724, USA; ³Department of Genetics, Institute for Cancer Research, Rikshospitalet-Radiumhospitalet Medical Center, 0310 Oslo, Norway; ⁴Faculty of Medicine, University of Oslo, 0316 Oslo, Norway; ⁵The Cancer Clinic, Rikshospitalet-Radiumhospitalet Medical Center, 0310 Oslo, Norway; ⁶Department of Surgery, Ullevål Univ. Hospital, 0407 Oslo, Norway; ⁷Memorial Sloan-Kettering Cancer Center, New York, New York 10021, USA; ⁸Karolinska Institutet, Department of Oncology-Pathology, 171 76 Stockholm, Sweden; ⁹Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, New York 10461, USA

Representational Oligonucleotide Microarray Analysis (ROMA) detects genomic amplifications and deletions with boundaries defined at a resolution of ~50 kb. We have used this technique to examine 243 breast tumors from two separate studies for which detailed clinical data were available. The very high resolution of this technology has enabled us to identify three characteristic patterns of genomic copy number variation in diploid tumors and to measure correlations with patient survival. One of these patterns is characterized by multiple closely spaced amplicons, or “firestorms,” limited to single chromosome arms. These multiple amplifications are highly correlated with aggressive disease and poor survival even when the rest of the genome is relatively quiet. Analysis of a selected subset of clinical material suggests that a simple genomic calculation, based on the number and proximity of genomic alterations, correlates with life-table estimates of the probability of overall survival in patients with primary breast cancer. Based on this sample, we generate the working hypothesis that copy number profiling might provide information useful in making clinical decisions, especially regarding the use or not of systemic therapies (hormonal therapy, chemotherapy), in the management of operable primary breast cancer with ostensibly good prognosis, for example, small, node-negative, hormone-receptor-positive diploid cases.

[Supplemental material is available online at www.genome.org and at <http://roma.cshl.edu>.]

As cancers evolve, their genomes undergo many alterations, including point mutations, rearrangements, deletions, and amplifications, which presumably alter the ability of the cancer cell to proliferate, survive, and spread in the host (Balmain et al. 2003; DePinho and Polyak 2004). An understanding of these changes will allow the design of more rational therapies and, by providing precise diagnostic criteria, allow fitting the correct therapy to each patient according to need. Primary breast cancers in particular exhibit a wide range of outcomes and degrees of benefit from systemic therapies, which are incompletely predicted by conventional clinical and clinico-pathological features. This is especially apparent in the case of small primaries without axillary lymph node involvement, which usually have a good prognosis but are sometimes associated with eventual metastatic dissemination and death.

Breast tumors have long been known to suffer multiple genomic rearrangements during their development, and thus it is reasonable to hypothesize that clinical heterogeneity may be

caused by the existence of genetically distinct subgroups. One common approach to the molecular characterization of breast cancer has been “expression profiling,” measuring the entire transcriptome by microarray hybridization. Expression profiling has been very effective at revealing phenotypic subtypes of breast cancer and clinically useful diagnostic patterns of gene expression in tumors (Perou et al. 2000; Sorlie et al. 2001; Ahr et al. 2002; van't Veer et al. 2002; Sotiriou 2003; Paik et al. 2004). Expression profiling does not look directly at underlying genetic changes, and its dependence on RNA, a fragile molecule, creates some problems in standardization and cross-validation of microarray platforms. Moreover, variation in the physiological context of the cancer within the host, such as the proportion of normal stroma and the degree of inflammatory response, or the degree of hypoxia, as well as methods used for extraction and preservation of sample, are all potentially useful but confounding factors (Edén et al. 2004).

Direct analysis of the tumor genome provides an alternative and perhaps, complementary, means of comparing breast tumors by revealing the genetic events accumulated during tumor progression. We have begun a long-term genomic study of clinically well-defined sets of breast cancer patients with a high-resolution

¹⁰Corresponding author.

E-mail hicks@cshl.edu; fax (516) 367-8381.

Article is online at <http://www.genome.org/cgi/doi/10.1101/gr.5460106>. Freely available online through the *Genome Research* Open Access option.

microarray technique called Representational Oligonucleotide Microarray Analysis (ROMA) (Lucito et al. 2003). ROMA is based on the principle that noise in microarray hybridization can be significantly reduced by reducing the complexity of the labeled DNA target in the hybridization mix. In its present configuration, ROMA uses a "representation" of the genome created by PCR amplification of the smallest fragments of a BglIII restriction digest. The representation contains <3% of the complexity of the normal human genome and is specifically matched with a unique microarray containing >83,000 oligonucleotide probes designed to pair with the amplified fragments. Coupled with an efficient edge-detection or segmentation algorithm, ROMA yields highly precise profiles of even closely spaced amplicons and deletions. At present, ROMA is capable of detecting the breakpoints of chromosomal events at a resolution of 50 kb. This study is intended to explore whether high resolution of the genetic events in tumors can form an additional basis for the clinical assessment of breast cancer.

The first global studies capable of resolving deletions and amplifications combined comparative genomic hybridization (CGH) and cytogenetics (Kallioniemi et al. 1992a,b,c), and this approach has been applied to breast tumors (Kallioniemi et al. 1994; Ried et al. 1997; Tirkkonen et al. 1998). Subsequently, microarray methods using CGH have increased resolution and reproducibility and 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 cytogenetics 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 *ERBB2*, as studied by fluorescence in situ hybridization (FISH) or Q-PCR, has been clinically useful (van de Vijver et al. 1987; Slamon et al. 1989; Menard et al. 2001). 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 sought to determine whether there were features in the genomes of tumor cells that correlated with clinical outcome in a uniform population of women with "diploid" breast cancers. We chose this population because a significant number of cases culminate in death despite their clinical and histo-pathological parameters that would predict a favorable outcome. Our population of 99 diploid cancers drew from a bank at the Karolinska Institute (KI), and was comprised of long-term and short-term survivors who were similar for node status, grade, and size. For part of our analysis, we draw on additional studies in progress, one using 41 aneuploid (defined as >2n DNA content) (see Methods) cancers from KI, and the other using an additional 103 can-

cers from the Oslo Micrometastasis Study, Oslo, Norway (OMS). The latter set was not scored for ploidy and has only an average of 8 yr follow up and is included in this study only for comparison of overall frequency of events. The individual genome profiles from the KI data set but not the OMS data set are in the Supplemental material and at (<http://roma.cshl.edu>). The OMS data set will be posted as part of a second paper specifically dealing with that group. The makeup of these sample sets with respect to clinical parameters is summarized in Table 1.

Our studies demonstrate a striking similarity of genome profiles from two different study populations, as well as the commonality of affected loci in aneuploid and diploid cancers. Significantly, we observe a different genome profile between diploid tumors with good and poor outcome. The complexity and the number of events, captured in a mathematical measure, has led to our working hypothesis that genomic profiling may be useful for the molecular staging of breast cancer, and, when validated by further studies, may have implications for clinical practice.

Results

The clinical makeup of the sample sets included in this study is summarized in Table 1. The KI tumor data set was assembled from a collection of >10,000 fresh frozen surgical tumor samples with detailed pathology profiles and long-term follow up. The patients in this study underwent surgery between 1987 and 1992 yielding follow-up data for survival of 15–18 yr. The sample set was assembled with the goal of studying a statistically significant population of otherwise rare outcomes, particularly diploid tumors that led to death within 7 yr, and aneuploid tumors with long-term survival (described in Methods). At the same time, the sample was balanced with respect to tumor size, grade, node involvement, and hormone receptor status. Treatment information is also available in the clinical table available in the Supplemental material; however, the sample set was not stratified according to treatment because the treatment groups are too fragmented to be significant. The Norwegian tumor set was selected from a trial previously described by Wiedswang et al. (2003) designed to identify markers associated with micrometastasis at the time of diagnosis (i.e., disseminating tumor cells in blood and bone marrow). The patients included in the study were recruited between 1995 and 1998, and fresh frozen tumors were available for a subset that was not selected for particular characteristics.

Processing individual cancer genomes

We examined all breast cancer genomes with ROMA, an array-based hybridization method that uses genomic complexity reduction based on representations. In the present case, we per-

Table 1. Distribution of patients and clinical parameters in the Swedish and Norwegian data sets

| Karolinska Institutet, Sweden | Total | Node (pos/neg) | Median age at diagnosis | Grade I/II/III | Size (mm) <20/>20 | PR ^a (+/-) | ER ^a (+/-) | <i>ERBB2</i> ^b 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 |
| Oslo Micrometastasis Study (OMS) | 103 | 52/46 | 63 | 10/50/41 | 44/55 | 43/57 | 58/44 | 27/76 |

Numbers will not add up exactly because of partial information on certain individual cases.

^aProgesterone (PR) and estrogen (ER) receptors measured by ligand binding; (pos) ≥ 0.5 fg/ μ g protein.

^b*ERBB2* amplification scored by ROMA as segmented ratio >0.1 above baseline.

formed comparative hybridization using BglIII representations, and arrays of 85,000 oligonucleotide (50-mer) probes with a Poisson distribution throughout the genome and a mean interprobe distance of 35 kb (Lucito et al. 2003). In all cases, we compared tumor DNA from a patient to a standard unrelated male human genome. We performed hybridizations in duplicate with color-reversal, and data were rendered as normalized ratios of probe hybridization intensity of tumor to normal.

The normalized ratios are influenced by many factors, including the signal-to-noise characteristics that differ for each probe, sequence polymorphisms in the genomes that affect the BglIII representation, DNA degradation of the sample, and other variation in reagents and protocols during the hybridization and scan. Statistical processing called “segmentation” identifies the most likely state for each block of probes, thus reducing the noise in the graphical presentation of the profile.

Within each raw ROMA profile, segmentation places consecutive probe intensity ratios into a series of distinct distributions, reflecting the alterations that occur when blocks of the genome are amplified, duplicated, or deleted. Several methods for segmentation have been published by us and others (Daruwala et al. 2004; Olshen et al. 2004), but in the present case, and in the interest of having very solid findings, we have used a simplified method that recognizes distinct distributions of ratio based on minimization of variance and a Kolmogorov-Smirnov test with P -values set at 10^{-5} (see Methods). All methods converge on roughly the same segmentation pattern, especially at the boundaries, or edges, of events, but the simplified method used herein does not consider short segments (sets of probes less than six). On average, the resolution of the edges of a gene copy number alteration event is ~50 kb under our present conditions. We report each probe ratio as the mean of the medians of the ratios within the segment to which that probe belongs, producing a “segmented profile” of each cancer. Both raw ratios and segmented ratios are posted on our Web site. Events less than six probes in length are, of course, visible in the unsegmented data and can be segmented by other methods, such as Hidden Markov Models (HMM); however, these very narrow events do not affect the conclusions of this report and are excluded from the statistical analysis for simplicity.

