Supporting Information

Van Loo et al. 10.1073/pnas.1009843107

SI Materials and Methods

Derivation of ASCAT Equations. We may express Log R and B Allele Frequency (BAF) at a given genomic location (SNP) as functions of the allele-specific copy numbers n_A and n_B, where n_A denotes the number of copies of the A allele and n_B denotes the number of copies of the B allele. For diploid and homogeneous (100% aberrant cells) samples, and ignoring measurement noise, Log R and BAF (r and b, respectively) are given to a very good approximation by:

$$r_i = \gamma \log_2\left(\frac{n_{A,i} + n_{B,i}}{2}\right) \quad [S1]$$

$$b_i = \frac{n_{B,i}}{n_{A,i} + n_{B,i}}, \quad [S2]$$

where i represents the genomic location and γ is a constant depending on the SNP array technology used [≈0.55 for Illumina (1)], implying a "compaction" of Log R profiles compared with the theoretically expected values. For reference, $r_i = 0$ in regions with copy number 2. In heterozygous regions (of copy number 2), $b_i = 0.5$, whereas in homozygous regions, $b_i = 0$ (homozygous A) or $b_i = 1$ (homozygous B). In case of no compaction ($\gamma = 1$), $r_i = -1$ in deleted regions (copy number 1).

Tumor aneuploidy causes a shift in the Log R value corresponding to copy number 2 (this will no longer equal 0), whereas BAF remains unaffected. We model this by adding the ploidy ψ of the sample to the denominator in the Log R equation:

$$r_i = \gamma \log_2\left(\frac{n_{A,i} + n_{B,i}}{\psi}\right) \quad [S3]$$

In a diploid sample (ploidy 2n, $\psi = 2$), Eq. S3 resolves to Eq. S1.

We model the involvement of nonaberrant cells by breaking up the copy number of both alleles in an aberrant tumor component and a nonaberrant component:

$$n_{\text{total}} = \rho n_{\text{aberrant}} + (1 - \rho) n_{\text{nonaberrant}}, \quad [S4]$$

where ρ represents the aberrant cell fraction (percentage of aberrant tumor cells) for the specimen (ρ ranges from 0 to 1). Furthermore, we assume that the nonaberrant cells have a total copy number of 2 for all loci (2). This results in:

$$r_i = \gamma \log_2\left(\frac{2(1-\rho) + \rho(n_{A,i} + n_{B,i})}{\psi}\right) \quad [S5]$$

$$b_i = \frac{1 - \rho + \rho n_{B,i}}{2 - 2\rho + \rho(n_{A,i} + n_{B,i})}. \quad [S6]$$

In Eq. S5, the ploidy is modeled by $\psi = 2(1 - \rho) + \rho\psi_t$, with ψ_t the tumor ploidy. Eq. S6 is valid for SNPs that are heterozygous in the germline (and thus nonaberrant cell) DNA. SNPs that are germline homozygous will remain homozygous in the tumor sample, resulting in $b_i = 0$ (homozygous A) or $b_i = 1$ (homozygous B), irrespective of the copy number in the aberrant tumor cells. Therefore the BAFs of these SNPs are not informative to infer allele-specific copy numbers and are omitted in this step of the algorithm.

Based on Eqs. S5 and S6, the allele-specific copy number estimates $\hat{n}_{A,i}$ and $\hat{n}_{B,i}$ can be expressed as functions of the Log R value r_i and the BAF value b_i, the parameters ρ and ψ_t (constant for one tumor specimen), and the platform-dependent "technology" parameter γ:

$$\hat{n}_{A,i} = \frac{\rho - 1 + 2^{\frac{r_i}{\gamma}}(1 - b_i)(2(1-\rho) + \rho\psi_t)}{\rho} \quad [S7]$$

$$\hat{n}_{B,i} = \frac{\rho - 1 + 2^{\frac{r_i}{\gamma}} b_i (2(1-\rho) + \rho\psi_t)}{\rho}. \quad [S8]$$

ASPCF Segmentation and Filtering Algorithm. The Allele-Specific Piecewise Constant Fitting (ASPCF) algorithm was developed in MATLAB. The input to the algorithm is (*i*) Log R data and (*ii*) BAF data. Probes with BAF > 0.7 or BAF < 0.3 in the matching blood (germline) data are considered as homozygous in the germline and are masked. The algorithm is applied separately to each of 40 distinct genomic regions, each corresponding to a chromosome, or a chromosome arm (in case of a large centromere): 1p, 1q, 2p, 2q, 3p, 3q, 4p, 4q, 5p, 5q, 6p, 6q, 7p, 7q, 8p, 8q, 9p, 9q, 10p, 10q, 11p, 11q, 12p, 12q, 13q, 14q, 15q, 16p, 16q, 17, 18p, 18q, 19p, 19q, 20p, 20q, 21q, 22q, Xp, and Xq. The ASPCF algorithm fits piecewise constant regression functions simultaneously to the Log R and the BAF data, forcing change points to occur at the same positions in both. This is an extension of the univariate PCF algorithm (3) available in CGH-Explorer (4). In ASPCF, the BAF data are mirrored around 0.5 (resulting in only one band) (5) before the determination of the change points. For a given sample and genomic region, let the data be given by $LRR = \{(x_i, r_i), i = 1, \ldots, n\}$ and $BAF = \{(x_i, b_i), i = 1, \ldots, n\}$. Here, $x_1 < x_2 < \ldots < x_n$ are the probe locations, $r_1, \ldots, r_n$ are the corresponding Log R values, and $b_1, \ldots, b_n$ are the corresponding BAF values. ASPCF seeks an optimal partitioning of the genomic region into segments, or equivalently of the probes into subsets $I_1, \ldots, I_Q$, each consisting of a number of consecutive probes along the genome. An optimal partitioning is one that minimizes the penalized optimization criterion

$$\sum_{j=1}^{Q} \sum_{i \in I_j} \left[w(r_i - ave(\{r_s\}_{s \in I_j}))^2 + (1-w)(b_i - ave(\{b_s\}_{s \in I_j}))^2 \right] + \lambda \cdot Q, \quad [S9]$$

where by default $w = 0.5$. In this expression, minimization is with respect to the number of segments Q as well as the assignment of probes to segments. The first term in the square brackets is the goodness of fit to the Log R data, whereas the second term is the goodness of fit to the BAF data. Here, $ave(\{r_s\}_{s \in I})$ denotes the average of the r_s for probes s in the segment I. The last term in the criterion is a penalty for discontinuities (change points) in the function. The constant $\lambda > 0$ controls the tradeoff between the goodness of fit and the penalty term. When change points have been determined and piecewise constant functions have been fitted to the Log R and BAF data, a final step is performed in which for each segment the mean deviation from 0.5 (called d) was calculated as well as the SD (called s). For a given constant $\tau > 0$, two values symmetric around 0.5 are returned for BAF if $d \geq \tau \cdot s$, and the single value 0.5 is returned otherwise. Input parameters used were: minimum segment length = 6, $\lambda = 50$, and $\tau = \sqrt{3}$.

