

A complete electronic version of this article and other services, including high-resolution figures, can be found at:

<http://stm.sciencemag.org/content/2/38/38ra47.full.html>

Supplementary Material can be found in the online version of this article at:

<http://stm.sciencemag.org/content/suppl/2010/06/29/2.38.38ra47.DC1.html>

Related Resources for this article can be found online at:

<http://stm.sciencemag.org/content/scitransmed/2/38/38ps32.full.html>

This article **cites 65 articles**, 19 of which can be accessed free:

<http://stm.sciencemag.org/content/2/38/38ra47.full.html#ref-list-1>

This article has been **cited by 2** articles hosted by HighWire Press; see:

<http://stm.sciencemag.org/content/2/38/38ra47.full.html#related-urls>

Information about obtaining **reprints** of this article or about obtaining **permission to reproduce this article** in whole or in part can be found at:

<http://www.sciencemag.org/about/permissions.dtl>

Genomic Architecture Characterizes Tumor Progression Paths and Fate in Breast Cancer Patients

Hege G. Russnes,^{1,2,3*} Hans Kristian Moen Vollan,^{1,3,4*} Ole Christian Lingjærde,⁵ Alexander Krasnitz,⁶ Pär Lundin,⁷ Bjørn Naume,⁸ Therese Sørli,¹ Elin Borgen,² Inga H. Rye,² Anita Langerød,¹ Suet-Feung Chin,⁹ Andrew E. Teschendorff,^{9,10} Philip J. Stephens,¹¹ Susanne Månér,⁷ Ellen Schlichting,⁴ Lars O. Baumbusch,^{1,2,5} Rolf Kåresen,⁴ Michael P. Stratton,^{11,12} Michael Wigler,⁶ Carlos Caldas,^{9,13} Anders Zetterberg,⁷ James Hicks,^{6†} Anne-Lise Børresen-Dale^{1,3†‡}

(Published 30 June 2010; Volume 2 Issue 38 38ra47)

Distinct molecular subtypes of breast carcinomas have been identified, but translation into clinical use has been limited. We have developed two platform-independent algorithms to explore genomic architectural distortion using array comparative genomic hybridization data to measure (i) whole-arm gains and losses [whole-arm aberration index (WAAI)] and (ii) complex rearrangements [complex arm aberration index (CAAI)]. By applying CAAI and WAAI to data from 595 breast cancer patients, we were able to separate the cases into eight subgroups with different distributions of genomic distortion. Within each subgroup data from expression analyses, sequencing and ploidy indicated that progression occurs along separate paths into more complex genotypes. Histological grade had prognostic impact only in the luminal-related groups, whereas the complexity identified by CAAI had an overall independent prognostic power. This study emphasizes the relation among structural genomic alterations, molecular subtype, and clinical behavior and shows that objective score of genomic complexity (CAAI) is an independent prognostic marker in breast cancer.

INTRODUCTION

Breast cancer is a heterogeneous disease as reflected by histopathology, molecular alterations, and clinical behavior. Substantial effort has been exerted toward identifying tumor groups with distinct molecular features to relate cellular and subcellular features to clinical parameters and outcome. Estrogen receptor (ER) status is a major discriminating factor of clinical importance (1). However, recent gene expression-based classifications have identified five different subgroups, where two were luminal cell-related (luminal A and B), one was myoepithelial cell-related (basal-like), one resembled normal breast tissue (normal-like), and one was erbB2-enriched (2–4). The erbB2-enriched group has frequent activation of the erbB2/

HER2 pathway and shows a high correlation to the basal-like centroid, and such tumors seem to be closely associated with the basal-like phenotype (4). Here, the erbB2-enriched and basal-like subtypes are called “basal-related tumors.” Basal-like and luminal A carcinomas have different etiologies and, for most purposes, may be considered as distinct diseases (4–7). This is also reflected in the genomic portraits defined by array comparative genomic hybridization (aCGH), and it seems that the history of molecular subgroups is inscribed in the DNA alterations (8–10). Despite the power of RNA- and DNA-based profiling, translating complex molecular classifications into clinical practice has proven to be a formidable challenge. Clinical cohorts are often selected to have tumors of a certain category and might not include all subtypes or outcome groups. The size of sample sets available for microarray studies has so far been limited, and combining sets to increase size has been challenging because various types of array platforms have been used.

aCGH does not reveal the chromosomal rearrangement patterns associated with copy number alterations; however, much can be inferred from cytogenetic studies (11). The genomic architectural changes in breast tumors revealed by karyotyping follow some main traits. One such event seen early in tumor progression is the loss or gain of whole chromosome arms (12). Alternative events involve more complex rearrangements, where several different chromosomes concomitantly undergo inversions, deletions, and amplifications (12).

Previously, we found that invasive breast tumors had different patterns of aCGH aberrations and could be grouped into three different categories: simplex, complex I “sawtooth,” and complex II “firestorm” (13). Tumors of the simplex type had few alterations, with loss or gain of whole arms dominating, whereas tumors of the complex type either had many chromosomes altered with multiple regions with low-level loss and gain (sawtooth pattern) or had a few selected regions with high copy number gains with intermittent losses (firestorms). We hy-

¹Department of Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, 0310 Oslo, Norway. ²Division of Pathology, Oslo University Hospital Radiumhospitalet, 0310 Oslo, Norway. ³Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, 0318 Oslo, Norway. ⁴Department of Breast and Endocrine Surgery, Division of Surgery and Cancer, Oslo University Hospital, 0450 Oslo, Norway. ⁵Biomedical Research Group, Department of Informatics, University of Oslo, P.O. Box 1080 Blindern, 0316 Oslo, Norway. ⁶Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA. ⁷Department of Oncology-Pathology, Karolinska Institutet, Cancer Center Karolinska, SE-171 76 Stockholm, Sweden. ⁸Department of Oncology, Division of Surgery and Cancer, Oslo University Hospital Radiumhospitalet, 0310 Oslo, Norway. ⁹Breast Cancer Functional Genomics, Cancer Research UK Cambridge Research Institute and Department of Oncology, University of Cambridge, Li Ka-Shing Centre, Robinson Way, Cambridge CB2 0RE, UK. ¹⁰UCL Cancer Institute, University College London, London WC1E 6BT, UK. ¹¹Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ¹²Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK. ¹³Cambridge Breast Unit, Addenbrookes Hospital and Cambridge National Institute for Health Research Biomedical Research Centre, Cambridge University Hospitals NHS Foundation Trust, Hills Road, Cambridge CB2 0QQ, UK.

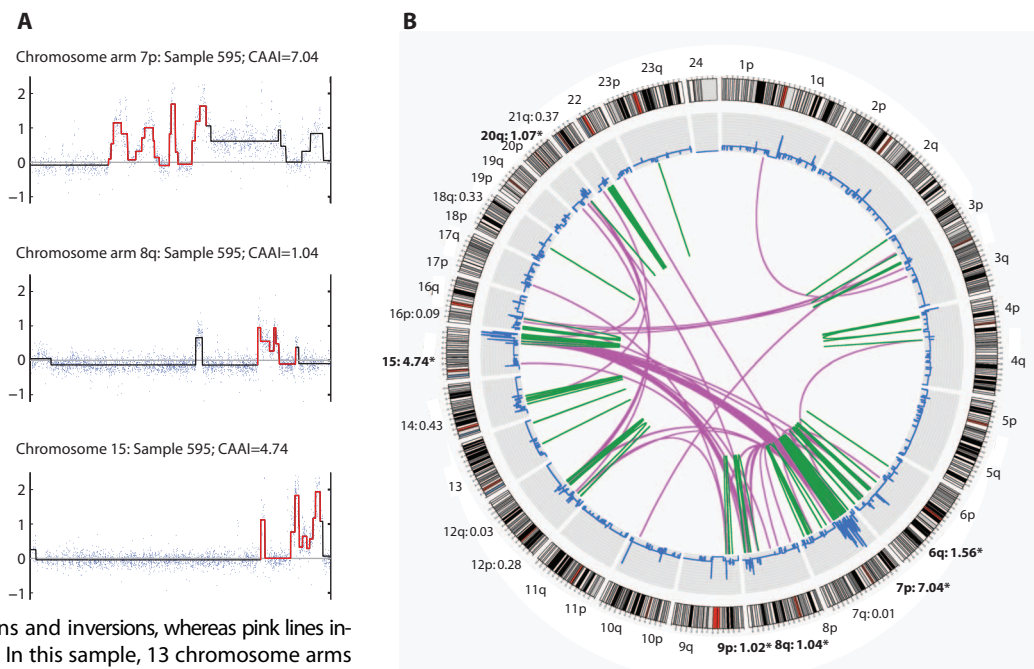
*These authors contributed equally to this work.

†These authors contributed equally to this work.

‡To whom correspondence should be addressed. E-mail: a.l.borresen-dale@medisin.uio.no

Fig. 1. CAAI values compared to structural rearrangements identified by paired-end sequencing.

(A) Raw (dots) and segmented (line) data for chromosome arms 7p and 8q and chromosome 15 from sample 595. Red segments correspond to the 20-Mb windows with highest CAAI; the corresponding CAAI was 7.04, 1.04, and 4.74, respectively. Chromosome arm 7p had an additional region with high-level CAAI, but because this score was lower than 7.04, it was not highlighted in red. (B) Structural sequence alterations identified by genome-wide paired-end sequencing for the same sample. The outer circle shows the cytobands for each chromosome, followed by a plot indicating the copy number alterations. The green bars in the center refer to smaller intrachromosomal changes such as duplications and inversions, whereas pink lines indicate interchromosomal translocations. In this sample, 13 chromosome arms had CAAI > 0; six of these had CAAI \geq 0.5 (these are in bold and marked with an asterisk). The two regions with most rearrangements showed the highest CAAI (chromosome arm 7p and chromosome 15). Areas with few rearrangements had low or zero CAAI.



pothesized that distinct molecular mechanisms underlie such patterns of aberrations. The simplex and complex (sawtooth and firestorm) classification proposed by Hicks *et al.* (13) was not obtained algorithmically; hence, no objective measure across platforms is available.

One aim of this study was to develop objective estimates of genome-wide architectural distortion. For each chromosome arm, two platform-independent scores were defined: one measures whole-arm deviations from normal copy number [whole-arm aberration index (WAAI)], and the other the degree of local distortion [complex arm aberration index (CAAI)]. Our marker of genomic complexity (CAAI) was hypothesized to have independent prognostic power in breast cancer. Our aim was to investigate this marker in a large series of breast carcinomas ($n = 595$) analyzed with aCGH and its relation to patient outcome. In addition, a semisupervised classifier was constructed on the basis of acknowledged genomic alterations in luminal A and basal-like tumors by combining information from the WAAI and CAAI estimates. Because three of the four tumor sets had extensive additional molecular data and clinical follow-up available, this approach presciently revealed distinct patterns of genomic architectural distortion associated with outcome. Thus, varying levels of genomic distortion and survival outcomes may reflect different paths of tumor progression.

RESULTS

Genomic architecture characterized by CAAI and WAAI

Two novel algorithms were constructed: one to identify complex architectural distortions characterized by physically tight clusters of break points with large changes of amplitude (CAAI), and another to recognize gains and loss of whole chromosome arms (WAAI). Seg-

mented data from one tumor with corresponding CAAI values are illustrated for selected chromosome arms (Fig. 1A). The Circos plot from paired-end sequencing of the same sample (Fig. 1B) shows that CAAI recognizes regions with structural complexity (14). Areas of complex rearrangements were found by selecting chromosome arms with CAAI \geq 0.5. Comparison in one cohort of *HER2* copy number gains estimated by fluorescence in situ hybridization (FISH) and the CAAI score showed that all but one sample with high CAAI had more than four copies of *HER2* (fig. S1).