Single nucleotide polymorphisms (SNPs), found in all profiles, are present in our methods that use restriction endonuclease-based representations. These are most often the result of sequence differences between sample and reference that alter the restriction sites used in the representation process. For purposes of this report, they merely contribute to noise and do not significantly affect segmentation. However, both rare copy number variants (CNVs) and more

prevalent copy number polymorphisms (CNPs) (Sebat et al. 2004) will be present in any high-resolution copy number scan, regardless of method, when comparing one person to another. All of our tumor profiles are obtained by comparison to an unrelated standard normal male. If these CNPs and CNVs are not masked, analysis could mistake either for a cancer lesion. We have compiled a list of common CNPs and rare CNVs by profiling healthy cells from 482 individuals, and we used these to mask the “normal” CNPs in our tumor profiles as described in Methods, yielding a “masked segmented profile.” We post the masked segmented profiles in the Supplemental material. The collection of CNPs used for masking includes but is not limited to Scandinavian individuals and represents at most a few hundred probes being removed from consideration for segmentation in any sample. A CNP falling under a larger (cancer-related) event does not affect the segmentation of that event. Both the Kolmogorov-Smirnov segmentation software and the CNP masking algorithms are posted at <http://roma.cshl.edu> in the forms of scripts interpretable by R or S+ statistical analysis software.

The mean ratios within segments are not directly proportional to true copy number. The unknown proportion of “normal” stroma in the surgical biopsies, the potential for clonal variation, and nonspecific hybridization background signal all

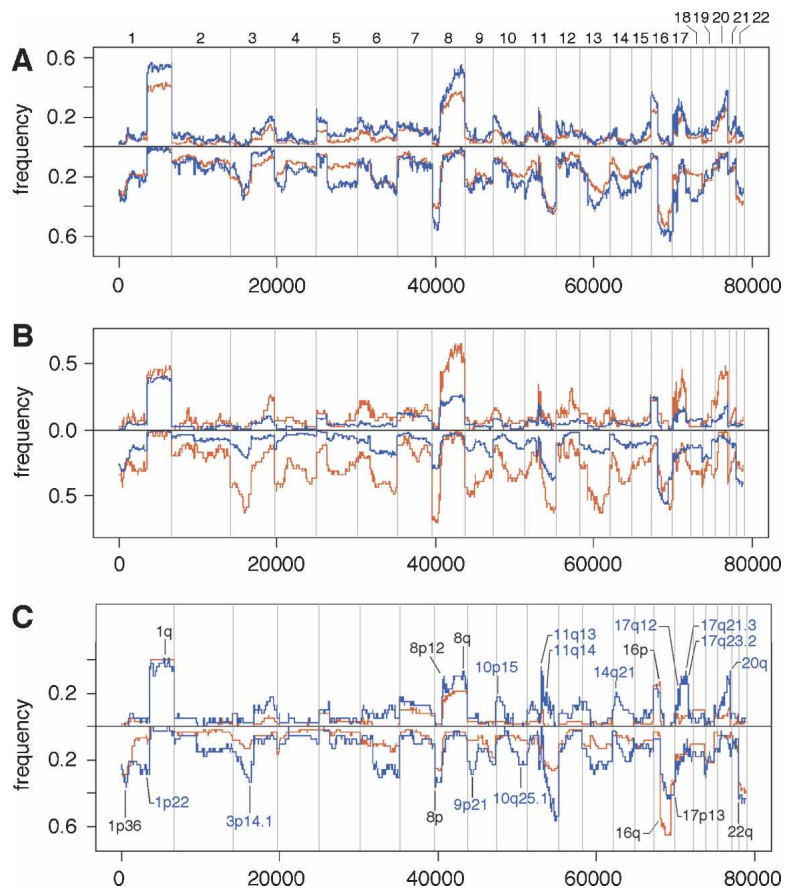


Figure 1. Comparative frequency plots of amplification (*up*) and deletion (*down*) in various data sets. Frequency calculated on normalized, segmented ROMA profiles using a minimum of six consecutive probes identifying a segment with a minimum mean of 0.1 *above* (amplification) or *below* (deletion) baseline. Frequencies are plotted only for chromosomes 1–22. (A) Total Swedish data set (red) versus total Norwegian data set (blue). (B) Swedish diploid subset (blue) versus total Swedish aneuploid subset (red). (C) Swedish diploid 7-yr survivors (red) versus Swedish diploid 7-yr nonsurvivors (blue).

contribute to a measured segment ratio below the actual copy number. Although ratios do not directly measure copy number, differences between the median ratios of segments do reflect differences in gene copy within a given experiment. This has been extensively validated by interphase FISH (see e.g., Fig. 3A,B below).

Event frequency plots in breast cancer and their correlation with outcome

Once all the individual profiles are accumulated, they can be examined and compared as subpopulations. A straightforward, albeit simplistic, view of genome alterations is the frequency plot, a measure at each probe of the frequency with which the probe is amplified or deleted above a threshold in the genome profiles of a set of cancers. To obtain an overview of breast cancer lesions, we show plots from the Swedish group, the Norwegian group, and for the combined set, plotting amplification frequencies as above the line and deletions below (Fig. 1A). Even at this crude view, it is evident that amplifications and deletions do not occur at random throughout the genome, and regions that are amplified tend not to be deleted, and vice versa. Many of the well-known loci known to be deleted or amplified, such as *TP53*, *CDKN2A*, *MYC*, *CCND1*, and *ERBB2*, are at or near the centers of frequently altered regions. Additionally, there are frequent “peaks” and “valleys” where none of the familiar suspects are found. The data are posted at our Web site, for detailed inspection by the interested reader.

The Swedish (combined aneuploid and diploid) and Norwegian breast cancers display similar frequency profiles, with slightly higher frequencies in the Norwegian set. This discrepancy is most likely explained by the high proportion of diploid cancers in the Swedish set. While the Norwegian set is sequential and unselected, the Swedish set is >70% pseudo-diploid, selected according to our working hypothesis that diploids would provide the most information about tumor development. When we compare the diploid to aneuploid Swedish cancers (Fig. 1B), we again observe similar profiles along with a similar difference in overall frequencies. This difference is not apparent when Swedish aneuploids are compared to the Norwegian group (data not shown). Thus the two cancer types, diploid and aneuploid, share the same loci of amplification and deletion.

The decreased frequency observed in the diploid set relative to the aneuploid set can be attributed to the presence of long-term survivors in the former group. Frequency plots comparing 7-yr (long-lived) survivors to those who do not survive as long (short-lived) is shown in Figure 1C. Clearly, designating a patient as a “survivor” or “nonsurvivor” at a specific time 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. We used 7 yr as a demarcation because it reflects the point at which the rate of death from cancer in the worst prognosis group drops to near zero. For the studies described in this paper, demarcation values between 7 yr and 10 yr can be used without changing the basic conclusions. It is quite apparent that there are fewer overall events, both amplifications and deletions, in the diploid survivors. Using 25 events as a divider, we obtain the most significant association of the long-lived versus the short-lived cancer patients, with a *P*-value of 4.2×10^{-4} by Fisher’s exact test.

Patterns of genome profiles

Visual inspection of segmented profiles suggests that they come in three basic patterns (Fig. 2), which we present as qualitative

heuristic tools for distinguishing apparently distinct processes of genomic rearrangement. The first profile pattern (Fig. 2A), 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. Simplex tumors make up ~60% of the diploid data set, while the rest fall into two distinct categories of “complex” patterns. One of these complex patterns is the “sawtooth” (Fig. 2B), characterized by many narrow segments of duplication and deletion, often alternating, more or less affecting all the chromosomes. Little of the genome remains at normal copy number, yet the events typically do not involve high copy number amplification. Note that the scale of the *y*-axis in Figure 2B is identical to that in Figure 2A. It should be further noted that the X-chromosome peak is often low in sawtooth profiles (e.g., WZ15 in Fig. 2B), indicating that the X chromosome is not exempt from frequent loss in these tumors.

The third pattern (Fig. 2C) resembles the simplex type except that the cancers contain at least one localized region of clustered, relatively narrow peaks of amplification, with each cluster confined to a single chromosome arm. We denote these clusters by the descriptive term “firestorms” because we believe that the clustering of multiple amplicons on single chromosome arms reflects a concerted mechanism of repeated recombination on that arm rather than a series of independent amplification events. The high copy number of these amplicons is reflected in the scale of the *y*-axis in Figure 2C.

The two complex patterns, firestorm (25%) and sawtooth (5%), make up ~30% of the diploid tumors in this data set. We cannot perfectly classify all profiles with this system, but the patterns appear to represent genomic lesions resulting from distinctly different mechanisms, and more than one mechanism may be operant to varying degrees within any given tumor.

A fourth type is the “flat” profile, in which we observe no clear amplifications or deletions other than copy number polymorphisms and single probe events, as discussed above, and the expected difference in the sex chromosomes. These examples are few in number (14/140) and are not presented graphically here. Some may result from the analysis of biopsies comprised mostly of stroma, or some may comprise a clinically relevant set of cancers with no detectable amplifications or deletions. Performing the analyses described in this paper with or without these flat profiles does not alter our conclusions; hence, we include them in the analyses presented here.

Firestorms

We used interphase FISH to validate that segmentation is not an artifact of ROMA or statistical processing of ROMA data. Either BAC clones or probes created by primer amplification were labeled and hybridized to preparations of the same frozen tumor specimens profiled by ROMA (Methods). Probes were selected from 33 loci representing both peaks and valleys in the ROMA profile. In each case, the segmentation values were confirmed by FISH. We show here representative instances of these data for the complex pattern of amplification we call “firestorms.”