Aberration Reliability Score. We assess the reliability of each identified aberration in an allele-specific copy number analysis of tumors (ASCAT) profile, by quantifying how much of the deviation in the data (Log R and BAF of the segment s) is explained by ASCAT's integer copy number predictions. The ASCAT integer copy number estimates are:

$$\hat{n}_{A,s}^{ASCAT} = \text{round}\left(\frac{\rho - 1 + 2^{\frac{r_s}{\gamma}}(1-b_s)(2(1-\rho)+\rho\psi_t)}{\rho}\right) \quad \text{[S10]}$$

$$\hat{n}_{B,s}^{ASCAT} = \text{round}\left(\frac{\rho - 1 + 2^{\frac{r_s}{\gamma}}b_s(2(1-\rho)+\rho\psi_t)}{\rho}\right) \quad \text{[S11]}$$

where the *round()* function rounds to the nearest nonnegative integer. On the basis of these estimates $\hat{n}_{A,s}^{\text{ASCAT}}$ and $\hat{n}_{B,s}^{\text{ASCAT}}$, a theoretical Log R and BAF value ($\hat{r}_s^{\text{ASCAT}}$ and $\hat{b}_s^{\text{ASCAT}}$, respectively) is calculated:

$$\hat{r}_s^{ASCAT} = \gamma\log_2\left(\frac{2(1-\rho)+\rho\left(\hat{n}_{A,s}^{ASCAT}+\hat{n}_{B,s}^{ASCAT}\right)}{2(1-\rho)+\rho\psi_t}\right) \quad \text{[S12]}$$

$$\hat{b}_s^{ASCAT} = \frac{1-\rho+\rho\hat{n}_{B,s}^{ASCAT}}{2-2\rho+\rho\left(\hat{n}_{A,s}^{ASCAT}+\hat{n}_{B,s}^{ASCAT}\right)}. \quad \text{[S13]}$$

Finally, both for Log R and BAF, an aberration reliability score ($l_{r,s}$ and $l_{b,s}$, respectively) is calculated as:

$$l_{r,s} = 1 - abs\left(\hat{r}_s^{ASCAT} - r_s\right)/abs(r_s) \quad \text{[S14]}$$

$$l_{b,s} = 1 - abs\left(\hat{b}_s^{ASCAT} - b_s\right)/abs(b_s - 0.5). \quad \text{[S15]}$$

In case of a copy number aberration without allelic imbalance [$abs(r_s) > 0.15$ and $b_s = 0.5$], the final aberration reliability score (percentage) l_s is given as $l_s = 100l_{r,s}$. In case of an allelic imbalance but no copy number aberration [$abs(r_s) \leq 0.15$ and $b_s \neq 0.5$; note that r_s and b_s are values obtained after ASPCF segmentation, which includes a check for one band with $b = 0.5$ vs. two bands symmetric around 0.5], $l_s = 100l_{b,s}$. In case of both a copy number aberration and an allelic imbalance [$abs(r_s) > 0.15$ and $b_s \neq 0.5$], $l_s = 50l_{r,s} + 50l_{b,s}$.

Hence, this aberration reliability score calculates for each aberration how well the ASCAT-predicted integer copy numbers match the data, compared with the hypothesis of no aberration. An aberration reliability score of 100% means ASCAT copy numbers perfectly explain the Log R and BAF data, whereas an aberration reliability score of 0 means the data are explained equally well by the ASCAT copy numbers as by the alternative hypothesis of no aberration.

Experimental Measurements of Tumor Ploidy. Imprints were made by lightly pressing the frozen tumor to a glass slide. By microscopic examination of intact cells both tumor cells and nontumoral cells, such as fibroblasts and lymphocytes, could be recognized.

The ploidy of each tumor was determined by measurement of DNA content of nontumoral and tumoral cells independently using Feulgen photocytometry (6, 7). The optical densities of intact nuclei on an imprint were measured, and a DNA index is calculated and displayed as a histogram (8). Normal cells and diploid tumors display a major peak at ploidy 2n, with a smaller peak of G_2 phase replicating cells that corresponds to the mitotic index. Highly aneuploid tumors display broad peaks that often center on 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 mode of the histogram.

FISH. FISH analysis was performed using imprints (i.e., on interphase cells), with nick-translated probes prepared from BACs. 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 removed, the slides washed twice in 4× sodium chloride/sodium phosphate/EDTA (SSPE) 37 °C and 47 °C, dehydrated in graded alcohol, hexan:isopropanol, and isopropanol, and rehydrated in graded alchohol and 0.1× PBS, then air-dried and mounted with antifade mounting medium containing DAPI (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 AxioCam MRM CCD camera and AxioVision software at minimum 21 *z*-levels. The signals were counted for a minimum of 20 cells in at least four areas of every slide.

Dilution Series of One Breast Carcinoma. DNA was isolated from a fresh frozen liver metastasis of a breast carcinoma, as well as for fresh frozen normal liver tissue. Both materials were mixed in four different (weight) ratios: (*i*) 100% tumor DNA; (*ii*) 80% tumor DNA, 20% host DNA; (*iii*) 65% tumor DNA, 35% host DNA; and (*iv*) 50% tumor DNA, 50% host DNA. The four resulting samples, as well as the normal liver DNA, were hybridized to Human-1 109K BeadChip (Illumina). Data from chromosome Y were removed, and the resulting SNP array data from the tumor dilutions and the normal liver (= germline) were used as input for ASCAT.