For most chromosome arms, the distribution of WAAI was nearly symmetric around zero (fig. S2). For some arms, however, WAAI was skewed toward positive values (1q, 8q, and 16p), and for others toward negative values (16q and 17p), reflecting a bias toward gain or loss. This pattern was seen in all cohorts, independently of platform (fig. S2). Arms with WAAI \geq 0.8 were defined as whole-arm gains and arms with WAAI \leq -0.8 as whole-arm losses. Whole-arm gain of 1q and whole-arm loss of 16q by aCGH in an ER-positive, diploid, invasive ductal carcinoma of histological grade 3 (Fig. 2A) were analyzed with FISH, illustrating a combination of probes, indicating a centromere-close translocation t(1q;16p) (Fig. 2B).

To determine the association between clinicopathological information and frequency of chromosomal aberrations, we analyzed four cohorts of breast cancer patients with WAAI and CAAI (Fig. 3A, table S1, and fig. S3). We used three previously published aCGH data sets: MicMa ($n = 125$) (13), WZ ($n = 141$) (15), and Chin-UCAM ($n = 162$) (16) (see Materials and Methods for details). In addition, aCGH data for 167 patients of the Ull cohort not previously published were analyzed. The four cohorts were merged for the analysis of association to clinicopathological information, and we observed an aberration pattern typical for breast cancer (Fig. 3A).

We found that the most frequent events such as gain of 1q and loss of 16q or 17p are whole-arm events, whereas most gains on 17q and losses on 11q have $CAAI \geq 0.5$ and are likely caused by

complex rearrangements (Fig. 3B). A few alterations such as gain on 8q and 20q displayed both whole-arm gain and high CAAI (Fig. 3B). These results suggest that the type of alteration and on which chromosome arm it occurs are of importance in breast cancer.

Defining subgroups based on genomic architecture

Several studies have shown that the number of genomic alterations and the regions preferentially altered differ between the molecular expression subtypes (8, 9, 16, 17). Luminal A or ER-positive tumors often have few alterations with gain of 1q and loss of 16q dominating, whereas basal-like tumors frequently have many alterations affecting most of the chromosomes (8, 9, 16–18). Loss on 5q and gain on 10p have been proposed as specific basal-like alterations (8, 9, 17, 19), similar to findings in breast carcinomas from *BRCA1* mutation carriers (20, 21). On the basis of this, we distinguished among four groups of tumors: those with whole-arm gain of 1q and/or loss of 16q (group A), those with regional loss on 5q and/or gain on 10p (group B), those with both (group AB), and those with neither (group C). The selection of these criteria was not influenced by knowledge of molecular or clinical parameters in the studied cohorts.

To further characterize these groups, we split each into two CAAI subgroups depending on the level of complex rearrangement: those with $CAAI < 0.5$ for all arms (low-level CAAI: A1, B1, AB1, and C1) and those with $CAAI \geq 0.5$ for at least one arm (high-level CAAI: A2, B2, AB2, and C2). The group distribution was similar for all four cohorts, except for the WZ cohort that had more samples of type C and fewer samples with high-level CAAI, most likely because of selection of diploid tumors (table S2) (13). The WAAI score was constructed to capture whole-arm events and not localized gains and losses; this is to reflect underlying defects in DNA maintenance, such as isochromosomes and centromere-close translocations. Localized

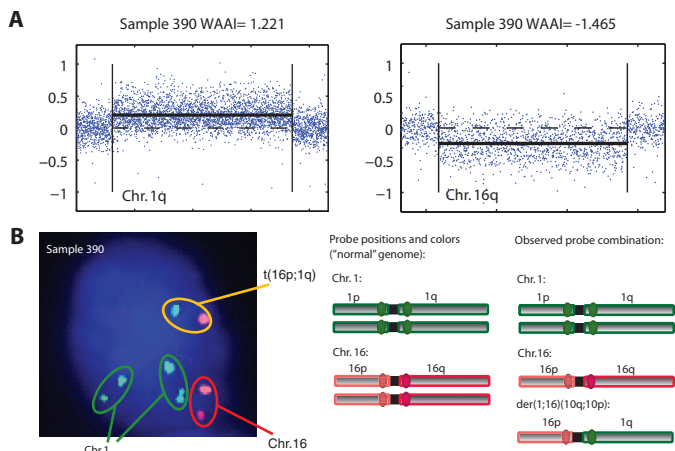


Fig. 2. WAAI and centromere-close translocation. **(A)** Plotted aCGH values for chromosome arms 1q and 16q from sample 390; unsegmented data (blue points) and PCF values (black line) showed whole chromosome arm gain of 1q and loss of 16q. This was reflected in the estimated WAAI ($WAAI = 1.221$ for 1q and $WAAI = -1.465$ for 16q). **(B)** Multigene FISH analyses with four selected probes derived from centromere-close BAC clones on chromosomes 1 and 16 were hybridized to tumor cells (imprint) from sample 390. Left: Tumor cell with all fluorescent probes superimposed revealing two green signals together, one orange and red and one green and orange (note that the probes will not be fused because of the large stretches of heterochromatin around the centromere). Right: Combination of fluorochromes observed in nuclei from lymphocytes with nontranslocated chromosomes 1 and 16, and illustration of the observed combination in the tumor cells probably demonstrating a translocation and a derivative chromosome [der(1;16)(10q;10p)].

Fig. 3. Genome-wide distribution of genomic loss and gain compared to frequencies of WAAI and CAAI in 595 breast carcinomas. **(A)** Frequency plot illustrating the percentage of samples with gain and loss genome wide (red, gain; green, loss). **(B)** The frequency of samples scored with whole-arm changes identified by WAAI and complex rearrangements scored by CAAI are shown in the heat map. The color indicates the percentage of samples with WAAI over and under the chosen threshold and the percentage of samples with CAAI higher than the threshold for each chromosome arm, with thresholds $WAAI \geq 0.8$ (red, top row), $WAAI \leq -0.8$ (green, middle row), and $CAAI \geq 0.5$ (blue, bottom row). The plot illustrates the nonrandom distribution of different types of genomic events.

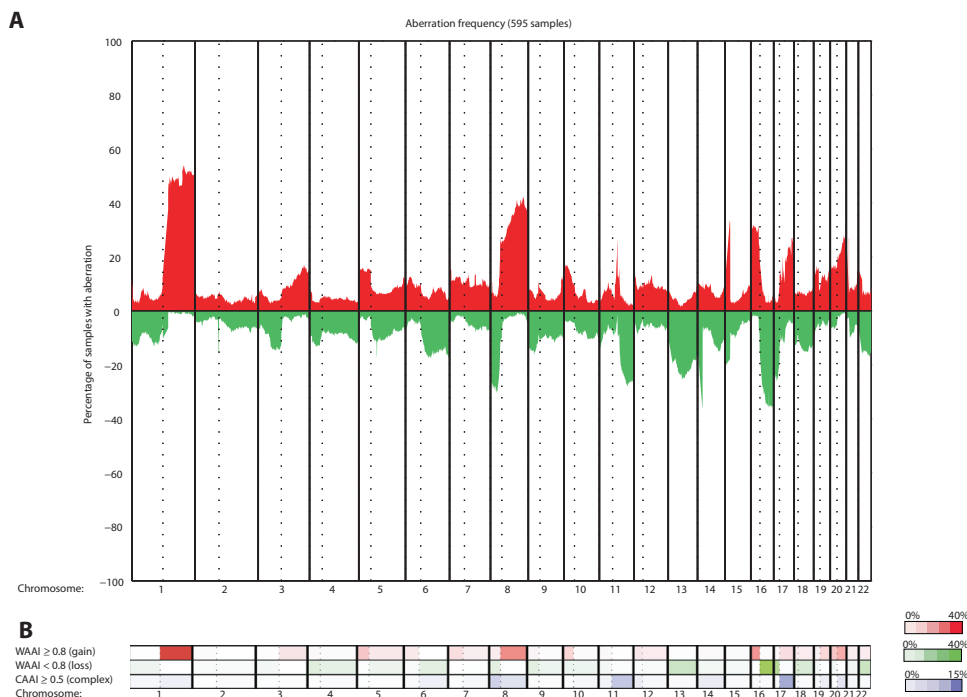
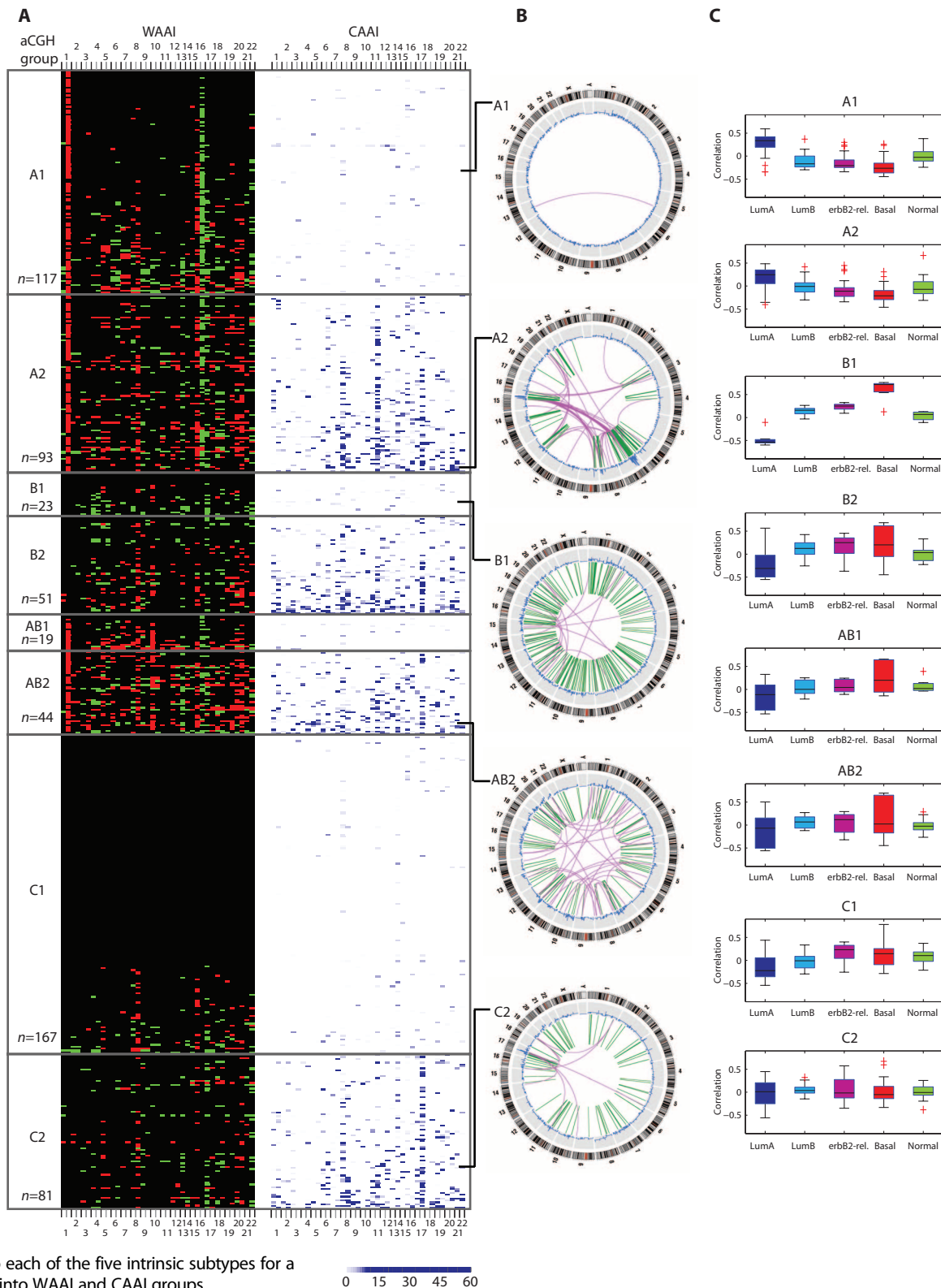


Fig. 4. Genome-wide distribution of WAAI and CAAI for all samples sorted into WAAI groups, examples of identified structural aberrations, and corresponding gene expression patterns. **(A)** The heat map illustrates the WAAI and CAAI score for all 595 samples sorted into A, B, AB, and C tumors and thereafter into groups of tumors with and without high-level CAAI on one chromosome arm or more. The sample sizes of the eight groups are indicated. Each row in the heat map corresponds to one sample, and each column to a chromosome arm (from 1p to 22). Left panel: WAAI alterations for each chromosome arm (red, WAAI ≥ 0.8 ; green, WAAI ≤ -0.8 ; black, $0.8 > \text{WAAI} > -0.8$). Right panel: Corresponding CAAI score for each chromosome arm for the same samples (no rearrangements = white). The CAAI scale is indicated below the figure. **(B)** Structural sequence alterations identified by genome-wide paired-end sequencing for selected samples from the various WAAI or CAAI groups. The outer circle shows the cytobands for each chromosome, followed by the copy number variation. The green bars in the center indicate smaller intrachromosomal changes, whereas pink lines indicate interchromosomal translocations. The lines indicate the position of the selected samples in the aCGH or CAAI groups. **(C)** Correlation to each of the five intrinsic subtypes for a total of 186 cases sorted into WAAI and CAAI groups.