Firestorms are represented in ROMA profiles as clustered narrow peaks of elevated copy number. The pattern is limited to one or a few chromosome arms in each tumor, with the remainder of the genome remaining more or less quiet, often indistinguishable from the simplex pattern. The individual amplicons in these firestorms are separated by segments that are not amplified,

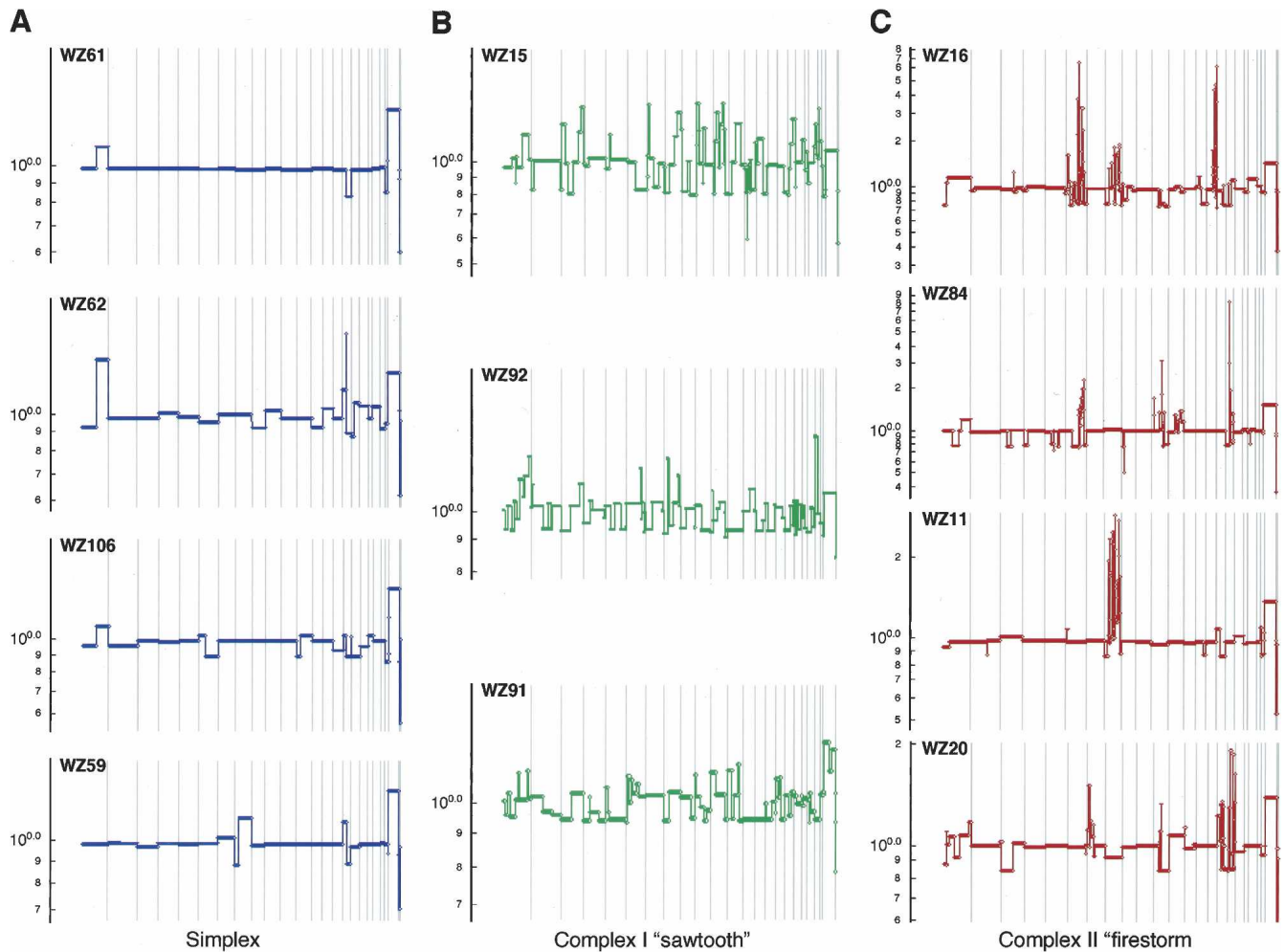


Figure 2. Major types of tumor genomic profiles. Segmentation profiles for individual tumors representing each category: (A) simplex; (B) complex type I or sawtooth; (C) complex type II or firestorm. Scored events consist of a minimum of six consecutive probes in the same state. The y-axis displays the geometric mean value of two experiments on a log scale. Note that the scale of the amplifications in C is compressed relative to A and B owing 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.

and are, in fact, often deleted, yielding a pattern of interdigitated amplification and LOH as shown for chromosome 8 (WZ11) in Figure 3A and chromosome 11q (WZ17) in Figure 3B. We infer from this that the phenomenon is a result of sequential replication and recombination events or breakage and rejoining events that occur on a particular chromosome arm rather than a general tendency toward amplification throughout the genome.

One might imagine that the individual peaks in a cluster arise from clonal subpopulations within the tumor. They do not. The FISH images of Figure 3 clearly indicate that amplifications at neighboring peaks of a cluster occur in the same cell. Moreover, they colocalize in the nucleus. In those cases in which a cell harbors two firestorms, each on different chromosomes, these too occur in the same cell, but individually segregate within the nucleus by chromosome arm, as shown in Figure 3C for *CCND1* (cyclin D1) on chromosome 11q and *ERBB2* (*HER-2/neu*) on 17q. A total of 18 BAC probes representing amplicons and intervening spaces were used in verifying the structure of chromosome 8 in WZ11 and 15 primer amplified probes were used for chromosome 11 in WZ17. Summary data for all probes are available in the Supplemental material.

Firestorms have been observed at least once on most chromosomes in the tumors we have analyzed, but certain arms clearly undergo this process more frequently (see Table 2). 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. Within the latter, the loci containing *CCND1* on 11q and *ERBB2* on 17q are most frequently amplified and may “drive” the selection of the events. Chromosomes 6, 8, and 20 have a comparable frequency of firestorms, but the “drivers” for these events are less obvious. However, these potential “driver” genes are likely not to be the sole reason for the complex amplification patterns seen in firestorms. The other peaks in the firestorms are not randomly distributed. Each chromosome appears to undergo selective pressure to gain or lose specific regions as exemplified by the frequency plot of chromosome 17 shown in Figure 4. The histogram of amplification (blue) or deletion (red) for 27 Grade II and Grade III tumors exhibiting firestorms on chromosome 17 from both Scandinavian data sets shows distinct peaks and valleys when compared to the equivalent histogram for a set of tumors of equivalent grade but without chromosome 17 firestorms (black and gray histograms). As shown in

Figure 4, there is a strong tendency for deletion of the distal p arm including *TP53* and for deletion of 17q21 including *BRCA1*. Conversely, there are at least four distinct peaks of high-frequency amplification on the long arm of 17 in addition to the peak containing *ERBB2*. As noted in the figure, several genes of interest for breast cancer are located near the epicenters of these peaks, including *TOB1* (transducer of *ERBB2*) and *BCAS3* (breast carcinoma amplified sequence). Furthermore, in contrast to accepted dogma (Jarvinen and Liu 2003), a fraction of the firestorms on 17q (5%–10%) do not include amplification of *ERBB2*, giving weight to the notion that other loci in the region may contribute to oncogenesis. In contrast, broad duplications and deletions are detectable in the non-firestorm subset, but they do not form clear peaks.

Frequently amplified and deleted loci

It is of interest to note the regions that are most frequently amplified or deleted in a large data set such as the one presented here. There is no single accepted algorithm for deciding which regions are of most interest, and the parameters used will depend on the goals of the individual researcher. In Table 3 we present the results of one such algorithm (see “Frequently Amplified and Deleted Loci” in Methods) that reflects a component of frequency at any locus plus a factor that gives weight to the inverse of the width of any given event. The latter is based on the rationale that narrow events centered on a given locus should carry more weight than a broad event that happens to encompass that locus. In the table, the relative value for each locus is shown in the Index column. Representative genes that have some potential relation to breast cancer are included for reference purposes, but we do not presume knowledge of the direct involvement of specific genes in tumorigenesis based on this analysis. While several specific amplicons have been reported previously for specific chromosomes, such as 11q (Ormandy et al. 2003) and the *ERBB2* region of 17q (Jarvinen and Liu 2003), we know of no other report cataloging a data set of comparable size and resolution permitting this level of detailed analysis. For example, Ormandy et al. (2003) report three narrow (<2 Mb) “core” amplicons in the 11q13 bands along with an independent 17-Mb amplicon spanning the other three. Our analysis yields roughly equivalent peaks of high significance (index value) at 11q13.3 and 13.4 in agreement with their data, along with at least 11 additional distinct peaks where repeated amplification events have occurred on that arm. A graphical version of this analysis is available in the Supplemental material.

Rearrangements in Grade I tumors

Tumors in which the cells maintain their differentiation as shown by histological examination are generally considered to be less aggressive and to have a good prognosis irrespective of migration to the lymph nodes. Ten examples of these so-called Grade I tumors were available from the Swedish samples and 13 from the Norwegian collection, including eight in which one or more nodes were affected. A single noninvasive DCIS (ductal carcinoma in situ) sample (MicMa245) was also present in the Norwegian set. All of the Swedish samples were medium to large tumors between 20 and 30 mm in size, while the Norwegian samples ranged from 0.5 to 25 mm.

Although the number of samples is small, the similarity in ROMA profiles among the 13 representative samples depicted in

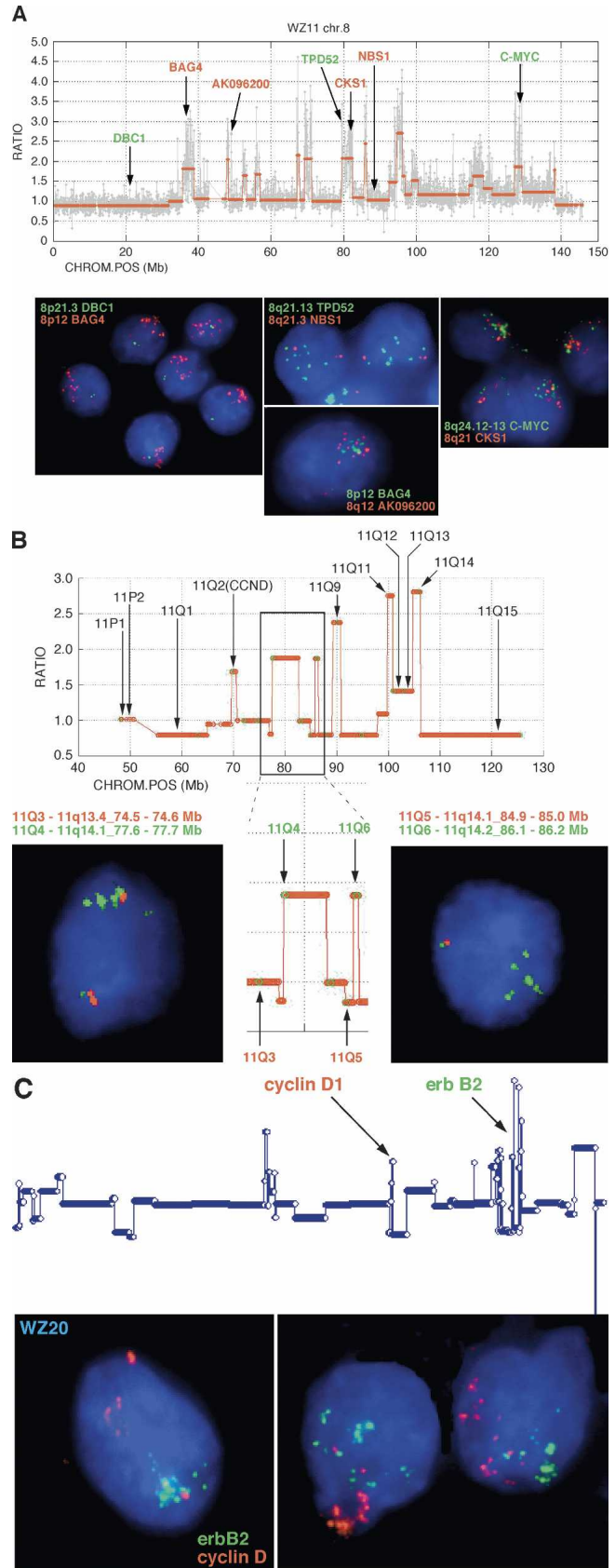


Figure 3. (Legend on next page)

Table 2. Occurrence of firestorms in the complete Swedish tumor set including both aneuploids and diploids, by chromosome arm, excluding X and Y

| chromosome arm | 1p/q | 2p/q | 3p/q | 4p/q | 5p/q | 6p/q | 7p/q | 8p/q | 9p/q | 10p/q | 11p/q |
|----------------|-------|------|------|------|-------|-------|-------|-------|-------|-------|-------|
| Firestorms | 2/3 | 0/3 | 0/1 | 0/0 | 2/0 | 3/8 | 1/1 | 6/8 | 0/0 | 0/3 | 1/16 |
| chromosome arm | 12p/q | 13q | 14q | 15q | 16p/q | 17p/q | 18p/q | 19p/q | 20p/q | 21q | 22q |
| Firestorms | 3/3 | 4 | 2 | 4 | 0/1 | 0/16 | 0/0 | 3/3 | 1/7 | 0 | 0 |

Firestorms are defined as three segmented events of any width over a threshold ratio of 0.1 on a single arm.