Frequency of Gains, Losses, LOH, and Copy Number-Neutral Events. The frequencies of gains and losses when using Log R thresholding (Fig. S6 *A1* and *B1–B5*) were calculated as follows. First, the raw Log R data were smoothed by a moving average filter (averaging over 21 consecutive SNPs). Next, SNPs were considered gained when Log R > 0.12 and lost when Log R < −0.12. The frequency of gains and losses was calculated for the 91 breast carcinomas for which we obtained an ASCAT profile (Fig. S6*A1*) and for a molecular subtype stratification thereof (Fig. S6 *B1–B5*).

For the calculation of frequency of gains and losses from ASCAT profiles (correcting for both nonaberrant cell involvement and aneuploidy; Fig. S6*A2*), SNPs with total copy number 0 or 1 were considered lost, and SNPs with total copy number ≥3 were considered gained.

For the calculation of frequency of gains and losses from ASCAT profiles, relative to tumor ploidy (hence correcting only for nonaberrant cell involvement; Fig. S6 *A3* and *B6–B10*), gains and losses were called relative to tumor ploidy. If the copy number of an SNP was more than 0.6 above tumor ploidy, the SNP was called as a gain. If the copy number of an SNP was more than 0.6 below tumor ploidy, the SNP was called as a loss.

For the calculation of the frequency of loss of heterozygosity (LOH), LOH was called when at least one allele had copy number 0.

A copy number-neutral aberration was called when the total copy number did not differ more than 0.6 from the tumor ploidy, and the copy number of A differed from the copy number of B.

1. Peiffer DA, et al. (2006) High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res* 16:1136–1148.
2. Qiu W, et al. (2008) No evidence of clonal somatic genetic alterations in cancer-associated fibroblasts from human breast and ovarian carcinomas. *Nat Genet* 40:650–655.
3. Baumbusch LO, et al. (2008) Comparison of the Agilent, ROMA/NimbleGen and Illumina platforms for classification of copy number alterations in human breast tumors. *BMC Genomics* 9:379.
4. Lingjaerde OC, Baumbusch LO, Liestøl K, Glad IK, Børresen-Dale AL (2005) CGH-Explorer: A program for analysis of array-CGH data. *Bioinformatics* 21:821–822.
5. Staaf J, et al. (2008) Segmentation-based detection of allelic imbalance and loss-of-heterozygosity in cancer cells using whole genome SNP arrays. *Genome Biol* 9:R136.
6. Forsslund G, Zetterberg A (1990) Ploidy level determinations in high-grade and low-grade malignant variants of prostatic carcinoma. *Cancer Res* 50:4281–4285.
7. Forsslund G, Nilsson B, Zetterberg A (1996) Near tetraploid prostate carcinoma. Methodologic and prognostic aspects. *Cancer* 78:1748–1755.
8. Kronenwett U, et al. (2004) Improved grading of breast adenocarcinomas based on genomic instability. *Cancer Res* 64:904–909.

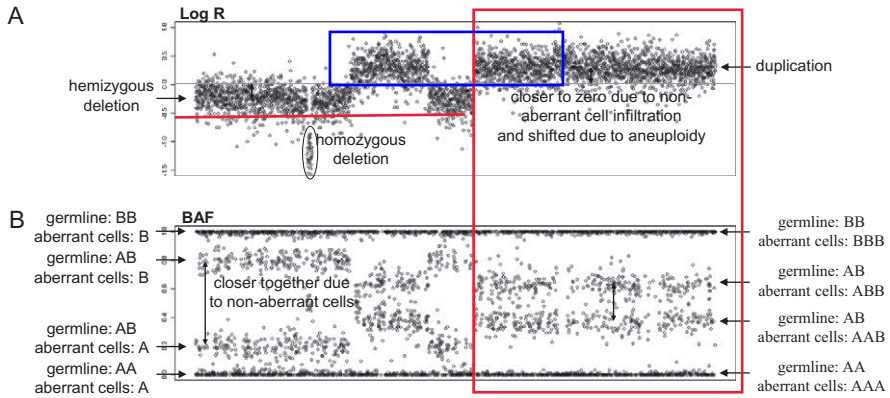

Fig. S1. Virtually all profiled breast carcinomas show evidence of the presence of nonaberrant cells. These nonaberrant cells can be either nontumoral cells in the tumor microenvironment (e.g., fibroblasts, endothelial cells, and infiltrating immune cells) (1) or normal cells in nontumoral regions of the biopsy. We also do not exclude the presence of a (sub)population of tumor cells with no visible aberrations. In addition, a considerable proportion of breast cancers show aneuploidy, most commonly an increased average copy number. For SNP-array platforms as well as for array-CGH (comparative genomic hybridization) platforms, profiling a sample with ploidy above 2n does not result in a higher average Log R. For example, a diploid sample before and after whole-genome duplication would show exactly the same Log R profile, because the amount of DNA per cell is unknown for most Log R preprocessing and normalization methods, and thus an average copy number of 2 is often implicitly assumed. This figure shows (*A*) Log R and (*B*) BAF of a chromosome arm of a breast carcinoma, demonstrating the effects of nonaberrant cell involvement and tumor cell aneuploidy. The nonaberrant cell involvement is most evident in the BAF track in regions with one copy lost (hemizygous deletions). Probes located in hemizygous deletions in a homogeneous sample (i.e., 100% aberrant cells) will appear on one of two narrow bands in the BAF track. One band is found at the bottom edge of the plot (BAF value close to 0) and corresponds to genotype A, and the other band is at the top edge of the plot (BAF value close to 1) and corresponds to genotype B. In the case of a carcinoma infiltrated with nonaberrant cells, two additional bands are observed. These correspond to a mixture of nonaberrant cells with genotype AB, and aberrant cells where A (top band) or B (bottom band) has been lost. The closer the two additional bands are, the smaller the relative signal of the aberrant cells. In the Log R track, the nonaberrant cell involvement is visible as a signal decay: whereas in a sample of aberrant cells only, the mean of Log R drops to −0.55 in case of a hemizygous deletion (2), this drop is smaller when nonaberrant cells are also present. The influence of nonaberrant cell involvement can be seen for other aberrations as well. For example, for duplications, Log R is lower and the BAF bands for the genotypes ABB and AAB are closer than for homogeneous samples. Aneuploidy does not affect BAF (for regions with unchanged copy number and for unchanged nonaberrant cell admixture) but shifts the Log R values. Ploidy above two (the most common case of aneuploidy) results in a downward shift of Log R.