Downloaded from stm.sciencemag.org on February 16, 2011

gains on 1q and losses on 16q would not be classified as A tumors according to our definition. This approach to classifying tumors is an advance over previous stratification paradigms because the criteria are

limited not only to specific genomic regions but also to the architectural type of rearrangements such as gain or loss of whole chromosome arms (9, 13, 22).

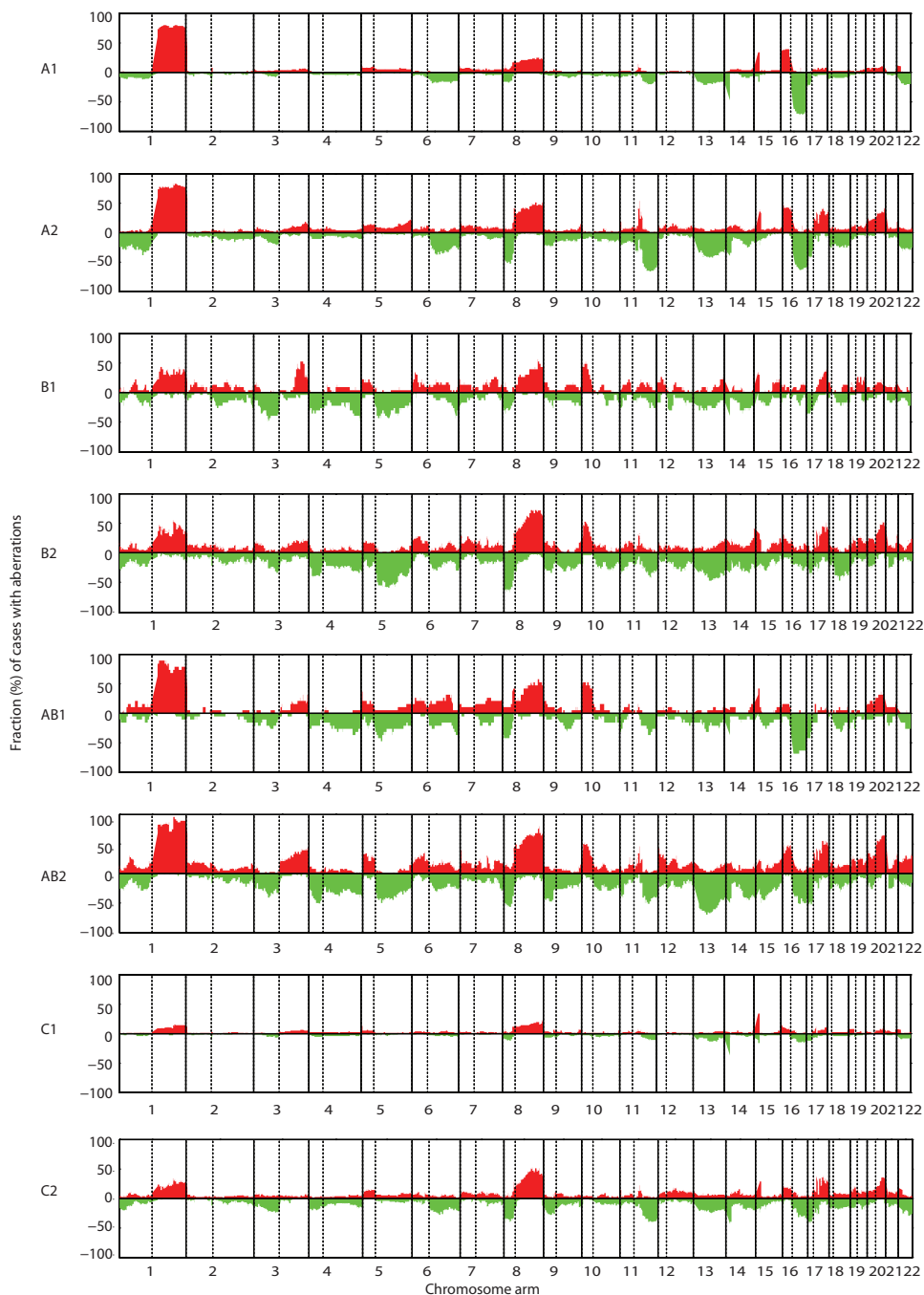


Fig. 5. Frequencies of gain and loss of the eight WAAI- or CAAI-defined groups. The figure shows frequency plots illustrating the percentage of samples with gains and losses within each WAAI or CAAI group (red, gain; green, loss). A1 tumors are dominated by gain on 1q and 16p and loss on 16q. These alterations are frequent in A2 tumors, in addition to gain on 8q, 17q, and 20q and loss on 6q, 8p, 11q, 13, and 17p. B1, B2, AB1, and AB2 tumors have similarities in the patterns of gain and loss with almost all chromosomes affected, a pattern dissimilar from aberrations in A1 and A2 tumors. C1 tumors have few alterations, with gain of 8q dominating. This is the most frequent aberration in C2 tumors as well, followed by gain on 1q, 17q, and 20q.

Patterns of genomic architecture in the WAAI and CAAI groups

WAAI and CAAI characteristics.

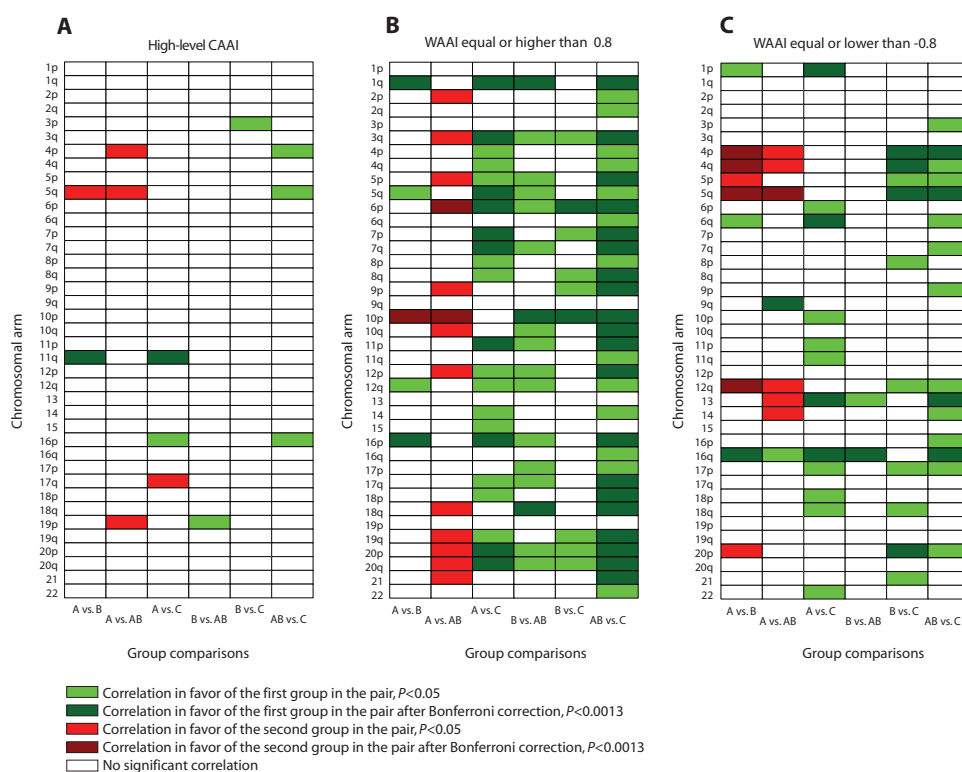
WAAI and CAAI revealed different chromosomal event and frequency distributions among the eight subgroups (Figs. 4A and 5). The subgroups displayed pronounced differences with respect to the number of whole chromosome arm loss or gain events (Fig. 4A and fig. S4). For each of the four WAAI groups, the tumors with complex rearrangements (that is, A2, B2, AB2, and C2) had more whole arms affected, mostly by gains (WAAI ≥ 0.8), than the corresponding group without complex rearrangements.

Tumors of type A were frequently ER-positive, of low or intermediate grade, and diploid and included most of the invasive lobular carcinomas (table S3). Group A was the only group with frequent alterations of whole chromosomes; particularly prominent were gain of 5, 7, 8, and 20 and loss of 18 (Fig. 4A), in line with previous cytogenetic findings (11, 23). A1 and A2 tumors had the same distributions of altered arms, and the increased number of gains seen in A2 tumors mainly affected 8q, 16p, 20p, and 20q (fig. S5). In tumors of type A2, complex rearrangements were most frequent on 11q and 8p, followed by 17q and 8q (fig. S6). The high-level amplifications on 8p and 11q include genes of interest such as *FGFR1* and *CCND1*, loci frequently amplified in ER-positive breast carcinomas (24–26).

Tumors of type B were more frequently of high grade, aneuploid, and *TP53*-mutated than tumors of type A (table S3). Tumors of type B1 were dominated by whole-arm losses, most frequently of 17p, 4p, 4q, and 5q, whereas tumors of type B2 had complex alterations often affecting many arms, most frequently 17q, followed by 8p and 20q (Fig. 4A and figs. S4 to S6). The overall frequencies of aberrations were similar in B1 and B2 (Fig. 5).

AB tumors had elements of both A and B tumors, were dominated by aneuploid tumors of intermediate or high grade, and had the highest frequency of whole-arm alterations (both gains and losses) (table S3 and figs. S4 and S5). The AB tumors with complex rearrangements had

Fig. 6. Pairwise comparison of WAAI and CAAI for the four groups across all chromosome arms. (A to C) A pairwise correlation of high-level CAAI (A), WAAI ≥ 0.8 (B), and WAAI ≤ -0.8 (C) among all eight WAAI or CAAI groups for all chromosome arms. Green indicates a correlation in favor of the first group in the pair, and red indicates a correlation in favor of the second group in the pair. Bright color indicates arms where the correlation reached a significant level ($P < 0.05$), and the dark color indicates arms where the correlation reached a significant level after Bonferroni correction ($P < 0.0013$).



a heterogeneous distribution pattern of arms with high level of CAAI (Fig. 4A and fig. S6).