Figure 5 is dramatic and may provide insight into some of the earliest events leading to invasive breast cancer. Four of the 23 Grade I samples yielded no detectable events (data not shown). Eighteen of the 19 tumors with any detectable events showed a characteristic rearrangement in chromosome 16 in which one copy of 16q appears to be deleted (assuming diploidy) and 16p is concomitantly duplicated. This rearrangement was also present in the DCIS sample (MicMa245 in Fig. 5B). The rearrangement of chromosome 16 is often coupled with either a converse rearrangement of the arms of chromosome 8 (8p deleted and 8q duplicated) 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 2C and, in fact, are among the most common events by frequency in all samples (see Fig. 1B).

Grade I tumors generally display relatively few genomic events but rarely show more complex patterns of advanced simplex tumors (see MicMa171 in Fig. 5B), indicating that despite a strong correspondence, there is not a strict relation between genomic state and histological grade. MicMa171 has progressed to the point of achieving the common amplicons at 8p12 (Garcia et al. 2005) and 17q11.2, both of which are noted in Table 3. The sole Grade I tumor not showing rearrangement of 16p/q (WZ43 in Fig. 5B) exhibits a different pattern with rearrangements of chromosome 20q and deletion of 22q, indicating that the 16p/q rearrangement is not the only pathway to tumorigenesis. Although certain of these rearrangements contain obvious candidate driver genes such as the duplication of *MYC* on 8q24 or the loss of the cadherin (*CDH*) complex on 16q, the actual target genes remain the target of further study.

Relation of patterns to clinical outcome

On first inspection, the highly rearranged “sawtooth” and “firestorm” patterns appeared to correlate with shorter survival in the diploid tumors, presumably because selection of novel genetic combinations afforded the cancer cells the opportunity for accelerated recombination. We sought to confirm this observation by rigorous mathematical and statistical analysis. Using the total number of segments, or events, as a measure does not clearly distinguish a sample with a single firestorm from the simplex pattern with a similar number of events, but the effects of the firestorm on survival are clear. We chose a mathematical measure that would separate the sawtooth and firestorm patterns from the flat and simplex patterns by scoring the close-packed spacing of the firestorm events, while at the same time incorporating the total number of events. The sum of the reciprocals of the mean of lengths of all adjacent segment pairs accomplishes this goal:

$$F = \sum_i \frac{2}{l_i^L + l_i^R} \quad (1)$$

where i enumerates all the discontinuities with a magnitude above a numerical threshold of 0.1 in the segmented profile, and where l_i^L (l_i^R) denotes the number of probes in the closest neighboring discontinuity on the right (left), or to a chromosome boundary, whichever is closer. We call this the “inverse adjacent segment length measure.” This calculation is performed after masking for CNPs, and does not include the X or Y chromosomes. The measure works equally well if absolute position in the genome is substituted for probe number. Using this algorithm, the sawtooth patterns achieve a high F because of the sheer num-

Figure 3. Validation of peaks and valleys in ROMA profiles by interphase FISH. (A) Expanded ROMA profile of a firestorm on chromosome 8 in the diploid tumor WZ11. The graph shows the normalized raw data (gray) and segmented profile (red) along with the genes for which the probes shown in the FISH images were constructed. Several distinct conditions are exemplified in the images. First, the ROMA profile indicates that the 8p arm is deleted distal to the 8p12 cytoband yielding a single copy of *DBC1* (green), but >10 tightly clustered copies of *BAG4*, which is located in the frequently amplified 8p12 locus (Garcia et al. 2005). Tight clusters of multiple copies corresponding to ROMA peaks are also shown in the FISH images for *CKS1A*, *MYC*, *TPD52*, and the uncharacterized ORF AK096200. Note that the FISH signals corresponding to distinct loci cluster together irrespective of their distance on the same arm (*CKS1A/MYC*) or across the centromere (*BAG4/AK096200*). Finally, the spaces between ROMA peaks on 8q, exemplified by *NBN* (formerly known as *NBS1*), uniformly show two copies as indicated by the ROMA profile. (B) Expanded view of the centromere and 11q arm from diploid tumor WZ17 showing correspondence of the copy number as measured by FISH with the copy number predicted by the ROMA profile. The y-axis represents the segmented ratios of sample versus control. Chromosome position on the x-axis is in megabases according to Freeze 15 (April 2003) on the UCSC Genome Browser (Karolchik et al. 2003). FISH probes were amplified from primers identified from specific loci using PROBER software (Methods). The insert outlined in black is magnified to show specific details. Comparative data for the probes shown in black are not shown but are available on our Web site. In the boxed region, note that in the nonamplified regions the ROMA profile predicts two copies of the arm proximal to the leftmost amplification. Consistent with the profile, the FISH image shows two copies of probe 11Q3, with one of the spots located in the cluster along with the amplified copies. The amplicon to the right yields four copies by FISH (probe 11Q4). The ROMA profile for the amplicon represented by probe 11Q6 suggests that it is in a region in which the surrounding nonamplified portion of the arm is deleted. This arrangement is commonly observed in firestorms and is confirmed by the FISH image showing one pair of the loci 11q5 and 11Q6 together, representing the intact arm, and no copy of probe 11Q5 in the amplified cluster of spots for 11Q6. (C) Profile of tumor WZ19 in which two firestorms are observed on chromosomes 11q and 17q. In contrast to the overlapping clusters shown in A, amplifications on unrelated arms visualized using FISH probes for *CCND1* and *ERBB2* cluster independently in the nucleus.

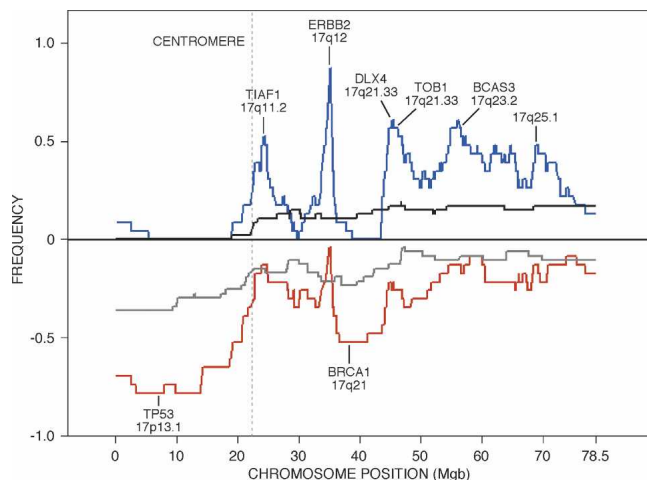


Figure 4. Frequency plots of amplification and deletions in tumors containing clustered amplifications (firestorms) on chromosome 17. Lines represent histograms of the number of events for each probe in segmented ROMA profiles over threshold as in Figure 1 for two subsets extracted from the combined Scandinavian data set. Blue and red lines represent amplifications and deletions, respectively, in the subset of 23 tumors containing firestorms on chromosome 17, each showing clear peaks (valleys) of activity. Black and gray lines represent equivalent events in a set of 53 tumors in which firestorms are not observed on chromosome 17.

ber of distributed events, while the firestorm patterns achieve high F -values even if only a single arm is affected because of the contribution of proximity (see WZ11 in Fig. 2C).

F is a robust measure separating the diploid cancers into two populations that have different survival rates. F ranges in value from zero to a maximum of -0.86 for the Swedish diploid group. For a range of values of F , from 0.08 to 0.1 we find both a significant and strong association between the discriminant value and survival beyond 7 yr. The optimum value for F separating by survival does not change appreciably when calculated for survival at 10 yr. As shown in Table 4, 0.08 and 0.09 yield the lowest P -values (2.8×10^{-7} and 5.9×10^{-7} by Fisher's exact test), with 0.09 showing the strongest association with the long-lived versus the short-lived cancer patients, with an odds ratio of 0.07. Analysis was performed using the `fisher.test` function in the R data analysis software, which computes an estimate of the odds ratio for a 2×2 contingency table using the conditional maximum likelihood estimate. In contrast, the divider based solely on the number of events without regard to size or proximity has a lower significance, with a P -value of 4.2×10^{-4} .

A strong association between F and survival is also found using an alternative statistical procedure that makes no explicit reference either to a particular discriminant value of F or to a particular survival time threshold: We divide the Swedish diploid set into quartiles with respect to F , then apply a log-rank test for differences in survival in these four groups. The four groups are found to have different survival properties, with a P -value of 10^{-7} . In Figure 6A, we display the Kaplan-Meier plots of survival for all Swedish diploids, with a range of discriminant values for F from 0.08 to 0.1. These plots show dramatically different rates of survival for tumors above or below the F -discriminant (F_d). The discriminatory power of F with respect to survival is even more dramatic when node-positive and node-negative cases are plotted separately as in Figure 6B, using $F = 0.09$.

While we find association between F and survival, we find

no significant association between F and either tumor size, lymph node status, grade, or expression of the estrogen (ER) and progesterone (PR) receptors (see Table 4; Methods). In other words, F is an independent clinical parameter. This result does not imply that these other parameters do not predict disease recurrence, or that in a random accrual F would not associate with them. Rather, it reflects that our two groups of diploids, short-term and long-term survivors, were picked to be balanced for lymph node status, tumor size, and so forth, and that F has predictive value independent of these traditional clinical measures. We do find significant association between F , on the one hand, and age at diagnosis and amplifications of the *CCND1*, *MYC*, and *ERBB2* loci, on the other hand. However, as we show in the following, F retains its predictive value for survival after adjustment for the effects of these four factors.

To further study the effect of F on survival, we fit our data to a Cox proportional hazards model, starting with a 63-case subset of the Swedish diploid data set for which we have complete information on all the clinical parameters listed in Table 4. A clinical parameter is considered significant for survival if the corresponding P -value is below 0.05. As shown in Table 5, we perform several rounds of analysis, each time removing from consideration clinical parameters not found significant in the previous round. This reduction in the number of parameters, in turn, allows us to increase the data set for which the information on the remaining parameters is complete. As a result, we find that F and the age at diagnosis are the only covariates that remain statistically significant through all the rounds of analysis. A fit to the entire Swedish diploid data set gives 4.4 as a hazard ratio for F , adjusted for the age at diagnosis.