1. Witz IP, Levy-Nissenbaum O (2006) The tumor microenvironment in the post-PAGET era. *Cancer Lett* 242:1–10.
2. Peiffer DA, et al. (2006) High-resolution genomic profiling of chromosomal aberrations using Infinium whole-genome genotyping. *Genome Res* 16:1136–1148.

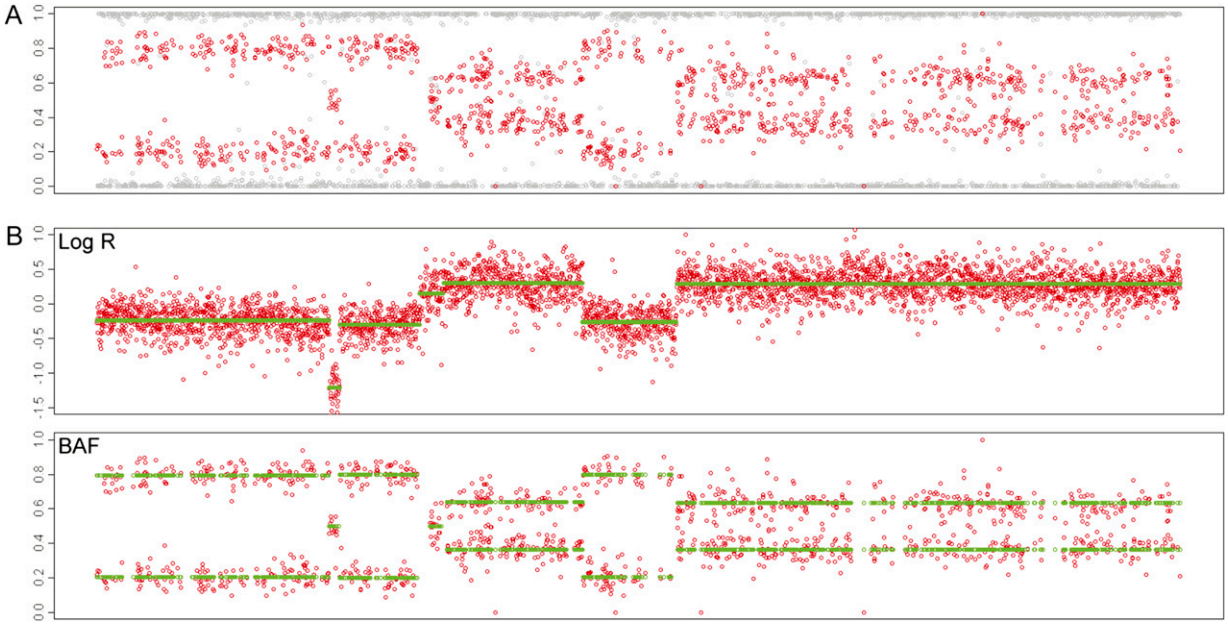

Fig. S2. ASPCF segmentation algorithm. (*A*) Probes homozygous in the germline DNA are removed from the BAF track (red, retained probes; gray, removed probes). BAF (*y* axis) is plotted for one chromosome arm and one sample. The probes are plotted in genomic sequence along the *x* axis. (*B*) ASPCF algorithm is applied to Log R and BAF (red, unprocessed data; green, data after application of the ASPCF algorithm).