Group C tumors had the fewest number of whole-arm alterations; however, the most frequent were observed gains of 8q and 16p and losses of 17p and chromosome 22 (Fig. 4A and figs. S4 and S5). This was seen in both C1 and C2 carcinomas, with 17p being more frequently lost in C2 than in C1. For C2 tumors, high level of CAAI was frequent on 17q but rare on 11q (Fig. 4A and fig. S6). The clinicopathological parameters among group C tumors were similar with those in the A group but had fewer ER-positive and more *TP53*-mutated tumors (table S3). Almost half of all tumors with histological grade 1 and most carcinomas of a special histological type such as lobular, tubulolobular, and mucinous were grouped as C by our method.

A pairwise comparison of WAAI and CAAI for the four groups across all chromosome arms showed that alterations of several arms distinguished A tumors from B, AB, and C tumors (Fig. 6). High-level CAAI values on 11q were associated with A tumors, in contrast to C and B tumors, and high-level CAAI values on 5q were characteristic of B and AB tumors and not A and C tumors. The high level of resemblance between CAAI distribution in B and AB was supported by WAAI alterations as well; no arms had significant differences in negative WAAI. This indicates that the whole chromosome arm losses characteristic of the B tumors are also present in AB tumors.

Paired-end sequencing. Paired-end sequencing was performed on a few selected samples representing distinct expression subgroups, and where sufficient amount of DNA was available, these analyses revealed genomic rearrangements down to the single-base level and identified both interchromosomal and intrachromosomal fusions (14) (Fig. 4B). Analysis of the A1 tumor showed a single rearrangement, in contrast to the A2 tumor that exhibited a larger number of complex inter-

chromosomal and intrachromosomal rearrangements, in line with the high-level CAAI. The 1q/16q translocation in the A1 tumor was missed because the paired-end sequencing method does not detect alterations involving centromere-close heterochromatin (27). The B1 tumor showed numerous smaller structural rearrangements (“mutator phenotype”), in contrast to the pattern seen in the A1 and A2 tumors. The AB2 tumor showed a mutator phenotype pattern but with more interchromosomal rearrangements than the B1 tumor. The C2 tumor had some segmental duplications or inversions in addition to complex rearrangements involving chromosome arm 17q.

Gene expression classification. For 186 tumors, gene expression data were available (28, 29). Because no gold standard for assigning samples to subtypes across microarray platforms exists and it has been shown that normalization across data sets with different proportion ER-positive samples affects subtyping (30), correlation to the subtype centroids was based on the original studies. Both A1 and A2 tumors showed strong correlation to the luminal A subtype (Fig. 4C). Luminal B tumors were more frequent in the A2 group, indicating that A2 tumors represent more advanced tumors with high proliferation and increased growth factor signaling than A1 (31) (table S4). This was also supported by ploidy data because the A2 group had a higher fraction of aneuploid and high-grade tumors (fig. S7). The B1 tumors were dominated by the basal-like subtype. The subtype correlation patterns of B2 and AB1/AB2 were similar, dominated by negative correlation to the luminal A subtype, and overall had a closer resemblance to B1 than to A1/A2. Most *erbb2*-enriched and normal-like tumors were classified as C tumors (29 of 45 and 19 of 34, respectively; table S4). Normal-like tumors are rare and often omitted from breast cancer expression classification studies (32). It is acknowledged that samples depleted of tumor cells frequently correlate closely with

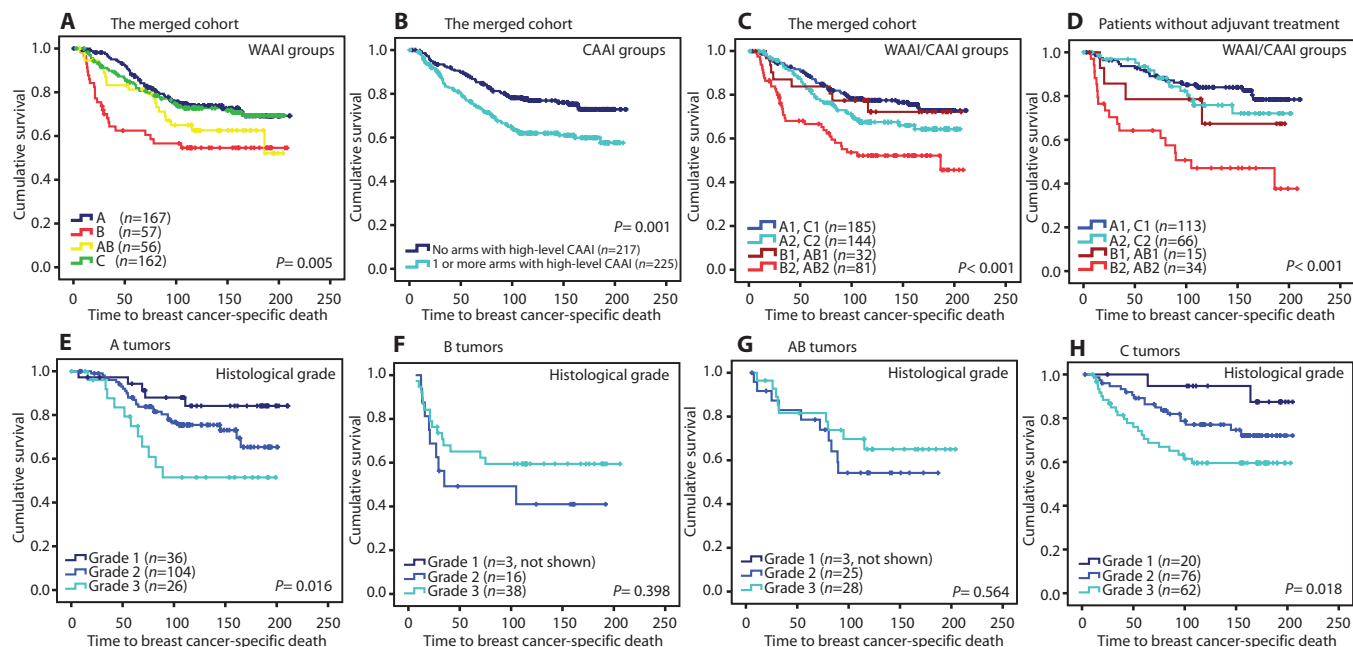


Fig. 7. CAAI and aCGH groups and breast cancer-specific survival in the merged clinical data set ($n = 451$ cases). (A to D) The Kaplan-Meier plots illustrate that breast cancer patients with B and AB tumors have the shortest survival (A), as do patients with high-level CAAI (B). The differences between the groups by combination of WAAI groups and high-level CAAI are shown in (C). The Kaplan-Meier curves show that B2 and AB2 have the worst survival not only in the merged cohort but also in patients who did not receive any adjuvant treatment (D). In (C) and (D),

groups with similar outcome and biology are collapsed. Kaplan-Meier plots of survival estimates of all eight groups are shown in fig. S8. (E to H) The different impact of histological grade in the four WAAI groups is illustrated. Patients with an A or C tumor were stratified into long-, intermediate-, and short-time survival by histological grade ($P = 0.02$ and $P = 0.03$), in contrast to patients with B and AB tumors where we could not show any difference in breast cancer-specific survival related to histological grade.

the normal-like centroid, and the existence of a normal-like subtype is disputed (32). However, normal-like tumors can be aggressive and highly proliferative with stem cell properties, and even be cultivated like the cell line PMC42 (33), and normal-like cell lines have shown enrichment in stem cell-related features (34). Almost 30% of all basal-like tumors were classified as C tumors. This reflects both cases with alterations other than the selected 5q loss and/or 10p gain, but in addition, several cases had almost a flat aCGH profile. The latter is in line with previous studies that have identified a subgroup of basal-like tumors having low genomic instability (16, 35). Data on cellularity of the tumor samples do not suggest that these flat profiles are due to contamination of normal cells or lymphocytes (table S5 and fig. S8). Although some basal-like tumors are shown to be polyclonal (36), it is unlikely that such tumors would result in flat profiles particularly with respect to amplifications.

WAAI and CAAI groups as prognostic markers

Both the WZ cohort, which was highly selected according to ploidy and outcome, and the ductal carcinoma in situ (DCIS) samples were omitted from survival and risk analyses to avoid bias, leaving 451 cases in the merged data set. Kaplan-Meier plots illustrate significant difference in breast cancer-specific death between the four WAAI groups and the two CAAI groups ($P = 0.005$ and $P = 0.001$, respectively; Fig. 7, A and B). By adding CAAI to the WAAI groups, additional prognostic information is obtained as a separation of the B and AB groups into a group with better prognosis and a group with worse prognosis ($P < 0.001$; Fig. 7C and fig. S9). In a multivariate Cox re-

gression analysis, patients with the B type of tumor had a doubled risk of dying of breast cancer compared to those with the A type, independent of lymph node status, tumor size, histological grade, and treatment [hazard ratio (HR), 2.14; 95% confidence interval (CI), 1.20 to 3.81; $P = 0.01$; Table 1A]. We also found an increased hazard rate (HR, 1.74; 95% CI, 1.18 to 2.55; $P = 0.005$) for breast cancer-specific death among patients with high-level CAAI compared to those without, independent of treatment, lymph node status, tumor size, histological grade, and WAAI class (Table 1B). We collapsed groups with similar outcome and biological features to be able to do a multivariate analysis of the combined WAAI and CAAI groups and not lose power (Fig. 7C and fig. S8). In a multivariate Cox analysis, patients with B2 or AB2 tumors had a 2.20 times higher risk (95% CI, 1.35 to 3.59; $P = 0.002$) and patients with A2 or C2 tumors had a 1.37 times higher risk (95% CI, 0.87 to 2.17; $P = 0.17$) of dying from breast cancer compared to patients with low-level CAAI (Table 1C). Survival curves for patients who did not receive any adjuvant therapy clearly show a worse predicted outcome and increase in breast cancer-related death for the B2/AB2 groups (Fig. 7D and table S6). This finding in therapy-naïve patients is important because it supports the view that the differences cannot be explained by therapy sensitivity but seem related to the biology of the tumors. These analyses suggest that high-level CAAI is an independent prognostic marker of breast cancer. A summary of the statistical analyses is presented according to the guidelines for reporting prognostic tumor markers (REMARK) (37) (table S7).

Because we regard several of the WAAI classes to represent distinct entities, we analyzed each of them separately with respect to commonly

used prognostic markers, including histological grade, tumor size, lymph node status, ER status, *TP53* mutation status, and expression-based subtype. Analyzing A tumors ($n = 166$) by univariate Cox regression analysis, we found that histological grade, tumor size, lymph node status, *TP53* mutation status, and high-level CAAI on more than two chromosomal arms were strong prognostic predictors (table S8A). This was in contrast to B tumors ($n = 57$), where only lymph node status and mutated *TP53* indicated an increased hazard rate (borderline significant P value; table S8B). High-level CAAI was a strong prognostic marker in AB tumors ($n = 55$), with an HR of 5.06 (95% CI, 1.38 to 18.59; $P = 0.015$), but only if more than two arms were affected. In C tumors, histological grade, tumor size, lymph node status, and *TP53* status were of importance ($n = 160$) (table S8, C and D). These results suggest that histological grade is prognostic only in A and C tumors (Fig. 7, E to G).

DISCUSSION

Genome-wide, high-resolution analyses of both DNA and RNA have brought novel insights into breast carcinoma classification (3, 9, 16), but conclusions have been limited by small sample sizes. By developing platform-independent algorithms, we could merge aCGH data from several clinical cohorts and perform DNA-based grouping of breast carcinomas using previous DNA and RNA classifications. This, combined with defined surrogate markers for luminal and basal-like breast cancer, revealed several distinct patterns of aberrant genomic architecture. Tumors of type A were dominated by ER-positive, luminal A tumors with large WAAI magnitude (both gains and losses) and by concomitant 1q gain and 16q loss probably caused by unbalanced centromere-close translocations between the two chromosomes (38). The same mechanism affecting other arms might explain the frequent losses and gains of whole chromosome arms in group A. Gain of 1q and/or loss of 16q are seen in different epithelial tumor types such as hepatocellular, ovarian, nasopharyngeal, and prostate carcinomas (39). Gain of 16p is almost always seen together with the loss of 16q, and the loss of 17p is common in a wide variety of tumors (39).