Discussion

To the best of our knowledge, this study represents the first large sample set of primary breast tumors profiled for copy number at a resolution of <50 kb, and using a set of probes designed specifically to cover the genome evenly without regard to gene position. Coupled with a segmentation algorithm that accurately reflects event boundaries, this design has allowed us to examine genome rearrangements in tumors at an unprecedented level of detail. At this resolution, narrow and closely spaced amplifications and deletions, some as narrow as 100 kb, are clearly distinguished, and can be validated as discrete events by interphase FISH.

Cataloging the events observed in these tumor sets has allowed us to create a high-resolution map of the regions most frequently affected in this collection of tumors as compiled in Table 3. Furthermore, examination of the ROMA patterns has led us to discern three distinct profile types, described as simplex, sawtooth, and firestorm, that provide insights into the natural history of tumor development and, moreover, provide prognostic and predictive information that may be of use in clinical practice.

ROMA profiles

Each of the three characteristic profiles shown by example in Figure 2 provides a different insight into the biology of primary breast tumors. Simplex profiles are characterized by multiple duplications and deletions of whole chromosomes or chromosome arms. Moreover, certain specific chromosome arm gains and losses are highly favored, and at least a subset appears in nearly

all simplex tumors, even those low-grade tumors with less than three total events (Fig. 5). These lesions, all of which have been reported elsewhere by various methods (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. Each of these shows high frequency in the set of diploid tumors (Fig. 1B). Not all of the events occur together in the same tumor, and there is 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) with subsets of tumors that have poor survival (and many more total events) (Fig. 1B). 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.

Comparing Figure 2, A and C, it is apparent that the complex firestorm profiles display a spectrum of whole arm events reminiscent of the simplex profiles, but with the notable difference that certain chromosomes are covered almost completely with high copy number, closely spaced amplicons. We call these features firestorms because they must be the result of violent disruptions of at least one homolog, probably involving multiple rounds of breakage, copying, and rejoining to form chains of many copies (up to 30 copies in some cases, as measured by FISH). The copies apparently remain contiguous since in all cases tested, FISH results indicate that the copies fall in tight clusters within the nucleus.

Firestorms might arise through one or more previously characterized genetic mechanisms that have been previously characterized in cultured cells, such as breaks at fragile sites (Coquelle et al. 1997; Hellman et al. 2002) or recombination at pre-existing palindromic sites (Tanaka et al. 2005), perhaps by shortened telomeres. Initial joining of chromatids or chromosomes can lead to breakage-fusion-bridge (BFB) processes first described by McClintock (1938, 1941). The process of chromatid fusion and bridge formation is often seen in tumor cells (Gisselsson et al. 2000; Shuster et al. 2000) and has the potential to result in repeated rounds of segmental amplification while remaining limited to a

single arm as we have documented for firestorm events. This in itself might be a mechanism for genetic instability that augurs poor outcome, for example, by enabling the cancer cell to “search” locally for combinations of genes that by amplification or deletion promote resistance to natural controls on cell growth, invasion, or metastasis.

Finally, the alternative complex pattern, which we call sawtooth, demonstrates the operation of a path to complex genomic alteration distinct from that leading to firestorms. In contrast to firestorms, the sawtooth pattern consists of up to 30 duplication or deletion events, mostly involving chromosomal segments significantly broader than firestorm amplicons and distributed nearly evenly across the genome. Sawtooth profiles seldom show high copy number amplification as noted by the difference in the *y*-axis scale between Figure 2, A and B, versus Figure 2C. Sawtooth

Table 3. Loci that undergo frequent amplification or deletion among members of the Swedish diploid tumor set

| Chromosome position | Band | Gene symbol | Index | miRNA |
|--------------------------------|--------|---------------------|-------|------------|
| Amplifications | | | | |
| Chr1: 142,883,026–145,311,463 | q21.1 | Various | 0.05 | |
| Chr3: 157,052,165–157,422,481 | q25.1 | <i>GMP5</i> | 0.11 | |
| Chr3: 197,059,401–199,326,099 | q29 | Various | 0.03 | |
| Chr4: 9,799,463–10,002,778 | p16.1 | None | 0.06 | |
| Chr5: 142,399–980,973 | p15.33 | Various | 0.05 | |
| Chr6: 15,331,503–16,100,229 | p22.3 | <i>JARID2</i> | 0.07 | |
| Chr6: 116,304,898–116,752,141 | q22.1 | <i>FRK</i> | 0.14 | |
| Chr6: 144,141,338–144,778,980 | q24.2 | <i>PLAGL1</i> | 0.07 | |
| Chr6: 151,805,890–152,531,243 | q25.1 | <i>ESR1</i> | 0.07 | |
| Chr7: 54,880,176–56,021,876 | p11.2 | <i>EGFR</i> | 0.07 | |
| Chr7: 81,363,861–81,906,266 | q21.11 | <i>CACNA2D1</i> | 0.065 | |
| Chr8: 31,389,288–32,073,293 | p12 | <i>NRG1</i> | 0.06 | |
| Chr8: 37,655,817–38,111,519 | p12 | <i>GPR124</i> | 0.17 | |
| Chr8: 48,351,903–48,797,073 | q11.21 | Unknown | 0.10 | |
| Chr8: 56,119,985–57,277,665 | q11.21 | <i>LYN</i> | 0.08 | |
| Chr8: 67,551,628–68,252,014 | q13.1 | Various | 0.08 | |
| Chr8: 95,078,426–96,623,917 | q22.1 | <i>CCNE2</i> | 0.08 | |
| Chr8: 127,391,153–127,771,453 | q24.21 | <i>FAM84B</i> | 0.07 | |
| Chr8: 128,345,346–129,528,851 | q24.21 | <i>MYC</i> | 0.06 | |
| Chr8: 138,413,221–138,669,893 | q24.23 | None | 0.10 | |
| Chr11: 50,335,199–56,087,807 | p11.2 | Olfactory receptors | 0.05 | |
| Chr11: 56,481,254–56,801,992 | q11.2 | <i>AGTRL1</i> | 0.08 | |
| Chr11: 57,968,971–58,155,437 | q12.1 | <i>LPXN</i> | 0.13 | |
| Chr11: 68,970,345–69,253,791 | q13.3 | <i>CCND1</i> | 0.35 | |
| Chr11: 69,301,635–69,776,764 | q13.3 | <i>FGF3</i> | 0.40 | |
| Chr11: 73,028,223–73,740,133 | q13.4 | <i>RAB6A</i> | 0.10 | |
| Chr11: 77,019,036–77,608,921 | q13.5 | <i>RSF1</i> | 0.19 | |
| Chr11: 78,852,218–79,294,501 | q14.1 | None | 0.09 | |
| Chr11: 82,552,236–83,111,027 | q14.1 | <i>DLG2</i> | 0.07 | |
| Chr11: 89,502,943–90,173,207 | q14.3 | Various | 0.09 | |
| Chr11: 92,206,944–92,432,032 | q21 | <i>FAT3</i> | 0.14 | |
| Chr11: 101,466,054–101,665,638 | q22.2 | <i>YAP1</i> | 0.10 | |
| Chr11: 105,134,747–105,674,579 | q22.3 | Various | 0.10 | |
| Chr11: 115,891,412–116,980,657 | q23.3 | Various | 0.05 | |
| Chr16: 15,064,442–16,759,687 | p13.11 | Various | 0.10 | mir-484 |
| Chr16: 32,082,910–33,715,287 | p11.2 | Various | 0.12 | |
| Chr16: 59,205,171–59,350,595 | q21 | None | 0.10 | |
| Chr17: 14,364,446–14,766,493 | p12 | Unknown | 0.09 | |
| Chr17: 20,559,616–21,208,425 | p11.2 | <i>MAP2K3</i> | 0.11 | |
| Chr17: 26,949,618–27,884,440 | q11.2 | Various | 0.10 | |
| Chr17: 34,786,206–35,245,713 | q21.1 | <i>ERBB2</i> | 0.18 | |
| Chr17: 44,669,729–45,499,914 | q21.32 | Various | 0.08 | |
| Chr17: 55,947,700–56,583,137 | q23.2 | <i>BCAS3</i> | 0.10 | |
| Chr20: 51,026,986–51,790,932 | q13.2 | <i>C20orf17</i> | 0.09 | |
| Chr20: 53,727,067–54,179,752 | q13.31 | <i>CBLN4</i> | 0.10 | |
| Chr20: 59,642,719–60,188,470 | q13.33 | <i>TAF4</i> | 0.14 | |
| Chr20: 60,787,319–62,306,895 | q13.33 | Various | 0.14 | mir-124a-3 |
| Chr21: 44,323,591–46,865,905 | q22.3 | Various | 0.06 | |

(continued)

profiles, like firestorms, are associated with a poor prognosis, but their relatively high *F* index comes from the sheer number of events rather than the close spacing of the amplicons in firestorms. Taken together, these differences indicate that a genome-wide instability has been established in these tumors, perhaps distinguishing a distinct ontogeny and pathway toward metastasis.

The “Firestorm Index”

The high resolution of the ROMA technique along with our segmentation algorithm has enabled us to visualize narrow and closely spaced chromosomal rearrangements, in particular, those that make up the complex firestorm patterns. The validity of the amplicon assignments, and hence of the Kolmogorov-Smirnov methodology, has been validated by FISH in all cases tested. Coupled with the long-term survival and ploidy data available for the Swedish data set, we derived a working hypothesis consistent with previously reported work (Al-Kuraya et al. 2004; Loo et al. 2004) that complexity of rearrangement is a negative prognostic factor, but with the novel addition that the closely spaced events in firestorms make a disproportionately large contribution to that prognosis.

We have, therefore, derived a molecular signature, *F*, that correlates with survival in a subset of tumors, namely, pseudodiploid tumors of patients from Scandinavia. The signature is a

simply defined mathematical measure that incorporates two features of the genome copy number profile, namely, the number of distinguishable amplification and deletion segments, and the close packing of these segments. It is easy to imagine that the number of distinguishable events can serve as a marker for malignant “progression.” A large number of events might reflect either an unstable genome, a cancer that has been growing for a longer time within the patient and hence has had more opportunity to metastasize, or a cancer that has undergone more selective events than a cancer with fewer “scars” in its genome. It is worth noting that even a single case of the clustered amplifications that we call firestorms appears to be a prognostic indicator of poor outcome.