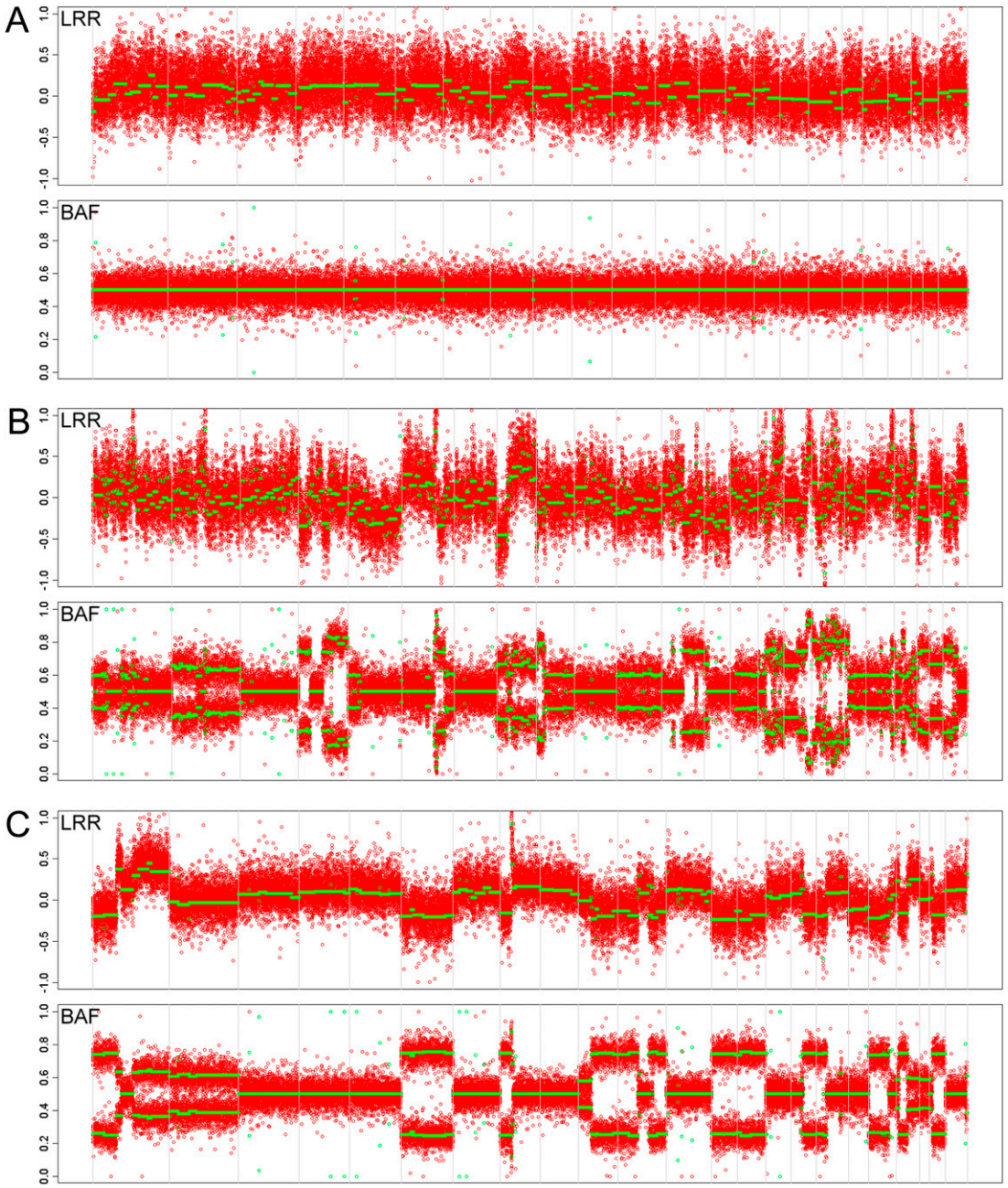

Fig. S3. Two cases excluded by the ASCAT algorithm because no acceptable solution could be identified, and one case for which ASCAT returned a solution (for comparison). Both Log R (depicted as LRR) and BAF are shown for the complete genome. In the BAF track, germline homozygous probes have been removed. Red, raw data; green, ASPCF processed data. (*A*) One case (out of seven in total) with a flat BAF profile, and a Log R profile that remains noisy even after ASPCF segmentation. (*B*) One case showing somewhat similar Log R noise problems but with a nonflat (and nonnoisy) BAF. (*C*) One example case for which ASCAT returned a solution, clearly showing less noise in the Log R profile compared with *A* and *B*.

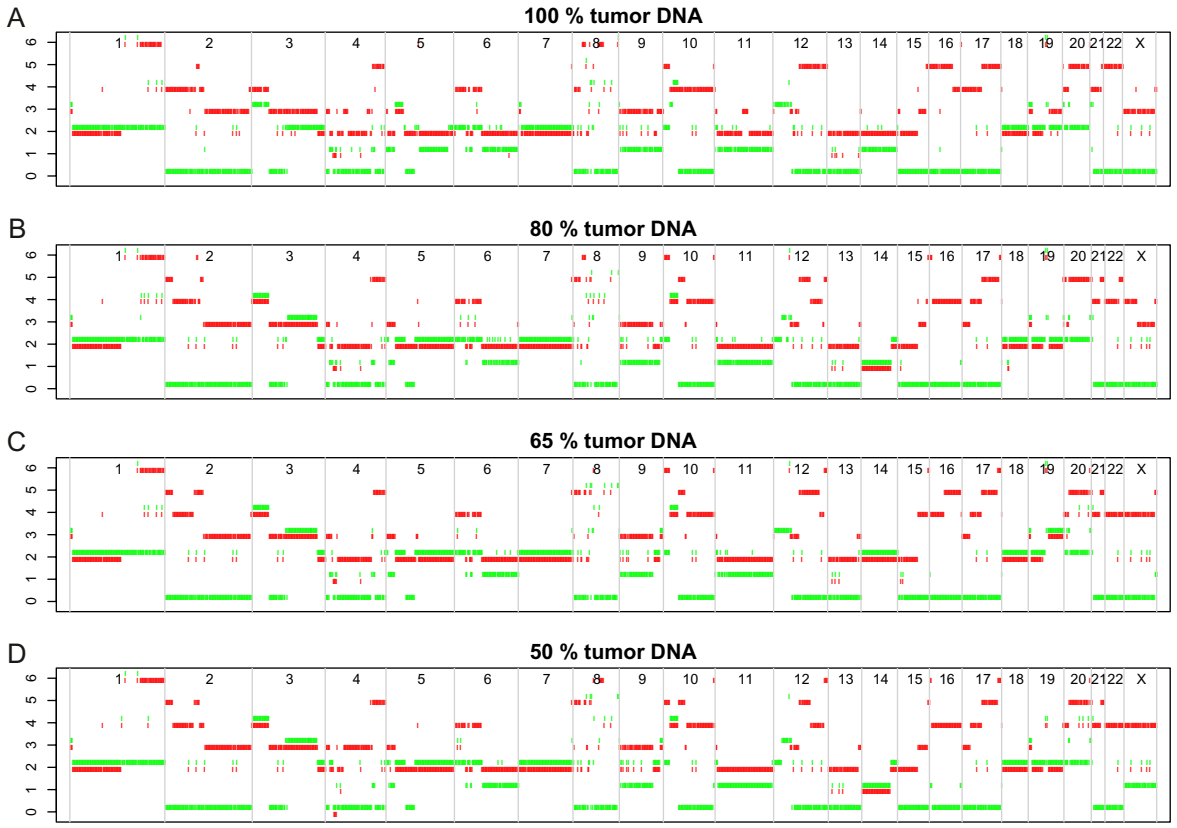

Fig. S4. Validation of ASCAT through a dilution series of a breast carcinoma. ASCAT profiles are shown for a dilution series of a highly aberrant breast carcinoma with ploidy 4.6n. Because the DNA mixes were produced by a total DNA weight ratio (i.e., not cell ratio), the annotated mixes correspond to (*A*) 100%, (*B*) 63%, (*C*) 45%, and (*D*) 30% aberrant tumor cells, assuming that the ploidy is close to 4.6n and the original tumor sample contained no nonaberrant cells. According to ASCAT, the samples contain (*A*) 83%, (*B*) 51%, (*C*) 46%, and (*D*) 32% aberrant tumor cells. Of all heterozygous probes, 64.8% (80% dilution), 60.3% (65% dilution), and 59.9% (50% dilution) showed exactly the same copy number for both alleles as the undiluted sample. For 95.3% (80% dilution), 93.9% (65% dilution), and 92.8% (50% dilution) of the heterozygous probes, the resulting copy numbers differ only slightly or not at all (a maximum copy number difference of 1 was allowed for each allele as well as for their sum).