Several studies have indicated that luminal tumors have a distinct progression path (40–43). This is reflected in our study by A2 tumors having more arms with high WAAI magnitude, being more frequently aneuploid, of high grade, and with worse outcomes than A1 tumors (Fig. 4A). Amplification is found to precede the development of aneuploidy in breast cancer cell lines (44), and our study indicates that the same switch also occurs in vivo. Progression from A1 to A2 seems to induce a shift in gene expression pattern, with a higher correlation to the luminal B centroid and worse outcome (Figs. 4C and 7C).

The B tumors had a completely different and more heterogeneous genomic pattern. Group B1 tumors were dominated by losses. The single B1 case investigated by paired-end sequencing had, in addition, the typical mutator phenotype pattern reflecting multiple segmental duplications (14). In two separate studies, we have found that a subgroup of basal-like tumors is characterized by losses and progress from hypodiploid to aneuploid, often with complex rearrangements (36, 45), in line with the B1 group being dominated by losses. Both AB and some C tumors had an expression pattern pointing toward a basal-like relation (Fig. 4C). In addition, AB2 and some C2 tumors had the greatest genomic distortion, were often aneuploid, and had short survival, and we hypothesize that B2, AB2, and some C2 cases reflect more advanced tumors (basal-like, erbb2-enriched, and normal-like).

Table 1. Multivariate Cox regression analysis, the risk of breast cancer-specific death measured by the defined parameters CAAI and WAAI.

Variables ($n = 389$)	Multivariate Cox regression			
	P	HR	95% CI	
			Lower	Upper
A. Four WAAI groups				
WAAI groups				
B tumors (versus A tumors)	0.010	2.14	1.20	3.81
AB tumors (versus A tumors)	0.283	1.38	0.78	2.48
C tumors (versus A tumors)	0.586	1.14	0.72	1.81
Lymph node status				
Positive (versus negative)	<0.001	1.99	1.36	2.91
Tumor size				
pT2 (versus pT1)	0.113	1.38	0.93	2.06
pT3 and pT4 (versus pT1)	0.001	2.83	1.53	5.21
Histological grade				
Grade 2 (versus grade 1)	0.078	2.05	0.92	4.54
Grade 3 (versus grade 1)	0.017	2.71	1.19	6.15
Adjuvant treatment				
Tamoxifen alone (versus no adjuvant)	0.119	0.70	0.45	1.10
Chemotherapy with/without tamoxifen (versus no adjuvant)	0.138	0.65	0.36	1.15
B. Samples with high-level CAAI				
CAAI (high- versus low-level)	0.005	1.74	1.18	2.55
Adjuvant treatment				
Tamoxifen alone (versus no adjuvant)	0.027	0.60	0.39	0.95
Chemotherapy with/without tamoxifen (versus no adjuvant)	0.162	0.67	0.38	1.18
Lymph node status				
Positive (versus negative)	<0.001	2.29	1.49	3.51
Tumor size				
pT2 (versus pT1)	0.067	1.45	0.97	2.16
pT3 and pT4 (versus pT1)	<0.001	3.02	1.63	5.58
Histological grade				
Grade 2 (versus grade 1)	0.082	2.04	0.91	4.56
Grade 3 (versus grade 1)	0.003	3.46	1.54	7.80
WAAI groups				
B tumors (versus A tumors)	0.037	1.85	1.04	3.29
AB tumors (versus A tumors)	0.551	1.20	0.66	2.20
C tumors (versus A tumors)	0.677	1.10	0.69	1.76
C. Combined WAAI and CAAI groups				
(grouped into four)				
A2, C2 (versus A1, C1)	0.172	1.37	0.87	2.17
B1, AB1 (versus A1, C1)	0.901	1.05	0.47	2.36
B2, AB2 (versus A1, C1)	0.002	2.20	1.35	3.59

continued on next page

Lymph node status					
Positive versus negative	0.001	1.89	1.29	2.76	
Tumor size					
pT2 (versus pT1)	0.145	1.35	0.90	2.01	
pT3 and pT4 (versus pT1)	0.001	2.79	1.51	5.16	
Histological grade					
Grade 2 (versus grade 1)	0.123	1.88	0.84	4.20	
Grade 3 (versus grade 1)	0.013	2.81	1.24	6.36	
Adjuvant treatment					
Tamoxifen alone (versus no adjuvant)	0.040	0.62	0.40	0.98	
Chemotherapy with/without tamoxifen (versus no adjuvant)	0.129	0.64	0.36	1.14	

We found that A and B tumors were different at the genomic, transcriptomic, and clinical level. It has been shown that amplifications on 8p/11q and 8q/17q occur preferentially in two phenotypically diverse groups of breast cancer (46), consistent with the different CAAI distribution in A and B tumors. In a study using high-resolution methylation arrays on one of the cohorts, we found patterns of methylation in A tumors resembling the CD24⁺/luminal cell relation and likewise a connection between B tumors and CD44⁺/progenitor cell methylation patterns (47). There are several indicators that molecular subgroups of breast cancer reflect transformation of different breast epithelial cell progenitors (48–50). Our study indicates that molecular subgroups can be recognized by differences in genomic architecture, possibly reflecting underlying subgroup-specific defects linked to different cells of origin. Basal-like carcinomas can be divided into several subtypes (51–53), and recent work indicates that a luminal progenitor on a *BRCA1*-deficient background may be the cell of origin of such tumors (54). We hypothesize that the heterogeneity seen in groups B, AB, and C with respect to the distribution of WAAI and CAAI indicates that tumors of these types descend from different but related early progenitors and that alternative combinations of repair defects define several progression paths.

Complex rearrangements as defined by CAAI occurred in all subgroups, and CAAI had a strong prognostic impact independent of other factors, even if it only occurred on one chromosomal arm. The mechanisms behind complex rearrangements are not completely understood, but one type can be explained by breakage-fusion-bridge cycles because of double-strand repair defects resulting in high-level amplicons with intermittent deletions (55, 56). Because high-level amplicons are seen even in DCIS (57) and in diploid tumors (13), this opens the possibility for a distinct subtype of carcinomas having complex alterations at an early stage of progression (“de novo complexity”).

The findings of this study are based on retrospective analyses of four previously collected data sets and are thus limited by ethnicity, sample size, and inclusion criteria. The acknowledged heterogeneity of breast cancer is evident among these 595 cases, and the results illustrate the need to stratify patients by genomic alterations before clinical risk assessment. Although we present a tool to merge data from different types of platforms, the results will always be imitated by the platform with the lowest resolution. Larger cohorts with long-time follow-up analyzed on high-resolution arrays will be advantageous to validate the findings and to explore the subgroups further.

The present study indicates that the type of genomic architectural distortion is of major importance in determining the tumor phenotype

and can be used to group tumors into luminal and basal-related tumors. This is of major importance because the value of established prognostic markers is subgroup-dependent. We also find that even in biological distinct subtypes of breast cancer, the addition of complex rearrangements seems to be of major importance for patient outcome. A strong hierarchical relation between subtypes of breast carcinomas is yet to be defined, but our findings provide a background for further functional studies aiming to elucidate the relation among genomic architecture, phenotypic traits, and the cell of origin in breast cancer.

MATERIALS AND METHODS

Patient samples and gene expression data

Five hundred and ninety-five patients from four clinical cohorts were included in this study. The aCGH data from three of these cohorts have previously been published and are reanalyzed here. For one cohort of 167 samples, we present new aCGH data. A summary of the cohorts with clinical and pathological data is found in table S1, and detailed clinical information is found in table S5. Previously published expression data were available for a subset of the samples: 112 Chin-UCAM, 113 MicMa, and 73 Ull (16, 28, 29). The subtype assignment for each sample and the centroid correlation values were as published in the original papers.

MicMa cohort. Fresh-frozen tumor biopsies were collected from 150 of the 920 patients included in the “Oslo Micrometastasis Project” from 1995 to 1998 (58). Expression data, *TP53* mutation status, and clinical data for these samples are described (29). One hundred and twenty-five of these samples were available for aCGH analyses and were partly part of a previous publication (13).

WZ cohort. A total of 141 frozen tumor specimens were selected from the archives of the Cancer Center of the Karolinska Institute from 1987 to 1991. Clinical and aCGH data are previously published (13). This cohort was retrospectively selected to represent most diploid cases, where half of the patients were long-term survivors and the other half were dead of breast cancer.

Chin-UCAM cohort. One hundred and sixty-two primary operable breast cancer specimens collected from 1990 to 1996 were obtained from the Nottingham Tenovus Primary Breast Cancer Series. Clinical, expression, and aCGH data are previously published (16).

Ull cohort. Tumor specimens from 212 patients with primary breast cancer were sequentially collected at Oslo University Hospital Ullevål from 1990 to 1994. Primary breast carcinoma tissue was collected at primary surgery and fresh-frozen at -80°C . DNA was isolated from tumor tissue with chloroform-phenol extraction followed by ethanol precipitation (Nuclear Acid Extractor 340A, Applied Biosystems) according to standard procedures. Clinical, *TP53* mutation, and expression data for 80 of these samples are described by Langerød *et al.* (28). Sufficient DNA for aCGH analysis was available from 167 tumors; these data are not previously published.

aCGH data

The raw and preprocessed data can be accessed from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) (<http://www.ncbi.nlm.nih.gov/geo/>) with accession numbers GSE8757 (Chin-UCAM), GSE20394 (Ull), and GSE19425 (MicMa and WZ).

DNA from the MicMa cohort was hybridized to the ROMA (representational oligonucleotide microarray analysis) 85K microarray,

developed at Cold Spring Harbor Laboratory (59). The method is based on oligonucleotide probes designed after the restriction fragments from digestion with Bgl II. The platform is manufactured by NimbleGen, and the experiments followed the ROMA/NimbleGen protocol as previously described (13). Probe intensities were read with the GenePix Pro 4.0 software and used for ratio calculation. The data from the MicMa and the WZ cohort were normalized with an intensity-based lowess curve fitting algorithm.

DNA from the Ull samples was analyzed with 244K CGH microarrays (Hu-244A, Agilent Technologies). This platform contains >236,000 mapped in situ-synthesized oligonucleotide probes representing coding and noncoding sequences of the genome (60). The standard Agilent protocol was used, without prelabeling amplification of input genomic DNA. Scanned microarray images were read and analyzed with Feature Extraction v9.5 (Agilent Technologies) with protocols (CGH-v4_95_Feb07 and CGH-v4 91 2) for aCGH preprocessing, which included linear normalization.

DNA from the Chin-UCAM cohort was as previously described (16) and analyzed with a customized oligonucleotide microarray containing 30K 60-nucleotide oligomer oligonucleotide probes representing 27,800 mapped sequences of the human genome (61). Signal intensities and fluorescent ratios were obtained with BlueFuse version 3.2 (BlueGnome). Raw data were preprocessed with the software R (62) and the Bioconductor package limma (63).

FISH analysis

FISH analysis was performed with imprints from a selected tumor (that is, interphase cells), with nick-translated probes prepared from bacterial artificial chromosomes (BACs) selected to be close to the centromere of chromosomes 1 and 16. The hybridization was performed as previously published (13). Evaluation of signals was carried out in an epifluorescence microscope. Selected cells were photographed in a Zeiss Axioplan 2 microscope equipped with an AxioCam MRm charge-coupled device camera and AxioVision software at minimum 21 z levels. The signals from lymphocytes served as controls. The combination of signals was evaluated and regarded as representative if they were observed in most cells.