Our preliminary analyses of this selected sample set indicate that prognoses in primary breast cancer, measured by the probability of overall survival, are correlated with the morphology of the gene copy number signature. Within the balanced group of our samples, the magnitude of the signature is independent of such established clinical markers as node status, histologic grade, and primary tumor size. Hence, it is reasonable to expect that the signature will contribute to the prediction of outcome, perhaps—as suggested by our data—in combination with other known factors. A particularly valuable role for the signature may be in the estimation of survival for patients with ostensibly good prognosis, node-negative breast cancer, a group that may or may not

benefit from systemic therapy. A clear potential application of such a measure is in the determination of prognosis, with a focus on the identification of patients with such excellent prognoses that systemic therapy is not required or, conversely, such poor prognoses—in spite of clinical measurements that might be misleading in this regard—that systemic treatment is absolutely indicated. For example, a patient with a small, estrogen-receptor-positive, node-negative primary breast cancer—all factors that usually indicate a good prognosis—might have an especially poor prognosis as predicted by our method. Further work with unselected sample sets will, of course, be required to extend these findings beyond the working hypothesis stage.

Event mapping

We expect further gains in outcome prediction that uses knowledge of which individual loci are amplified or deleted in a specific cancer. Indeed, there are clearly loci, such as 1q, 8p and 8q, 16p and 16q, and 22q that are present in both outcome groups with almost equal frequency, and others, such as 1p12–13, 11q12 and 11q13, 9p, 10q, 17q, and 20q that are present predominantly in the cancers from patients with poor outcomes. We can improve the separation of the two groups in our own data set by adding rules that proscribe amplification or deletion at specific loci or combina-

Table 3. Continued

| Chromosome position | Band | Gene symbol | Index | miRNA |
|--------------------------------|--------|------------------|-------|-----------------|
| Deletions | | | | |
| Chr1: 13,706,706–14,067,130 | p36.21 | <i>PRDM2</i> | 0.09 | |
| Chr1: 117,882,599–118,416,501 | p12 | <i>WDR3</i> | 0.06 | |
| Chr1: 145,686,817–146,572,267 | q21.1 | Various | 0.17 | |
| Chr3: 63,833,723–69,246,170 | p14.1 | Various | 0.02 | |
| Chr3: 112,531,083–113,299,667 | q13.3 | Various | 0.07 | |
| Chr4: 4,307–2,356,621 | p16.3 | Various | 0.07 | |
| Chr5: 105,121,999–105,651,166 | q21.3 | None | 0.07 | |
| Chr6: 108,995,171–109,511,112 | q21 | <i>FOXO3A</i> | 0.08 | |
| Chr7: 153,286–2,760,544 | p22.3 | Various | 0.08 | |
| Chr8: 6,644,897–7,789,182 | p23.1 | Various | 0.07 | |
| Chr9: 21,534,743–22,602,390 | p21.3 | <i>CDKN2A</i> | 0.07 | |
| Chr11: 56,865,377–57,532,499 | q12.1 | <i>CTNND1</i> | 0.08 | mir-130a |
| Chr11: 71,383,633–71,895,665 | q13.5 | Various | 0.07 | |
| Chr11: 84,354,772–84,783,036 | q14.1 | Unknown | 0.07 | |
| Chr11: 117,818,491–119,647,340 | q23.3 | Various | 0.05 | |
| Chr12: 129,600,721–132,216,957 | q24.33 | Various | 0.04 | |
| Chr13: 31,797,266–33,180,891 | q13.1 | <i>BRCA2</i> | 0.04 | |
| Chr13: 87,304,169–88,578,303 | q31.2 | None | 0.04 | |
| Chr14: 18,212,915–19,603,016 | q11.1 | <i>ACTBL1</i> | 0.07 | |
| Chr14: 93,367,136–94,452,890 | q32.13 | Various | 0.05 | |
| Chr15: 89,220,113–89,661,514 | q26.1 | Various | 0.06 | |
| Chr15: 99,340,489–100,206,128 | q26.3 | Various | 0.05 | |
| Chr16: 59,364,195–60,612,397 | q21 | <i>CDH8</i> | 0.07 | |
| Chr17: 6,584,338–9,759,236 | p13.1 | <i>TP53</i> | 0.04 | mir-195,497,324 |
| Chr17: 11,490,353–12,494,377 | p12 | <i>MAP2K4</i> | 0.06 | |
| Chr17: 14,864,271–16,460,839 | p12 | Various | 0.06 | |
| Chr17: 56,600,423–57,012,081 | q23.2 | <i>TBX2/TBX4</i> | 0.08 | |
| Chr17: 76,951,018–78,569,870 | q25.3 | Various | 0.10 | |
| Chr18: 20,839,509–21,648,403 | p11.2 | Unknown | 0.04 | |
| Chr19: 226,336–4,793,685 | p13.3 | Various | 0.08 | mir-7–3 |
| Chr20: 14,024,068–15,010,799 | q12.1 | <i>FLRT3</i> | 0.04 | |
| Chr22: 14,858,033–20,363,383 | q11.1 | Various | 0.05 | mir-185,130b |
| Chr22: 25,251,830–27,941,420 | q12.1 | <i>CHEK2</i> | 0.05 | |
| Chr22: 31,255,407–32,147,191 | q12.3 | <i>TIMP3</i> | 0.05 | |
| Chr22: 41,881,035–42,584,718 | q13.2 | <i>SCUBE1</i> | 0.11 | |

The Index represents a relative measure that combines frequency and the inverse width of the amplicon or deletion (Methods). Loci in the table were selected to have an index of 0.05 or greater.

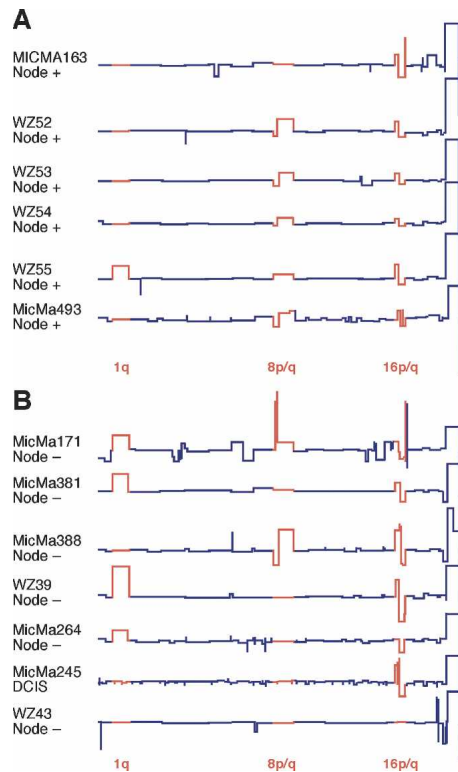


Figure 5. Comparison of Grade I and DCIS tumors by ROMA. Segmented ROMA profiles of six node-positive (Fig. 5A) and seven node-negative (Fig. 5B) Grade I or DCIS tumors, representing a total of 24 examples from the combined Swedish and Norwegian collections. Most frequent rearrangements are depicted in red.

tions of loci. However, despite exhaustive attempts, we could not convince ourselves that additional improvement in outcome prediction based on knowledge of specific loci was more than one would expect by chance, given overall event frequencies. The literature does contain many reports that specific amplifications or deletions correlate with poor prognosis (Berns et al. 1995; Jarvinen and Liu 2003; Al Kuraya et al. 2004; Chunder et al. 2004; Knoop et al. 2005; Madjd et al. 2005). While these reports may, indeed, be correct, they may also be a consequence of the larger picture, namely, that there are more lesions in “progressed” cancers. The copy numbers of specific genes may also be useful in

clinical decision-making, following the clear demonstration that *ERBB2* amplification—now determined by FISH—conveys both prognostic and therapeutic information. For example, patients with amplified *ERBB2*, as determined by FISH, are now treated with Herceptin. This determination can be made as well by ROMA or other methods for genome profiling, and such profiling may be more informative about which patients have amplifications and which benefit from such treatment. Other events in the genome can also indicate different choices of therapy. For example, two of the patients in our study exhibit amplification at the *EGFR* locus rather than *ERBB2*, and such patients might benefit from treatment with drugs targeted to that oncogene such as Tarceva. There are other such examples in the data set. More data than we now have will be needed to fully test a better outcome predictor model based on specific loci.

Scandinavian tumor sets

In the course of this study, and to gain a perspective, we have compared ROMA profiles from two independent sets of tumors from Sweden and Norway, and shown a basic similarity in the profiles independent of source or collection method. It is noteworthy that the diploid tumors with poor outcome show a very similar overall profile to the aneuploid tumors. Thus, whether or not the two classes of tumors, diploid and aneuploid, have different mechanisms for malignant genome evolution, a subset of loci occurred in amplifications and deletions in both types.

It is perhaps not surprising that the tumors from Swedish and Norwegian populations selected for this study have very similar frequency profiles, given the ethnic and environmental homogeneity in Scandinavia. It is unclear to us at the moment whether these populations will show similarity to other breast tumor sample sets. In any event, the ability to profile cancers from populations of restricted ethnicity and environment adds a new tool for those who wish to study the effects of genetics and environment on cancer. It will be of great interest to assess genome profiles of other geographically defined groups, with particular attention to the possibility of inherited patterns of disease susceptibility or gene–environment interactions.

Future directions

In this study, we have focused on a restricted question, the relationship between complex genomic rearrangements and tumor progression as determined by eventual outcome in breast cancer.

Table 4. Association of clinical parameters with the F_d measure in the Swedish diploid subset

| F_d value | Clinical parameter | Discriminating principle | P -value from Fisher's exact test | Odds ratio |
|-------------|----------------------------|---|-------------------------------------|------------|
| 0.08 | Survival | Above or below 7 yr | 2.8×10^{-7} | 0.073 |
| 0.09 | Survival | Above or below 7 yr | 5.9×10^{-7} | 0.070 |
| 0.1 | Survival | Above or below 7 yr | 8.2×10^{-6} | 0.073 |
| 0.09 | Grade | 2 vs. 3 | 0.39 | 0.58 |
| 0.09 | Node condition | Negative or positive | 1.0 | 0.96 |
| 0.09 | Size | Smaller or larger than 29 mm | 0.38 | 0.62 |
| 0.09 | ER status | Above or below 0.05 fg/ μ g protein | 0.73 | 0.77 |
| 0.09 | PR status | Above or below 0.05 fg/ μ g protein | 0.75 | 0.70 |
| 0.09 | <i>ERBB2</i> amplification | Above or below segment threshold | 0.0010 | 0.12 |
| 0.09 | <i>CCND1</i> amplification | Above or below segment threshold | 8.3×10^{-4} | 0.11 |
| 0.09 | <i>MYC</i> amplification | Above or below segment threshold | 0.0020 | 0.20 |
| 0.09 | Age at diagnosis | Above or below 57 yr | 0.0066 | 0.26 |
| 0.09 | Adjuvant therapy | –/+ | 0.44 | 0.64 |
| 0.09 | Radiation therapy | –/+ | 1.0 | 1.1 |

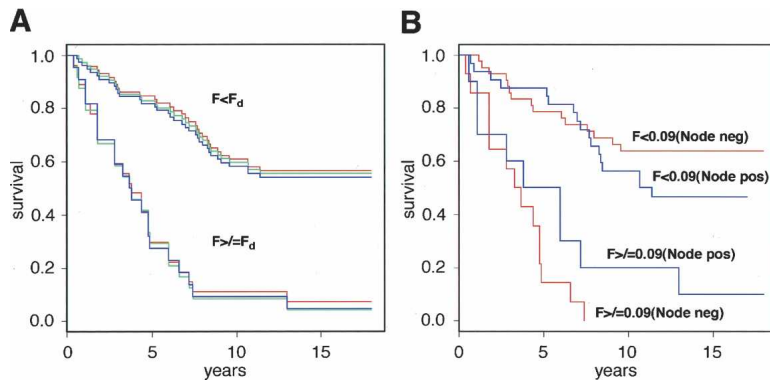


Figure 6. Kaplan-Meier plots of the Swedish diploid subset grouped according to the Firestorm Index (*F*). (A) Complete Swedish diploid data set grouped according to three different discriminator settings (F_d) of *F*: $F_d = 0.08$ (red); $F_d = 0.09$ (blue); $F_d = 0.1$ (green). (B) Swedish diploid data set separated into node-negative (red) and node-positive (blue) subsets with F_d set to 0.09.