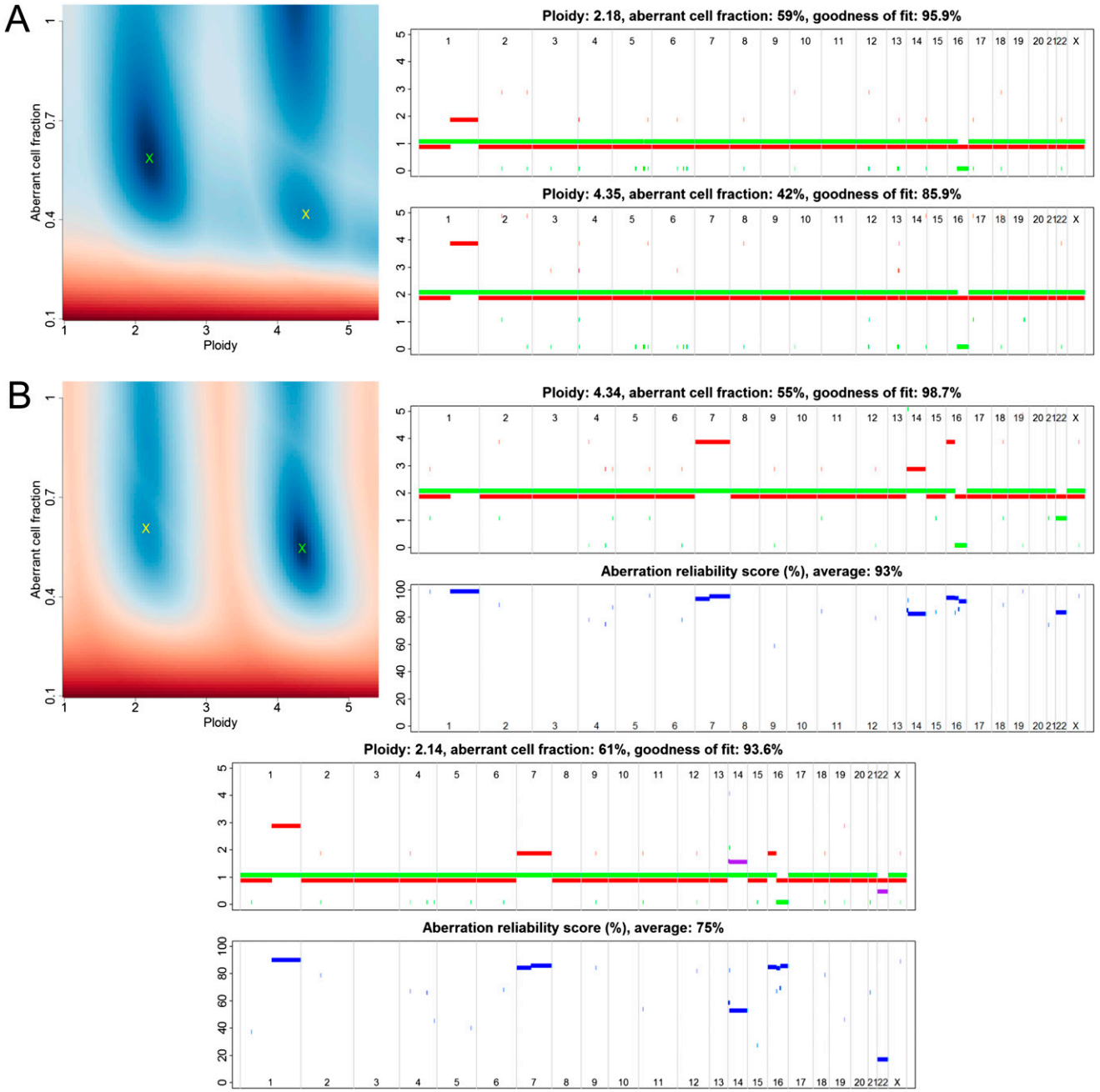

Fig. S5. ASCAT predictions for two cases in which ASCAT's ploidy predictions differ from the experimental ploidy measurements. (*A*) A breast carcinoma identified as nearly diploid according to ASCAT, which according to the ploidy measurements is nearly tetraploid. *Left:* Total distance to nonnegative whole numbers as a function of the aberrant cell fraction ρ and the tumor ploidy ψ_t (red, high distance; blue, low distance). The optimal solution is annotated by a green cross. An alternative tetraploid solution (not selected because of its lower goodness of fit) is annotated by a yellow cross. *Upper Right:* ASCAT profile of the optimal diploid solution. *Lower Right:* ASCAT profile of the alternative tetraploid solution. (*B*) A sample that is nearly diploid according to our experimental ploidy measurements, whereas it is detected as tetraploid by ASCAT. *Upper Left:* Total distance to nonnegative whole numbers as a function of ρ and ψ_t. The optimal solution is annotated by a green cross. An alternative diploid solution (not selected because of its lower goodness of fit) is annotated by a yellow cross. *Upper Right:* ASCAT profile of the optimal tetraploid solution and its corresponding aberration reliability score plot. *Lower:* ASCAT profile of the alternative diploid solution and its corresponding aberration reliability score plot. In the ASCAT profile, for two regions where the allele-specific copy number does not match whole numbers, the allele-specific copy numbers without rounding are plotted in purple (whereas the aberration reliability scores are still based on the rounded copy number predictions). Despite these two regions, our FISH experiments (Table S1) as well as the ploidy measurement suggest that this alternative solution is correct. We hypothesize that the two aberrations denoted in purple (gain of chromosome 14 and loss of chromosome 22) are present only in a subset (≈50%) of the aberrant cells. This tumor heterogeneity causes ASCAT to wrongfully interpret this tumor as nearly tetraploid.