Measurements of tumor ploidy

The ploidy of each tumor was determined by measurement of DNA content of nontumoral and tumoral cells independently with Feulgen photometry (64, 65). The optical densities of intact nuclei on an imprint were measured, and a DNA index was calculated and displayed as a histogram. Normal cells and diploid tumors display a major peak at ploidy 2n. Highly aneuploid tumors display broad peaks that often center on ploidy 4n but may include cells from 2n to 6n or above. The histograms were visually interpreted to assign one number to the tumor ploidy. This was done in a nonarbitrary way by selecting the value for which the maximum is reached.

Pathology data and survival analyses

For each series, an experienced pathologist reviewed hematoxylin and eosin slides and immunohistochemical staining for ER and progesterone receptor from all tumors and reclassified them according to the World Health Organization classification guidelines for breast cancer as previously published (13, 16, 28).

The endpoint for the survival analysis was breast cancer-specific death measured from the date of surgery to death of the disease or other-

wise censored at the time of the last follow-up visit or noncancer-related death. Kaplan-Meier survival curves for time to breast cancer-specific death were constructed, and *P* values were calculated by the log-rank test. Cox proportional hazard regression analysis was used for the univariate estimation of prognostic impact for the available clinical parameters, including expression classes, in the four WAAI groups. Three multivariate Cox proportional hazard regression analyses were performed by a backward conditional strategy, including the WAAI groups, CAAI groups, and the combined WAAI and CAAI groups, respectively. Because of the number of events, a selection of clinical variables for inclusion in the models was made, including tumor size (pT), lymph node status, treatment, and histological grade. The same analyses were also performed on nonadjuvant-treated patients separately. All calculations were made with SPSS 16.0, and details about the analyses are in the REMARK document (table S7).

Statistical methods and analytical tools

Segmentation into regions of constant copy number. For each sample, a piecewise constant regression function was fitted to the log-transformed aCGH data with the PCF algorithm (15, 66). For each probe, a fitted value ("PCF value") was thus obtained. The user controls the sensitivity of the method (via a "penalty parameter" γ) and the least allowed number of probes in a segment (k_{\min}). In our case, segmentation was performed on data from three different platforms with relative probe densities (average number of probes per unit distance) of 0.12 (Chin-UCAM), 0.34 (MicMa/WZ), and 1.00 (244K Ull). Because we aimed to pool all the segmented aCGH profiles, we scaled the parameters γ and k_{\min} to obtain roughly equal segmentation resolutions in the three platforms based on the theoretical resolution (thus essentially favoring variance reduction over bias reduction in the estimated copy number profiles for increasing probe densities) (15). Values for γ and k_{\min} were chosen to be 100 and 20 for Ull, 34 and 7 for MicMa/WZ, and 16 and 3 for Chin-UCAM. We acknowledge that the theoretical and the actual resolution may differ in different parts of the genome and that the theoretical functional resolution may be estimated as proposed by Coe *et al.* (67). In our study, ResCalc was not applied. Visual inspection of different segmentations with varying parameter choices indicated that this was a minor problem. The hypothesis of uniform distribution of aberrations is unlikely, and some arrays, such as Agilent 244K, are even constructed to be gene-centered. Furthermore, probes in repetitive regions of the genome will be sparsely spaced to maintain specificity. If such probes were removed, ResCalc would increase the functional resolution; however, the coverage would be reduced (68).

Centering of copy number estimates. To center the segmented data, we found the density of the PCF values using a kernel smoother with an Epanechnikov kernel and a window size of 0.03. The three tallest peaks— P_1 , P_2 , and P_3 —in the density were considered in decreasing order of height (if there was less than three peaks, we replicated the highest one to obtain three peaks). For each, we found the location and relative height (that is, the absolute height of the peak divided by the sum of the heights of the three highest peaks). Between P_1 and P_2 , the peak P with location closest to the median of the PCF values was selected. If the relative height of P was at least 0.2, then the PCF values were centered by subtracting the location of P ; otherwise, the PCF values were centered by subtracting the location of the tallest of all the three peaks.

Whole-arm aberration index. WAAI was found separately for each arm and sample. The normalized PCF (NPCF) values were defined as the centered PCF values divided by the residual SD. The var-

iable s was obtained by averaging NPCF over all probes. If $s > 0$, WAAI was the 5% quantile of NPCF; if $s \leq 0$, WAAI was the 95% quantile of NPCF (in practice, constrained to a predefined grid). Arms with WAAI ≥ 0.8 were called whole-arm gains, and arms with WAAI ≤ -0.8 were called whole-arm losses (see fig. S2 for an example).

Complex arm aberration index. CAAI was found separately for each arm and sample. For each break point found by PCF, we calculated three scores P , Q , and W that reflected the proximity to neighboring break points, the magnitude of change, and a weight of importance:

$$P = \tanh\left(\frac{\alpha}{L_1 + L_2}\right)$$

$$Q = \tanh(|H_2 - H_1|)$$

$$W = \frac{1}{2} \left[1 + \frac{\tanh(10(P - \frac{1}{2}))}{\tanh(5)} \right]$$

where α was a constant, L_1 and L_2 were the numbers of nucleotides, and H_1 and H_2 were properly scaled PCF values for the segments joined at the break point. For any genomic subregion R , we defined

$$S_R = \sum W \cdot \min(P, Q)$$

by summing over all break points in R . We defined CAAI as the maximal value of S_R across all subregions R of a predefined size of 20 Mb. The reason for using a window rather than calculating a score across the whole arm is to avoid spurious calls because of accumulation of isolated events not related to complex rearrangements. The size was a compromise between ensuring a local measure and including enough breakpoints to capture complex rearrangements. Table S5 contains all calculated WAAI and CAAI scores and the group designation for each sample.

The software used here is written in Java and is available at <http://www.ifi.uio.no/bioinf/Projects/GenomeArchitecture>. A guide for the bioinformatic analysis is included on the Web page.

SUPPLEMENTARY MATERIAL

www.sciencetranslationalmedicine.org/cgi/content/full/2/38/38ra47/DC1

Fig. S1. Validation of CAAI.

Fig. S2. Arm-wise distribution of WAAI.

Fig. S3. Frequencies of gains and loss in the four cohorts.

Fig. S4. Frequencies of whole-arm alterations in the WAAI and CAAI groups.

Fig. S5. Chromosome-wise frequencies of whole-arm alterations in the WAAI and CAAI groups.

Fig. S6. Chromosome-wise frequencies of high-level CAAI in the WAAI and CAAI groups.

Fig. S7. Ploidy measurements and histological grade in the WAAI and CAAI groups.

Fig. S8. The tumor cell percentage related to expression subclasses and the WAAI and CAAI groups.

Fig. S9. The prognostic impact of the combined WAAI and CAAI groups in all samples and the WAAI groups, CAAI groups, and combined WAAI and CAAI groups in nonadjuvant-treated patients.

Table S1. Demographic data for the four cohorts.

Table S2. Distribution between the WAAI and CAAI groups in the four cohorts.

Table S3. Clinicopathological characteristics of the four WAAI groups.

Table S4. Correlation between intrinsic subgroups and WAAI and CAAI groups.

Table S5. Clinical data and WAAI and CAAI scores.

Table S6. Multivariate Cox regression analysis of patients who did not receive adjuvant therapy, the risk of breast cancer-specific death measured by the defined parameters CAAI and WAAI.

Table S7. REMARK profile of the study.

Table S8. Univariate Cox regression analysis showing hazard rates for different parameters in the four individual WAAI groups.

REFERENCES AND NOTES

1. Early Breast Cancer Trialists' Collaborative Group (EBCTCG), Effects of chemotherapy and hormonal therapy for early breast cancer on recurrence and 15-year survival: An overview of the randomised trials. *Lancet* **365**, 1687–1717 (2005).
2. C. M. Perou, T. Sørlie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, C. A. Rees, J. R. Pollack, D. T. Ross, H. Johnsen, L. A. Akslen, O. Fluge, A. Pergamenschikov, C. Williams, S. X. Zhu, P. E. Lønning, A. L. Børresen-Dale, P. O. Brown, D. Botstein, Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
3. T. Sørlie, C. M. Perou, R. Tibshirani, T. Aas, S. Geisler, H. Johnsen, T. Hastie, M. B. Eisen, M. van de Rijn, S. S. Jeffrey, T. Thorsen, H. Quist, J. C. Matese, P. O. Brown, D. Botstein, P. Eystein Lønning, A. L. Børresen-Dale, Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 10869–10874 (2001).
4. T. Sørlie, R. Tibshirani, J. Parker, T. Hastie, J. S. Marron, A. Nobel, S. Deng, H. Johnsen, R. Pesich, S. Geisler, J. Demeter, C. M. Perou, P. E. Lønning, P. O. Brown, A. L. Børresen-Dale, D. Botstein, Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 8418–8423 (2003).
5. L. A. Carey, C. M. Perou, C. A. Livasy, L. G. Dressler, D. Cowan, K. Conway, G. Karaca, M. A. Troester, C. K. Tse, S. Edmiston, S. L. Deming, J. Geradts, M. C. Cheang, T. O. Nielsen, P. G. Moorman, H. S. Earp, R. C. Millikan, Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA* **295**, 2492–2502 (2006).
6. R. C. Millikan, B. Newman, C. K. Tse, P. G. Moorman, K. Conway, L. G. Dressler, L. V. Smith, M. H. Labbok, J. Geradts, J. T. Bensen, S. Jackson, S. Nyante, C. Livasy, L. Carey, H. S. Earp, C. M. Perou, Epidemiology of basal-like breast cancer. *Breast Cancer Res. Treat.* **109**, 123–139 (2008).
7. G. S. Dalgin, G. Alexe, D. Scandfeld, P. Tamayo, J. P. Mesirov, S. Ganesan, C. DeLisi, G. Bhanot, Portraits of breast cancer progression. *BMC Bioinformatics* **8**, 291 (2007).
8. A. Bergamaschi, Y. H. Kim, P. Wang, T. Sørlie, T. Hernandez-Boussard, P. E. Lønning, R. Tibshirani, A. L. Børresen-Dale, J. R. Pollack, Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes Cancer* **45**, 1033–1040 (2006).
9. K. Chin, S. DeVries, J. Fridlyand, P. T. Spellman, R. Roydasgupta, W. L. Kuo, A. Lapuk, R. M. Neve, Z. Qian, T. Ryder, F. Chen, H. Feiler, T. Tokuyasu, C. Kingsley, S. Dairkee, Z. Meng, K. Chew, D. Pinkel, A. Jain, B. M. Ljung, L. Esserman, D. G. Albertson, F. M. Waldman, J. W. Gray, Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *Cancer Cell* **10**, 529–541 (2006).
10. S. F. Chin, Y. Wang, N. P. Thorne, A. E. Teschendorff, S. E. Pinder, M. Vias, A. Naderi, I. Roberts, N. L. Barbosa-Morais, M. J. Garcia, N. G. Iyer, T. Kranjac, J. F. Robertson, S. Aparicio, S. Tavaré, I. Ellis, J. D. Brenton, C. Caldas, Using array-comparative genomic hybridization to define molecular portraits of primary breast cancers. *Oncogene* **26**, 1959–1970 (2007).
11. M. R. Teixeira, N. Pandis, S. Heim, Cytogenetic clues to breast carcinogenesis. *Genes Chromosomes Cancer* **33**, 1–16 (2002).
12. B. Dutrillaux, M. Gerbault-Seureau, B. Zafrani, Characterization of chromosomal anomalies in human breast cancer. A comparison of 30 paradiplod cases with few chromosome changes. *Cancer Genet. Cytogenet.* **49**, 203–217 (1990).
13. J. Hicks, A. Krasnitz, B. Lakshmi, N. E. Navin, M. Riggs, E. Leib, D. Esposito, J. Alexander, J. Troge, V. Grubor, S. Yoon, M. Wigler, K. Ye, A. L. Børresen-Dale, B. Naume, E. Schlichting, L. Norton, T. Hägerström, L. Skoog, G. Auer, S. Månér, P. Lundin, A. Zetterberg, Novel patterns of genome rearrangement and their association with survival in breast cancer. *Genome Res.* **16**, 1465–1479 (2006).
14. P. J. Stephens, D. J. McBride, M. L. Lin, I. Varela, E. D. Pleasance, J. T. Simpson, L. A. Stebbings, C. Leroy, S. Edkins, L. J. Mudie, C. D. Greenman, M. Jia, C. Latimer, J. W. Teague, K. W. Lau, J. Burton, M. A. Quail, H. Swerdlow, C. Churcher, R. Natrajan, A. M. Sieuwerts, J. W. Martens, D. P. Silver, A. Langerød, H. E. Russnes, J. A. Foekens, J. S. Reis-Filho, L. van 't Veer, A. L. Richardson, A. L. Børresen-Dale, P. J. Campbell, P. A. Futreal, M. R. Stratton, Complex landscapes of somatic rearrangement in human breast cancer genomes. *Nature* **462**, 1005–1010 (2009).
15. L. O. Baumbusch, J. Aarøe, F. E. Johansen, J. Hicks, H. Sun, L. Bruhn, K. Gunderson, B. Naume, V. N. Kristensen, K. Liestøl, A. L. Børresen-Dale, O. C. Lingjaerde, Comparison of the Agilent, ROMA/NimbleGen and Illumina platforms for classification of copy number alterations in human breast tumors. *BMC Genomics* **9**, 379 (2008).
16. S. F. Chin, A. E. Teschendorff, J. C. Marioni, Y. Wang, N. L. Barbosa-Morais, N. P. Thorne, J. L. Costa, S. E. Pinder, M. A. van de Wiel, A. R. Green, I. O. Ellis, P. L. Porter, S. Tavaré, J. D. Brenton, B. Ylstra, C. Caldas, High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol.* **8**, R215 (2007).
17. J. Adélaïde, P. Finetti, I. Bekhouche, L. Repellini, J. Geneix, F. Sircoulomb, E. Charafe-Jauffret, N. Cervera, J. Desplans, D. Parzy, E. Schoenmakers, P. Viens, J. Jacquemier, D. Birnbaum, F. Bertucci, M. Chaffanet, Integrated profiling of basal and luminal breast cancers. *Cancer Res.* **67**, 11565–11575 (2007).
18. F. Farabegoli, M. A. Hermsen, C. Ceccarelli, D. Santini, M. M. Weiss, G. A. Meijer, P. J. van Diest, Simultaneous chromosome 1q gain and 16q loss is associated with steroid receptor presence and low proliferation in breast carcinoma. *Mod. Pathol.* **17**, 449–455 (2004).