There are many other interesting questions that we do not address in the present paper. We do not examine the related question of genomic and molecular markers for survival among aneuploid cancers. We have not analyzed what the collective profiles teach us about the location of candidate oncogenes and tumor suppressors. The latter is a deceptively complex problem that we will address subsequently. In the meantime, we post our genome profiles and associated data on our Web site (<http://roma.cshl.edu>) for others to explore. It is evident from even superficial inspection that many recurrent events encompass known oncogenes (such as *ERBB2*, *CCND1*, *MYC*) and tumor suppressors (such as *CDKN2A* and *TP53*), but many do not, such as a commonly amplified and very narrow region at 8p12, for which the driver gene has not been definitively identified (marked with a probe for *BAG4* in Fig. 3A; Garcia et al. 2005). We are also currently analyzing the important question of whether certain lesions show covariance.

Finally, it is becoming clear through the identification of gene copy number alterations in tumors in numerous CGH studies, that there is likely to be a genetic pathway, albeit a complex one, at work in the evolution of tumors. As the collection of tumor genomic profiles increases and can be compared with

treatment regimes as well as patient outcomes, that prognostic information regarding clinical outcome will likely become apparent. Thus existence of some systematic organization to the genomic events in these tumors raises the intriguing possibility that we may soon be able to dissect the pathways that determine the bridge from noninvasive to invasive to metastatic cancer.

Methods

Patient samples

A total of 140 frozen tumor specimens was selected from the 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. From a total of 5782 cases, analyzed for ploidy at the Division for Cellular and Molecular Pathology at the Karolinska Hospital at the time of primary diagnosis (1987–1991), 1601 pseudo-diploids were available with complete clinical information including ploidy, grade, node status, and clinical follow up for 14 to 18 yr. Of these, 4.0% or 64 cases were node-negative nonsurvivors at 7 yr, and 8.0% or 127 cases were node-positive nonsurvivors. Of these, 47 cases were locally available as frozen tissue and made up the group of node-negative and node-positive nonsurvivors. The diploid survivor group was selected from the remainder of the samples in order to match tumor size and grade.

From the Oslo Micrometastasis study (OMS) (Wiedswang et al. 2003), fresh frozen samples from the primary tumor from 103 cases were available for analyses by 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 DNA for classification as receptor positive for the Swedish samples. For the Norwegian samples, automatic immunostaining

Table 5. Multivariate analysis of clinical parameters shown in Table 3

| Clinical parameter | Discriminating principle | (P) | HR | CI | (P) | HR | CI | (P) | HR | CI | (P) | HR | CI |
|--------------------|----------------------------------|--------------------|------|----------|--------------------|-----|----------|------|-----|----------|------|-----|---------|
| <i>F</i> | Above or below 0.09 | 5×10^{-6} | 9.5 | 3.6:25.0 | 2×10^{-6} | 8.4 | 3.5:20.3 | 6e-5 | 5.3 | 2.4:12.1 | 9e-7 | 4.4 | 2.4:7.8 |
| AD | Above or below 57 yr | 6×10^{-3} | 3.0 | 1.4:6.7 | 0.02 | 2.3 | 1.2:4.5 | 0.04 | 2.3 | 1.1:5.0 | 7e-3 | 2.2 | 1.2:3.8 |
| <i>MYC</i> amp. | Above or below segment threshold | 0.02 | 0.26 | 0.08–0.8 | NS | | | | | | | | |
| ER status | +/- | NS | | | | | | | | | | | |
| PR status | +/- | NS | | | | | | | | | | | |
| Size | Above or below 29 mm | NS | | | | | | | | | | | |
| Node condition | +/- | NS | | | | | | | | | | | |
| Grade | I, II, or III | NS | | | | | | | | | | | |
| <i>ERBB2</i> amp. | Above or below segment threshold | NS | | | | | | | | | | | |
| <i>CCND1</i> amp. | Above or below segment threshold | NS | | | | | | | | | | | |

Discriminating values for AD and size were chosen to maximize their association with survival. (HR) Hazard Ratio; (CI) 95% confidence interval for HR; (NS) not significant. Columns 3 through 5: all the clinical parameters listed were used in the fit; columns 6 through 8: *F*, AD and *MYC* amp. were used in the fit; columns 9 through 14: *F* and AD were used in the fit. Results in columns 3 through 11 are based on a 63-case subset of the Swedish diploid set for which all the clinical parameters used were available. Results in columns 12 through 14 are based on the entire Swedish diploid set.

was performed using mouse monoclonal antibodies against ER and PgR (clones 6F11 and 1A6, respectively; Novocastra). Immunopositivity was recorded if $\geq 10\%$ of the tumor cell nuclei were immunostained. Amplification of the *ERBB2* gene was assessed by FISH on tissue microarray sections using the PathVysion HER-2 DNA Probe kit (Vysis Inc.).

ROMA DNA microarray analysis

ROMA was performed on a high-density oligonucleotide array containing $\sim 85,000$ features, manufactured by Nimblegen. 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 adaptor-mediated PCR of genomic DNA (Sebat et al. 2004). 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 the 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, $5 \times \text{SSC}$, and 0.1% SDS) and 10 μL of labeled DNA. Samples were denatured in an MJ Research Tetrad for 5 min at 95°C, and then pre-annealed for 30 min at 37°C. This solution was then applied to the microarray and hybridized under a coverslip for 14–16 h at 42°C. After hybridization, slides were washed for 1 min in 0.2% SDS/0.2 \times SSC, 30 sec in 0.2 \times SSC, and 30 sec in 0.05 \times 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. Log ratio values obtained from color reversal experiments were averaged and displayed as presented in the figures.

Statistics and 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, is 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 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.

Statistical analysis of segmented data was performed using R and S+ statistical languages. In particular, the R Survival package was used for survival analysis.

Masking of frequent CNPs

A large fraction of our collection of genome profiles are of a self-nonsel type, that is, a cancer genome and a reference genome originate in different individuals. As a result, not all of the relative copy number variation in the cancer genome is due to cancer: Some of it reflects copy number polymorphisms (CNPs) present in the healthy genome of the affected individual. This noncancerous signal can potentially contaminate subsequent analysis and must be filtered out. To this end, we examine our collection of ROMA profiles derived from cancer-free genomes (~ 500 cases in our most recent study). From that collection we determine the contiguous regions (here to be understood as series of consecutive ROMA probes) in the genome where CNP frequencies satisfy two conditions: (1) These frequencies are higher than certain f_c everywhere in the region; (2) these frequencies are higher than certain $f_s \geq f_c$ somewhere in the region. This determination is done separately for the amplification and for the deletion CNPs. With our present cancer-free collection, the optimal values are $f_c = 0.006$, $f_s = 0.03$. Once the mask, that is, the set of CNP-prone regions of the genome, is known, it is used for masking likely noncancerous CNPs in cancer genome profiles. Here we describe the masking algorithm for amplifications; the algorithm for deletions is completely analogous. If an amplified segment in a cancer genome profile falls entirely within a mask, a point (a probe) is selected at random in the segment, and the neighboring segments on the right and on the left are extended to that point. If one of the segment’s endpoints is at a chromosome boundary, the neighboring segment is extended from the other endpoint to the boundary. In effect, the CNPs are excised from the profile in a minimally intrusive fashion.

Frequently amplified and deleted loci

For the purpose of compiling a list of frequently amplified loci, amplification events are defined as follows. First, the logarithm of the relative copy number is computed for every segment in the genome (the segmentation method is described earlier in this section). Denote the resulting piecewise constant function $L(x)$, where x is the genome position. Next, (1) the values of $L(x)$ below a threshold t are replaced by 0. Then (2) we identify event blocks, that is, contiguous intervals of the genome such that $L(x) > 0$ everywhere within the interval. For every block, (3) an event extending over the entire block is added to the list of events. Next (4) a minimal nonzero value of $L(x)$ is found in each block, and that value is subtracted from $L(x)$ within that block. The steps (1) through (4) are iterated as long as $L(x) > 0$ anywhere in the genome. The event counting rule for deletions is completely analo-

gous, with obvious sign changes made throughout the description. We used a value of 0.1 for t in the present study. Once the events have been identified, we compute for every position in the genome an event density measure, defined as the sum of inverse lengths of all the events containing that position. We then identify positions with the highest event density in every chromosome arm.

Fluorescence in situ hybridization

FISH analysis was performed using interphase cells, and probes were prepared either from 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 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.). Denaturation of probe and target DNA was performed for 5 min at 90°C, followed by hybridization in a humidity chamber overnight at 47°C. The cover glasses were then removed, and the slides were washed in $2 \times$ SSC for 10 min at 72°C, and slides were dehydrated in graded alcohol. The slides were mounted with antifade 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 an Axio Cam MRM CCD camera and Axio Vision software.

Probe design for FISH

Hybridization probes for FISH were constructed in one of two methods. For the interdigitation analysis, probes were created from bacterial artificial chromosomes (BAC) selected using the UCSD Genome Browser. For the determination of copy number in the deletions and amplifications of the aneuploid tumors, probes were made with PCR amplification of primers identified through the PROBER algorithm designed in this laboratory (Navin et al. 2006). Genomic sequences of 100 kb containing target amplifications were tiled with 50 probes (800–1400 bp).

Oligonucleotide primers were ordered in 96-well plates from Sigma Genosys and resuspended to 25 μ M. Probes were amplified with the PCR Mastermix kit from Eppendorf (Cat. 0,032,002.447) from EBV immortalized cell line DNA (Chp-Skn-1) DNA (100 ng) with 55°C annealing, 72°C extension, 2 min extension time, and 23 cycles. Probes were purified with Qiagen PCR purification columns (Cat. 28,104) and combined into a single probe cocktail (10–25 μ g 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 photometry (Forsslund and Zetterberg 1990; Forsslund et al. 1996). The optical densities of the nuclei in a sample are measured and a DNA index is 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 G₂-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

KI samples were collected from patients undergoing radical mastectomy at the Karolinska Institutet between 1984 and 1991. This project was approved by the Ethical Committee of the Karolinska Institute, Stockholm, Sweden (772003). Samples in the OMS set were collected during 1995–1998 after informed written consent and analysis protocols approved by the Regional Committee for Research Ethics, Health Region II, Oslo, Norway (approval S97103).