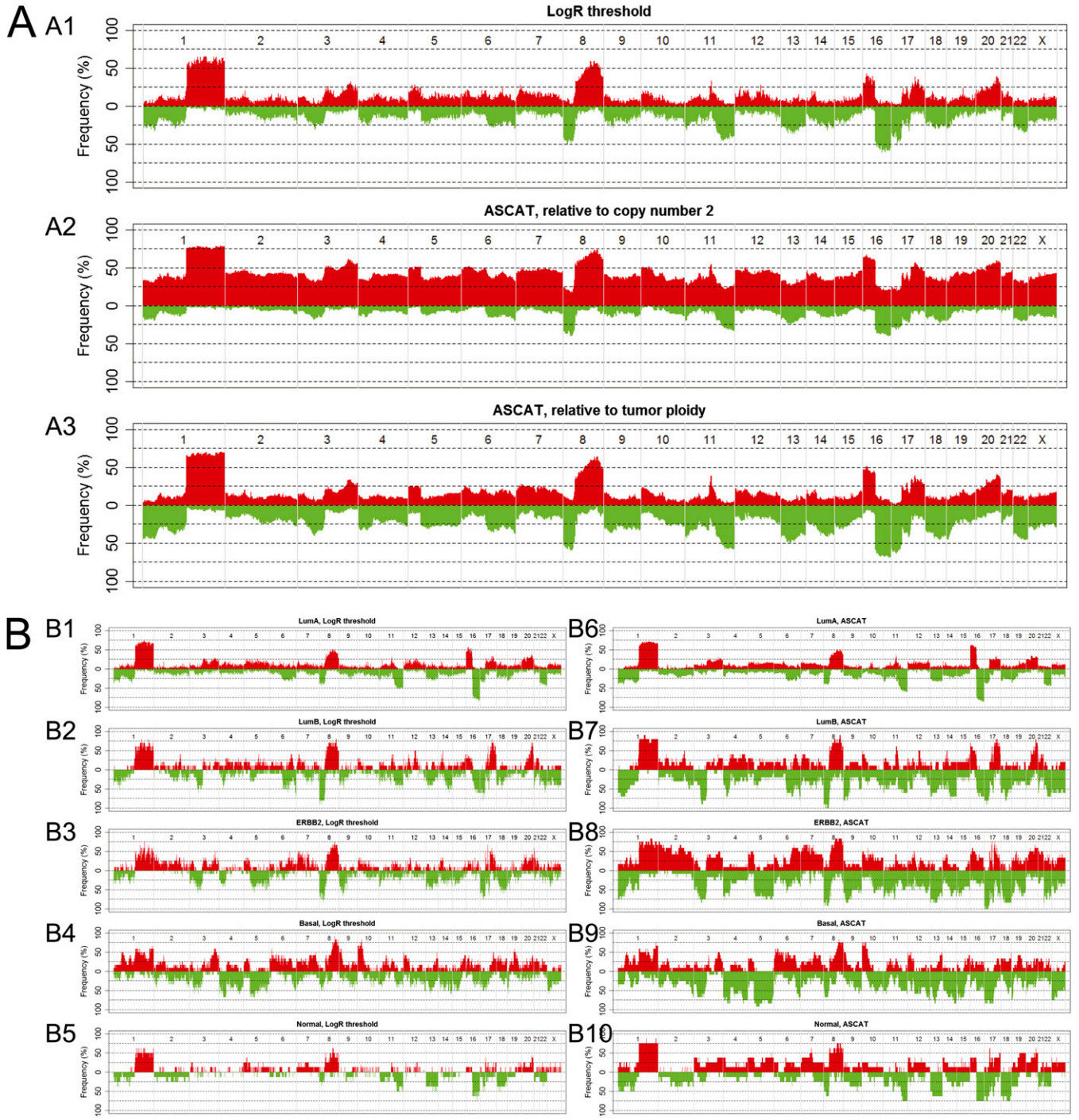

Fig. S6. Recurring gains and losses, assayed using ASCAT profiles. (*A*) Recurring gains and losses on the entire breast carcinoma series. Red, gains; green, losses (depicted downward). Probes are shown in genomic order along the *x* axis, from chromosome 1 to chromosome X, where different chromosomes are delimited by gray lines. (*A1*) Frequency of gains and losses as they would have been detected by common CGH arrays (applying a threshold on Log R) (i.e., without adjusting for ploidy and percentage of aberrant tumor cells). (*A2*) Frequency of gains and losses using ASCAT profiles, relative to a copy number of 2. Consistent with our observation that nearly half of the breast carcinomas have a ploidy of 2.7n or more, we observe a much higher average frequency of gains (copy number above 2) and a lower average frequency of losses (copy number below 2) when using ASCAT profiles. (*A3*) Frequency of gains and losses using ASCAT profiles, relative to the actual estimated ploidy state of the sample. When taking aneuploidy of tumors into account, one may consider not defining gains and losses relative to copy number 2 (e.g., should a locus with copy number 3 be called a gain in a tetraploid tumor?) but rather to define a gain/loss as being significantly above/below the ploidy of the tumor. This approach in a way cancels the correction for aneuploidy from the ASCAT profiles, while keeping the correction for nonaberrant cell involvement. Comparing the results of this approach to approaches not correcting for nonaberrant cell involvement (*A1*), one observes both an increase in frequency of gains and losses, and a decrease in noise. (*B*) Frequency of gains and losses, stratified by molecular breast cancer subtype. (*B1–B5*) Applying a threshold on Log R to identify gains and losses; (*B6–B10*) Using ASCAT profiles to determine gains and losses, relative to the actual estimated ploidy state of the sample; (*B1, B6*) Luminal A subtype ($n = 45$); (*B2, B7*) Luminal B subtype ($n = 10$); (*B3, B8*) ERBB2 subtype ($n = 12$); (*B4, B9*) Basal-like subtype ($n = 12$); (*B5, B10*) Normal-like subtype ($n = 8$). ASCAT profiles show more pronounced genomic frequency distributions, especially for the ERBB2 and Normal-like subtypes, for which both aneuploidy and nonaberrant cell infiltration are highest. Not taking these two factors into consideration, previous reports have described these two subtypes to harbor only a limited number of aberrations (1–3), whereas using ASCAT, this is clearly not the case.

1. Bergamaschi A, et al. (2006) 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.
2. Nordgard SH, et al. (2008) Genome-wide analysis identifies 16q deletion associated with survival, molecular subtypes, mRNA expression, and germline haplotypes in breast cancer patients. *Genes Chromosomes Cancer* 47:680–696.
3. Chin K, et al. (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 10:529–541.