19. A. Vincent-Salomon, N. Gruel, C. Lucchesi, G. MacGrogan, R. Dendale, B. Sigal-Zafrani, M. Longy, V. Raynal, G. Pierron, I. de Mascarel, C. Taris, D. Stoppa-Lyonnet, J. Y. Pierga, R. Salmon, X. Sastre-Garau, A. Fourquet, O. Delattre, P. de Cremoux, A. Aurias, Identification of typical medullary breast carcinoma as a genomic sub-group of basal-like carcinomas, a heterogeneous new molecular entity. *Breast Cancer Res.* **9**, R24 (2007).
20. H. K. Johannsdottir, G. Jonsson, G. Johannsdottir, B. A. Agnarsson, H. Eerola, A. Arason, P. Heikkilä, V. Ejlsson, H. Olsson, O. T. Johannsson, H. Nevanlinna, A. Borg, R. B. Barkardottir, Chromosome 5 imbalance mapping in breast tumors from BRCA1 and BRCA2 mutation carriers and sporadic breast tumors. *Int. J. Cancer* **119**, 1052–1060 (2006).
21. M. Tirkkonen, O. Johannsson, B. A. Agnarsson, H. Olsson, S. Ingvarsson, R. Karhu, M. Tanner, J. Isola, R. B. Barkardottir, A. Borg, O. P. Kallioniemi, Distinct somatic genetic changes associated with tumor progression in carriers of BRCA1 and BRCA2 germ-line mutations. *Cancer Res.* **57**, 1222–1227 (1997).
22. J. Fridlyand, A. M. Snijders, B. Ylstra, H. Li, A. Olshen, R. Segraves, S. Dairkee, T. Tokuyasu, B. M. Ljung, A. N. Jain, J. McLennan, J. Ziegler, K. Chin, S. Devries, H. Feiler, J. W. Gray, F. Waldman, D. Pinkel, D. G. Albertson, Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* **6**, 96 (2006).
23. R. Molist, Y. Remvikos, B. Dutrillaux, M. Muleris, Characterization of a new cytogenetic subtype of ductal breast carcinomas. *Oncogene* **23**, 5986–5993 (2004).
24. A. Letessier, F. Sircoulomb, C. Ginestier, N. Cervera, F. Monville, V. Gelsi-Boyer, B. Esterni, J. Geneix, P. Finetti, C. Zemmour, P. Viens, E. Charafe-Jauffret, J. Jacquemier, D. Birnbaum, M. Chaffanet, Frequency, prognostic impact, and subtype association of 8p12, 8q24, 11q13, 12p13, 17q12, and 20q13 amplifications in breast cancers. *BMC Cancer* **6**, 245 (2006).
25. A. L. Paterson, J. C. Pole, K. A. Blood, M. J. Garcia, S. L. Cooke, A. E. Teschendorff, Y. Wang, S. F. Chin, B. Ylstra, C. Caldas, Co-amplification of 8p12 and 11q13 in breast cancers is not the result of a single genomic event. *Genes Chromosomes Cancer* **46**, 427–439 (2007).
26. J. S. Reis-Filho, K. Savage, M. B. Lambros, M. James, D. Steele, R. L. Jones, M. Dowsett, Cyclin D1 protein overexpression and CCND1 amplification in breast carcinomas: An immunohistochemical and chromogenic in situ hybridisation analysis. *Mod. Pathol.* **19**, 999–1009 (2006).
27. P. J. Campbell, P. J. Stephens, E. D. Pleasance, S. O'Meara, H. Li, T. Santarius, L. A. Stebbings, C. Leroy, S. Edkins, C. Hardy, J. W. Teague, A. Menzies, I. Goodhead, D. J. Turner, C. M. Clee, M. A. Quail, A. Cox, C. Brown, R. Durbin, M. E. Hurler, P. A. Edwards, G. R. Bignell, M. R. Stratton, P. A. Futreal, Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing. *Nat. Genet.* **40**, 722–729 (2008).
28. A. Langerød, H. Zhao, Ø. Borgan, J. M. Nesland, I. R. Bukholm, T. Ikdhah, R. Kåresen, A. L. Børresen-Dale, S. S. Jeffrey, TP53 mutation status and gene expression profiles are powerful prognostic markers of breast cancer. *Breast Cancer Res.* **9**, R30 (2007).
29. B. Naume, X. Zhao, M. Synnestevedt, E. Borgen, H. G. Russnes, O. C. Lingjaerde, M. Strømberg, G. Wiedswang, G. Kvalheim, R. Kåresen, J. M. Nesland, A. L. Børresen-Dale, T. Sorlie, Presence of bone marrow micrometastasis is associated with different recurrence risk within molecular subtypes of breast cancer. *Mol. Oncol.* **1**, 160–171 (2007).
30. L. Lusa, L. M. McShane, J. F. Reid, L. De Cecco, F. Ambrogio, E. Biganzoli, M. Gariboldi, M. A. Pierotti, Challenges in projecting clustering results across gene expression-profiling datasets. *J. Natl. Cancer Inst.* **99**, 1715–1723 (2007).
31. S. Loi, C. Sotiriou, B. Haibe-Kains, F. Lallemand, N. M. Conus, M. J. Piccart, T. P. Speed, G. A. McArthur, Gene expression profiling identifies activated growth factor signaling in poor prognosis (luminal-B) estrogen receptor positive breast cancer. *BMC Med. Genomics* **2**, 37 (2009).
32. J. Peppercorn, C. M. Perou, L. A. Carey, Molecular subtypes in breast cancer evaluation and management: Divide and conquer. *Cancer Invest.* **26**, 1–10 (2008).
33. A. Git, I. Spiteri, C. Blenkiron, M. J. Dunning, J. C. Pole, S. F. Chin, Y. Wang, J. Smith, F. J. Livesey, C. Caldas, PMCA2, a breast progenitor cancer cell line, has normal-like mRNA and microRNA transcripts. *Breast Cancer Res.* **10**, R54 (2008).
34. A. M. Sieuwerts, J. Kraan, J. Bolt, P. van der Spoel, F. Elstrodt, M. Schutte, J. W. Martens, J. W. Gratama, S. Sleijfer, J. A. Foekens, Anti-epithelial cell adhesion molecule antibodies and the detection of circulating normal-like breast tumor cells. *J. Natl. Cancer Inst.* **101**, 61–66 (2009).
35. O. A. Stefansson, J. G. Jonasson, O. T. Johannsson, K. Olafsdottir, M. Steinarsdottir, S. Valgeirsdottir, J. E. Eyfjord, Genomic profiling of breast tumours in relation to BRCA abnormalities and phenotypes. *Breast Cancer Res.* **11**, R47 (2009).
36. N. Navin, A. Krasnitz, L. Rodgers, K. Cook, J. Meth, J. Kendall, M. Riggs, Y. Eberling, J. Troge, V. Grubor, D. Levy, P. Lundin, S. Månér, A. Zetterberg, J. Hicks, M. Wigler, Inferring tumor progression from genomic heterogeneity. *Genome Res.* **20**, 68–80 (2010).
37. L. M. McShane, D. G. Altman, W. Sauerbrei, S. E. Taube, M. Gion, G. M. Clark; Statistics Subcommittee of the NCI-EORTC Working Group on Cancer Diagnostics, Reporting recommendations for tumour MARKer prognostic studies (REMARK). *Br. J. Cancer* **93**, 387–391 (2005).
38. H. Tsarouha, N. Pandis, G. Bardi, M. R. Teixeira, J. A. Andersen, S. Heim, Karyotypic evolution in breast carcinomas with i(1)(q10) and der(1;16)(q10;p10) as the primary chromosome abnormality. *Cancer Genet. Cytogenet.* **113**, 156–161 (1999).
39. M. Baudis, Genomic imbalances in 5918 malignant epithelial tumors: An explorative meta-analysis of chromosomal CGH data. *BMC Cancer* **7**, 226 (2007).
40. H. Buerger, E. C. Mommers, R. Littmann, R. Simon, R. Diallo, C. Poremba, B. Dockhorn-Dwornczak, P. J. van Diest, W. Boecker, Ductal invasive G2 and G3 carcinomas of the breast are the end stages of at least two different lines of genetic evolution. *J. Pathol.* **194**, 165–170 (2001).
41. E. Korsching, J. Packeisen, M. W. Helms, C. Kersting, R. Voss, P. J. van Diest, B. Brandt, E. van der Wall, W. Boecker, H. Bürger, Deciphering a subgroup of breast carcinomas with putative progression of grade during carcinogenesis revealed by comparative genomic hybridisation (CGH) and immunohistochemistry. *Br. J. Cancer* **90**, 1422–1428 (2004).
42. T. M. Abdel-Fatah, D. G. Powe, Z. Hodi, J. S. Reis-Filho, A. H. Lee, I. O. Ellis, Morphologic and molecular evolutionary pathways of low nuclear grade invasive breast cancers and their putative precursor lesions: Further evidence to support the concept of low nuclear grade breast neoplasia family. *Am. J. Surg. Pathol.* **32**, 513–523 (2008).
43. R. Natrajan, M. B. Lambros, F. C. Geyer, C. Marchio, D. S. Tan, R. Vatcheva, K. K. Shiu, D. Hungermann, S. M. Rodriguez-Pinilla, J. Palacios, A. Ashworth, H. Buerger, J. S. Reis-Filho, Loss of 16q in high grade breast cancer is associated with estrogen receptor status: Evidence for progression in tumors with a luminal phenotype? *Genes Chromosomes Cancer* **48**, 351–365 (2009).
44. K. Rennstam, B. Baldetorp, S. Kytölä, M. Tanner, J. Isola, Chromosomal rearrangements and oncogene amplification precede aneuploidization in the genetic evolution of breast cancer. *Cancer Res.* **61**, 1214–1219 (2001).
45. P. van Loo, S. H. Nordgard, O. C. Lingjaerde, H. G. Russnes, I. H. Rye, W. Sun, V. J. Weigman, P. Marynen, A. Zetterberg, B. Naume, C. M. Perou, A. L. Børresen-Dale, V. N. Kristensen, Allele-specific copy number analysis of tumors, unpublished.
46. F. Courjal, M. Cuny, J. Simony-Lafontaine, G. Louason, P. Speiser, R. Zeillinger, C. Rodriguez, C. Theillet, Mapping of DNA amplifications at 15 chromosomal localizations in 1875 breast tumors: Definition of phenotypic groups. *Cancer Res.* **57**, 4360–4367 (1997).
47. S. Kamalakaran, H. G. Russnes, A. Janevski, D. Levy, J. Kendall, V. Varadan, M. Riggs, N. Banerjee, M. Synnestevedt, E. Schlichting, R. Kåresen, R. Lucito, M. Wigler, N. Dimitrova, B. Naume, A. L. Børresen-Dale, J. Hicks, Elucidation of tumor aneuploidy, non-aberrant cell infiltration and LOH from genome wide SNP data in breast carcinomas. Presentation at the annual meeting of the American Society of Human Genetics, Honolulu, 21 October 2009 (<http://www.ashg.org/2009meeting/abstracts/fulltext/f21187.htm>).
48. G. Dontu, D. El-Ashry, M. S. Wicha, Breast cancer, stem/progenitor cells and the estrogen receptor. *Trends Endocrinol. Metab.* **15**, 193–197 (2004).
49. K. Polyak, Breast cancer: Origins and evolution. *J. Clin. Invest.* **117**, 3155–3163 (2007).
50. A. H. Sims, A. Howell, S. J. Howell, R. B. Clarke, Origins of breast cancer subtypes and therapeutic implications. *Nat. Clin. Pract. Oncol.* **4**, 516–525 (2007).
51. J. Kao, K. Salari, M. Bocanegra, Y. L. Choi, L. Girard, J. Gandhi, K. A. Kwei, T. Hernandez-Boussard, P. Wang, A. F. Gazdar, J. D. Minna, J. R. Pollack, Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS One* **4**, e6146 (2009).
52. R. M. Neve, K. Chin, J. Fridlyand, J. Yeh, F. L. Baehner, T. Fevr, L. Clark, N. Bayani, J. P. Coppe, F. Tong, T. Speed, P. T. Spellman, S. DeVries, A. Lapuk, N. J. Wang, W. L. Kuo, J. L. Stilwell, D. Pinkel, D. G. Albertson, F. M. Waldman, F. McCormick, R. B. Dickson, M. D. Johnson, M. Lippman, S. Ethier, A. Gazdar, J. W. Gray, A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**, 515–527 (2006).
53. A. E. Teschendorff, A. Miremadi, S. E. Pinder, I. O. Ellis, C. Caldas, An immune response gene expression module identifies a good prognosis subtype in estrogen receptor negative breast cancer. *Genome Biol.* **8**, R157 (2007).
54. E. Lim, F. Vaillant, D. Wu, N. C. Forrest, B. Pal, A. H. Hart, M. L. Asselin-Labat, D. E. Gyorki, T. Ward, A. Partanen, F. Feleppa, L. I. Huschtscha, H. J. Thorne, KConFab, S. B. Fox, M. Yan, J. D. French, M. A. Brown, G. K. Smyth, J. E. Visvader, G. J. Lindeman, Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* **15**, 907–913 (2009).
55. B. McClintock, The behavior in successive nuclear divisions of a chromosome broken at meiosis. *Proc. Natl. Acad. Sci. U.S.A.* **25**, 405–416 (1939).
56. B. McClintock, The stability of broken ends of chromosomes in *Zea mays*. *Genetics* **26**, 234–282 (1941).
57. V. V. Iakovlev, N. C. Ameson, V. Wong, C. Wang, S. Leung, G. Iakovleva, K. Warren, M. Pintilie, S. J. Done, Genomic differences between pure ductal carcinoma in situ of the breast and that associated with invasive disease: A calibrated aCGH study. *Clin. Cancer Res.* **14**, 4446–4454 (2008).
58. G. Wiedswang, E. Borgen, R. Kåresen, G. Kvalheim, J. M. Nesland, H. Qvist, E. Schlichting, T. Sauer, J. Janbu, T. Harbitz, B. Naume, Detection of isolated tumor cells in bone marrow is an independent prognostic factor in breast cancer. *J. Clin. Oncol.* **21**, 3469–3478 (2003).
59. R. Lucito, J. Healy, J. Alexander, A. Reiner, D. Esposito, M. Chi, L. Rodgers, A. Brady, J. Sebat, J. Troge, J. A. West, S. Rostan, K. C. Nguyen, S. Powers, K. Q. Ye, A. Olshen, E. Venkatraman, L. Norton, M. Wigler, Representational oligonucleotide microarray analysis: A high-resolution method to detect genome copy number variation. *Genome Res.* **13**, 2291–2305 (2003).