Acknowledgments

This work was supported by grants to M.W. from the National Institutes of Health 5R01-CA078544-07; Department of the Army W81XWH04-1-0477; W81XWH-05-1-0068; W81XWH-04-0905; The Simons 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 0046-B05-39XBC), from the Stockholm Cancer Society (grant number 03:17), and from the Swedish Research Council (grant number K2006-31X-20081-01-3). The OMS study has been supported by the Norwegian Cancer Society. We are also grateful for critical review by Knut Liestøl, Institute for Informatics, University of Oslo and for useful comments by Xiaoyue Zhao, Cold Spring Harbor Laboratory.

References

- Ahr, A., Karn, T., Solbach, C., Seiter, T., Strebhardt, K., Holtrich, U., and Kaufmann, M. 2002. Identification of high risk breast-cancer patients by gene expression profiling. *Lancet* **359**: 131–132.
- Albertson, D.G. 2003. Profiling breast cancer by array CGH. *Breast Cancer Res. Treat.* **78**: 289–298.
- Al Kuraya, K., Schraml, P., Torhorst, J., Tapia, C., Zaharieva, B., Novotny, H., Spichtin, H., Maurer, R., Mirlacher, M., Kochli, O., et al. 2004. Prognostic relevance of gene amplifications and coamplifications in breast cancer. *Cancer Res.* **64**: 8534–8540.
- Balmain, A., Gray, J., and Ponder, B. 2003. The genetics and genomics of cancer. *Nat. Genet.* **33**: 238–244.
- Berns, E.M., de Klein, A., van Putten, W.L., van Staveren, I.L., Bootsma, A., Klijn, J.G., and Foekens, J.A. 1995. Association between RB-1 gene alterations and factors of favourable prognosis in human breast cancer, without effect on survival. *Int. J. Cancer* **64**: 140–145.
- Chunder, N., Mandal, S., Roy, A., Roychoudhury, S., and Panda, C.K. 2004. Analysis of different deleted regions in chromosome 11 and their interrelations in early- and late-onset breast tumors: Association with cyclin D1 amplification and survival. *Diagn. Mol. Pathol.* **13**: 172–182.
- 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**: 215–225.
- Daruwala, R.S., Rudra, A., Ostrer, H., Lucito, R., Wigler, M., and Mishra, B. 2004. A versatile statistical analysis algorithm to detect genome copy number variation. *Proc. Natl. Acad. Sci.* **101**: 16292–16297.
- DePinho, R.A. and Polyak, K. 2004. Cancer chromosomes in crisis. *Nat. Genet.* **36**: 932–934.
- Edén, P., Ritz, C., Rose, C., Ferno, M., and Peterson, C. 2004. “Good Old” clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. *Eur. J. Cancer* **40**: 1837–1841.
- Forsslund, G. and Zetterberg, A. 1990. Ploidy level determinations in high-grade and low-grade malignant variants of prostatic carcinoma. *Cancer Res.* **50**: 4281–4285.
- Forsslund, G., Nilsson, B., and Zetterberg, A. 1996. Near tetraploid prostate carcinoma. Methodologic and prognostic aspects. *Cancer* **78**: 1748–1755.
- Garcia, M.J., Pole, J.C., Chin, S.F., Teschendorff, A., Naderi, A., Ozdag, H., Vias, M., Kranjac, T., Subkhankulova, T., Paish, C., et al. 2005. A 1 Mb minimal amplicon at 8p11-12 in breast cancer identifies new

- candidate oncogenes. *Oncogene* **24**: 5235–5245.
- 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**: 5357–5362.
- Hellman, A., Zlotorynski, E., Scherer, S.W., Cheung, J., Vincent, J.B., Smith, D.I., Trakhtenbrot, L., and Kerem, B. 2002. A role for common fragile site induction in amplification of human oncogenes. *Cancer Cell* **1**: 89–97.
- Jarvinen, T.A. and Liu, E.T. 2003. HER-2/neu and topoisomerase II α in breast cancer. *Breast Cancer Res. Treat.* **78**: 299–311.
- 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**: 818–821.
- 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**: 190–193.
- Kallioniemi, O.P., Kallioniemi, A., Kurisu, W., Thor, A., Chen, L.C., Smith, H.S., Waldman, F.M., Pinkel, D., and Gray, J.W. 1992c. ERBB2 amplification in breast cancer analyzed by fluorescence in situ hybridization. *Proc. Natl. Acad. Sci.* **89**: 5321–5325.
- 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**: 2156–2160.
- Knoop, A.S., Knudsen, H., Balslev, E., Rasmussen, B.B., Overgaard, J., Nielsen, K.V., Schonau, A., Gunnarsdottir, K., Olsen, K.E., Mouridsen, H., et al. 2005. Retrospective analysis of topoisomerase IIa amplifications and deletions as predictive markers in primary breast cancer patients randomly assigned to cyclophosphamide, methotrexate, and fluorouracil or cyclophosphamide, epirubicin, and fluorouracil: Danish Breast Cancer Cooperative Group. *J. Clin. Oncol.* **23**: 7483–7490.
- 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**: 904–909.
- Lage, J.M., Leamon, J.H., Pejovic, T., Hamann, S., Lacey, M., Dillon, D., Segraves, R., Vossbrinck, B., Gonzalez, A., Pinkel, D., et al. 2003. Whole genome analysis of genetic alterations in small DNA samples using hyperbranched strand displacement amplification and array-CGH. *Genome Res.* **13**: 294–307.
- Loo, L.W., Grove, D.I., Williams, E.M., Neal, C.L., Cousens, L.A., Schubert, E.L., Holcomb, I.N., Massa, H.F., Glogovac, J., Li, C.I., et al. 2004. Array comparative genomic hybridization analysis of genomic alterations in breast cancer subtypes. *Cancer Res.* **64**: 8541–8549.
- Lucito, R., Healy, J., Alexander, J., Reiner, A., Esposito, D., Chi, M., Rodgers, L., Brady, A., Sebat, J., Troge, J., et al. 2003. Representational oligonucleotide microarray analysis: A high-resolution method to detect genome copy number variation. *Genome Res.* **13**: 2291–2305.
- Madjd, Z., Spendlove, I., Pinder, S.E., Ellis, I.O., and Durrant, L.G. 2005. Total loss of MHC class I is an independent indicator of good prognosis in breast cancer. *Int. J. Cancer* **117**: 248–255.
- 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**: 315–376.
- McClintock, B. 1941. The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**: 234–282.
- Menard, S., Fortis, S., Castiglioni, F., Agresti, R., and Balsari, A. 2001. HER2 as a prognostic factor in breast cancer. *Oncology* **61**: 67–72.
- Navin, N., Grubor, V., Hicks, J., Leib, E., Thomas, E., Troge, J., Riggs, M., Lundin, P., Maner, S., Sebat, J., et al. 2006. PROBER: Oligonucleotide FISH probe design software. *Bioinformatics* **22**: 2437–2438.
- Nessling, M., Richter, K., Schwaenen, C., Roerig, P., Wrobel, G., Wessendorf, S., Fritz, B., Bentz, M., Sinn, H.-P., Radwimmer, B., et al. 2005. Candidate genes in breast cancer revealed by microarray-based comparative genomic hybridization of archived tissue. *Cancer Res.* **65**: 439–447.
- Olshen, A.B., Venkatraman, E.S., Lucito, R., and Wigler, M. 2004. Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**: 557–572.
- Ormandy, C.J., Musgrove, E.A., Hui, R., Daly, R.J., and Sutherland, R.L. 2003. Cyclin D1, EMS1 and 11q13 amplification in breast cancer. *Breast Cancer Res. Treat.* **78**: 323–335.
- Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., et al. 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N. Engl. J. Med.* **351**: 2817–2826.
- Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Aksien, L.A., et al. 2000. Molecular portraits of human breast tumours. *Nature* **406**: 747–752.
- 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**: 12963–12968.
- Ried, T., Just, K.E., Holgreve-Grez, H., Du Manoir, S., Speicher, M.R., Schröck, E., Latham, C., Blegen, H., Zetterberg, A., Cremer, T., et al. 1995. Comparative genomic hybridization of formalin-fixed, paraffin-embedded breast tumors reveals different patterns of chromosomal gains and losses in fibroadenomas and diploid and aneuploid carcinomas. *Cancer Res.* **5**: 5415–5423.
- Ried, T., Liyanage, M., Du Manoir, S., Heselmeyer, K., Auer, G., Macville, M., and Schrock, E. 1997. Tumor cytogenetics revisited: Comparative genomic hybridization and spectral karyotyping. *J. Mol. Med.* **75**: 801–814.
- Sebat, J., Lakshmi, B., Troge, J., Alexander, J., Young, J., Lundin, P., Maner, S., Massa, H., Walker, M., Chi, M., et al. 2004. Large-scale copy number polymorphism in the human genome. *Science* **305**: 525–528.
- Shuster, M.I., Han, L., Le Beau, M.M., Davis, E., Sawicki, M., Lese, C.M., Park, N.H., Colicelli, J., and Gollin, S.M. 2000. A consistent pattern of RIN1 rearrangements in oral squamous cell carcinoma cell lines supports a breakage-fusion-bridge cycle model for 11q13 amplification. *Genes Chromosomes Cancer* **28**: 153–163.
- 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. 1989. Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. *Science* **244**: 707–712.
- Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.D., van de Rijn, M., Jeffrey, S.S., et al. 2001. Gene expression patterns of carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci.* **98**: 10869–10874.
- Sotiriou, C. 2003. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc. Natl. Acad. Sci.* **100**: 10393–10398.
- Tanaka, H., Bergstrom, D.A., Yao, M.C., and Tapscott, S.J. 2005. Widespread and nonrandom distribution of DNA palindromes in cancer cells provides a structural platform for subsequent gene amplification. *Nat. Genet.* **37**: 320–327.
- 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**: 177–184.
- van de Vijver, M., van de Bersselaar, R., Devilee, P., Cornelisse, C., Peterse, J., and Nusse, R. 1987. Amplification of the neu (c-erbB-2) oncogene in human mammary tumors is relatively frequent and is often accompanied by amplification of the linked c-erbA oncogene. *Mol. Cell. Biol.* **7**: 2019–2023.
- van't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., et al. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**: 530–536.
- Wiedswang, G., Borgen, E., Karesen, R., Kvalheim, G., Nesland, J.M., Qvist, H., Schlichting, E., Sauer, T., Janbu, J., Harbitz, T., et al. 2003. Detection of isolated tumor cells in bone marrow is an independent prognostic factor in breast cancer. *J. Clin. Oncol.* **21**: 3469–3478.

Received April 4, 2006; accepted in revised form September 14, 2006.