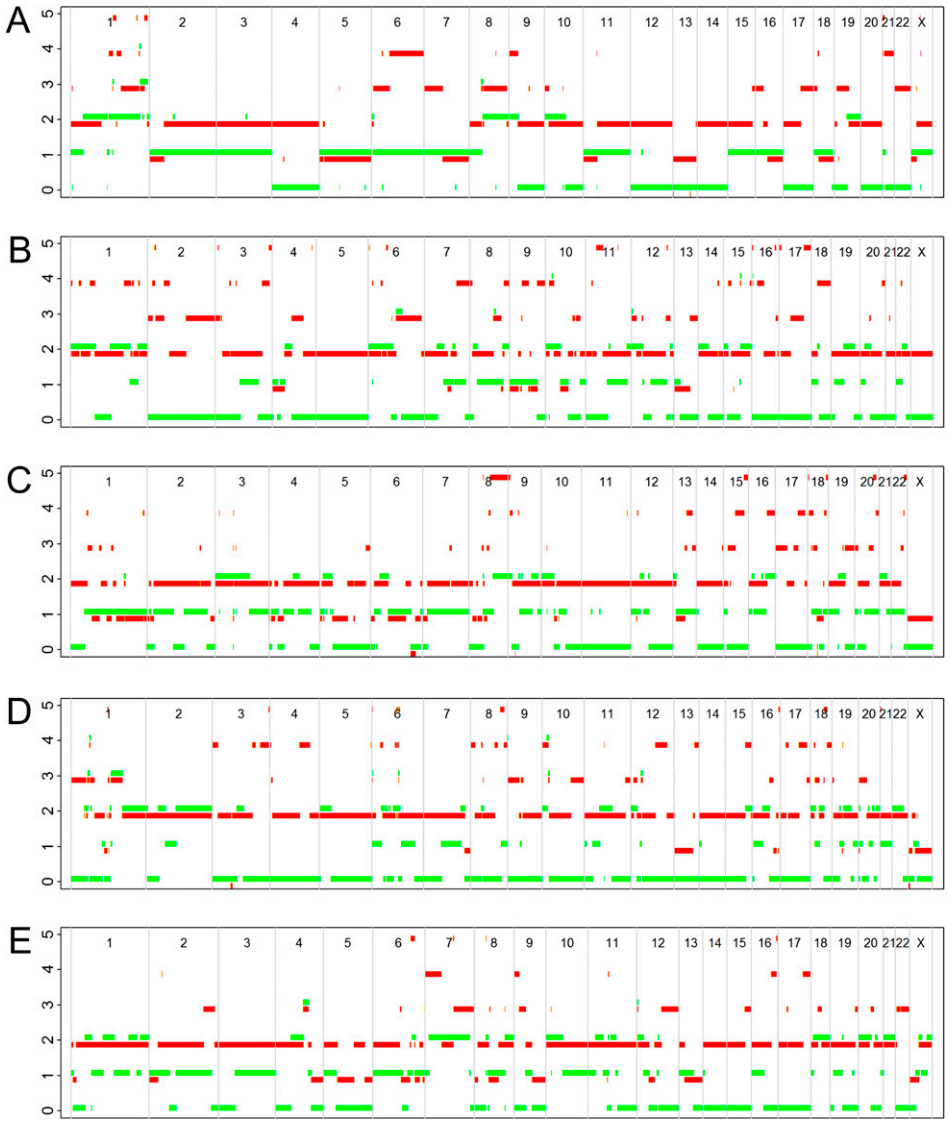

Fig. S7. ASCAT profiles of basal-like carcinomas with ploidy around 3n. (*A*–*E*) These cases are hypothesized to have undergone a whole-genome duplication late in carcinoma development, visible through the frequent occurrence of regions with even copy number of both alleles (mostly copy number 0 for one allele and copy number 2 for the other allele). In cases *B* and *D*, such regions occur particularly frequent, whereas odd allele-specific copy numbers are rare, suggesting that in these cases the whole-genome duplication event was a very late event in the development of these carcinomas.

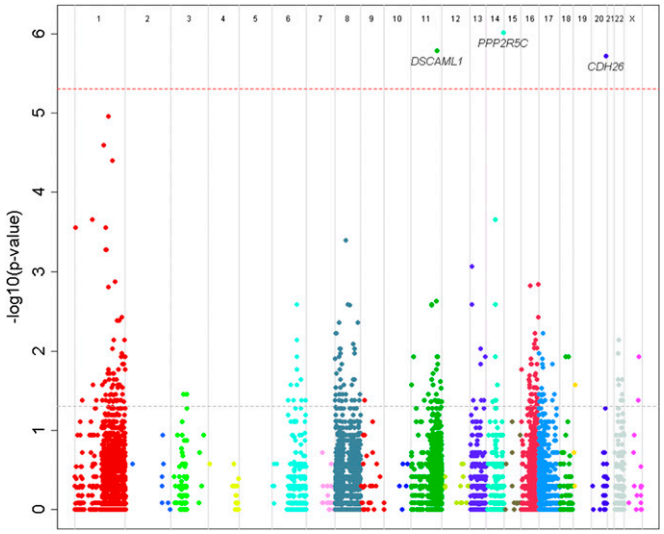

Fig. S8. Manhattan plot of genome-wide allelic skewness, combining gains, losses, and copy number-neutral events. We used binomial statistics to calculate the significance of allelic skewness. For each SNP, we selected the cases from our cohort that are germline heterozygous. We counted how many cases showed gains (of A vs. of B), losses (of A vs. of B), and copy number-neutral events (with gain of A and loss of B vs. with gain of B and loss of A). We combined the counts for gain of A, loss of B, and copy number-neutral events with gain of A and loss of B, and the counts for gain of B, loss of A, and copy number-neutral events with gain of B and loss of A, and next performed a two-sided binomial test. Several probes displayed a preferential gain of the one allele and loss of the other, implying a selective advantage for the gained, and a disadvantage for the lost allele in breast carcinomas. Probes are shown in genomic order along the *x* axis, from chromosome 1 to chromosome X. Only probes with a total of at least 20 gains, losses, and copy number-neutral events are shown. The *y* axis shows the −log10 transformed *P* values for allelic skewness. The gray dotted line indicates the nominal *P* value threshold, whereas the red dotted line shows the significance threshold after Bonferroni multiple-testing correction.

Table S1. Results of FISH validation experiments performed on three different genomic locations for 11 breast carcinomas

	1p		16p		17q (HER2)	
Sample	ASCAT	FISH	ASCAT	FISH	ASCAT	FISH
1	2	2	2	2	1	1
2	2	2	2	2	2	2
3	2	2	3	3	5	NA
4	3	2	6	4	2	2
5	5	3	4	3	13	11
6	2	2	2	2	2	NA
7*	4	2	6	3	4	2
8	3	2	5	3	3	2
9	2	2	2	2	2	2
10	2	2	2	2	2	2
11	2	2	6	4	5	3

The samples and genomic locations were selected such that a good range of copy numbers is covered.

*For this breast carcinoma, ASCAT predicted a ploidy close to 4n, whereas our experimental ploidy measurements found a ploidy close to 2n (one of the red triangles in Fig. 2, main text). Fig. S5*B* shows the ASCAT output for this case, confirming that this error is likely caused by heterogeneity of the aberrant tumor cells.