60. M. T. Barrett, A. Scheffer, A. Ben-Dor, N. Sampas, D. Lipson, R. Kincaid, P. Tsang, B. Curry, K. Baird, P. S. Meltzer, Z. Yakhini, L. Bruhn, S. Laderman, Comparative genomic hybridization using oligonucleotide microarrays and total genomic DNA. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17765–17770 (2004).
61. P. van den Ijssel, M. Tijssen, S. F. Chin, P. Eijk, B. Carvalho, E. Hopmans, H. Holstege, D. K. Bangarusamy, J. Jonkers, G. A. Meijer, C. Caldas, B. Ylstra, Human and mouse oligonucleotide-based array CGH. *Nucleic Acids Res.* **33**, e192 (2005).
62. R Development Core Team, *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, Vienna, Austria, 2006).
63. R. C. Gentleman, V. J. Carey, D. M. Bates, B. Bolstad, M. Dettling, S. Dudoit, B. Ellis, L. Gautier, Y. Ge, J. Gentry, K. Hornik, T. Hothorn, W. Huber, S. Iacus, R. Irizarry, F. Leisch, C. Li, M. Maechler, A. J. Rossini, G. Sawitzki, C. Smith, G. Smyth, L. Tierney, J. Y. Yang, J. Zhang, Bioconductor: Open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
64. G. Forsslund, B. Nilsson, A. Zetterberg, Near tetraploid prostate carcinoma. Methodologic and prognostic aspects. *Cancer* **78**, 1748–1755 (1996).
65. G. Forsslund, A. Zetterberg, Ploidy level determinations in high-grade and low-grade malignant variants of prostatic carcinoma. *Cancer Res.* **50**, 4281–4285 (1990).
66. O. C. Lingjaerde, L. O. Baumbusch, K. Liestøl, I. K. Glad, A. L. Børresen-Dale, CGH-Explorer: A program for analysis of array-CGH data. *Bioinformatics* **21**, 821–822 (2005).
67. B. P. Coe, B. Ylstra, B. Carvalho, G. A. Meijer, C. Macaulay, W. L. Lam, Resolving the resolution of array CGH. *Genomics* **89**, 647–653 (2007).
68. C. Curtis, A. G. Lynch, M. J. Dunning, I. Spiteri, J. C. Marioni, J. Hadfield, S. F. Chin, J. D. Brenton, S. Tavaré, C. Caldas, The pitfalls of platform comparison: DNA copy number array technologies assessed. *BMC Genomics* **10**, 588 (2009).
69. **Acknowledgments:** We thank E. U. Due for technical assistance. **Funding:** Norwegian Research Council grants 155218/V40 and 175240/S10 and Norwegian Cancer Society grant D99061 (A.-L.B.-D.); Radiumhospitalets Legater (H.G.R. and I.H.R.); Cancer Research UK grant CRUK C507/A3086, Cancer Research UK National Institute for Health Research Cambridge Biomedical Research Centre, and Cambridge Experimental Cancer Medicine Centre (C.C.); Communities Sixth Framework Programme project DISMAL, contract no. LSHC-CT-2005-018911 (B.N. and L.O.B.); Norwegian Cancer Association, Helse Sør-Øst,

and Research Council of Norway (B.N.); Norwegian Research Council, Norwegian Cancer Society, and Helse Sør-Øst (E.B.); Swedish Cancer Society, Swedish Research Council, and Stockholm Cancer Society (A.Z.); and Norwegian Research Council (O.C.L.). **Author contributions:** H.G.R., J.H., and A.-L.B.-D. conceived and designed the study, with valuable input from H.K.M.V. and O.C.L. H.G.R., H.K.M.V., O.C.L., and A.-L.B.-D. developed CAAI, WAAI, and the WAAI or CAAI classification. O.C.L. developed Java software for PCF and centering, with valuable input from H.G.R. and H.K.M.V. H.G.R. performed statistical analyses, with valuable input from H.K.M.V., O.C.L., A.-L.B.-D., C.C., and B.N. H.K.M.V., J.H., A.K., M.W., L.O.B., A.E.T., S.-F.C., and C.C. planned and performed experiments and/or contributed aCGH data. B.N., R.K., E.S., A.L., E.B., and A.-L.B.-D. collected samples and clinical data. E.B. and H.G.R. contributed pathology data. A.Z., H.G.R., I.H.R., P.L., and S.M. planned and performed FISH experiments, with input from A.-L.B.-D. and J.H. H.G.R. and A.Z. planned and performed ploidy experiments, with input from A.-L.B.-D. and J.H. M.P.S. and P.J.S. contributed paired-end sequencing data. A.L., T.S., S.-F.C., A.-L.B.-D., A.E.T., and C.C. contributed gene expression data. A.L. and T.S. contributed TP53 sequencing data. H.G.R., H.K.M.V., O.C.L., A.-L.B.-D., J.H., A.Z., C.C., B.N., and L.O.B. provided valuable discussion. H.G.R., H.K.M.V., O.C.L., and A.-L.B.-D. wrote the paper. **Competing interests:** J.H. is a cofounder of and scientific advisor for GenDx, a company that has licensed genomic technology (ROMA) from Cold Spring Harbor Laboratory. The other authors declare that they have no competing interests. **Accession numbers:** The raw data and preprocessed data can be accessed from NCBI's GEO (<http://www.ncbi.nlm.nih.gov/geo/>), with accession numbers GSE8757 (Chin-UCAM), GSE20394 (Ull), and GSE19425 (MicMa and WZ).

Submitted 10 November 2009

Accepted 4 June 2010

Published 30 June 2010

10.1126/scitranslmed.3000611

Citation: H. G. Russnes, H. K. M. Vollen, O. C. Lingjaerde, A. Krasnitz, P. Lundin, B. Naume, T. Sørli, E. Borgen, I. H. Rye, A. Langerød, S.-F. Chin, A. E. Teschendorff, P. J. Stephens, S. Månér, E. Schlichting, L. O. Baumbusch, R. Kåresen, M. P. Stratton, M. Wigler, C. Caldas, A. Zetterberg, J. Hicks, A.-L. Børresen-Dale, Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci. Transl. Med.* **2**, 38ra47 (2010).