Inference of tumor evolution during chemotherapy by computational modeling and in situ analysis of cellular diversity for genetic and phenotypic features

Vanessa Almendro, Yu-Kang Cheng, Amanda Randles, Shalev Itzkovitz, Andriy Marusyk, Elisabet Ametller, Xavier Gonzalez-Farre, Montse Muñoz, Hege G. Russnes, Åslaug Helland, Inga H. Rye, Anne-Lise Borresen-Dale, Reo Maruyama, Alexander van Oudenaarden, Mitchell Dowsett, Robin L. Jones, Jorge Reis-Filho, Pere Gascon, Mithat Gönen, Franziska Michor, and Kornelia Polyak

Supplemental Figures

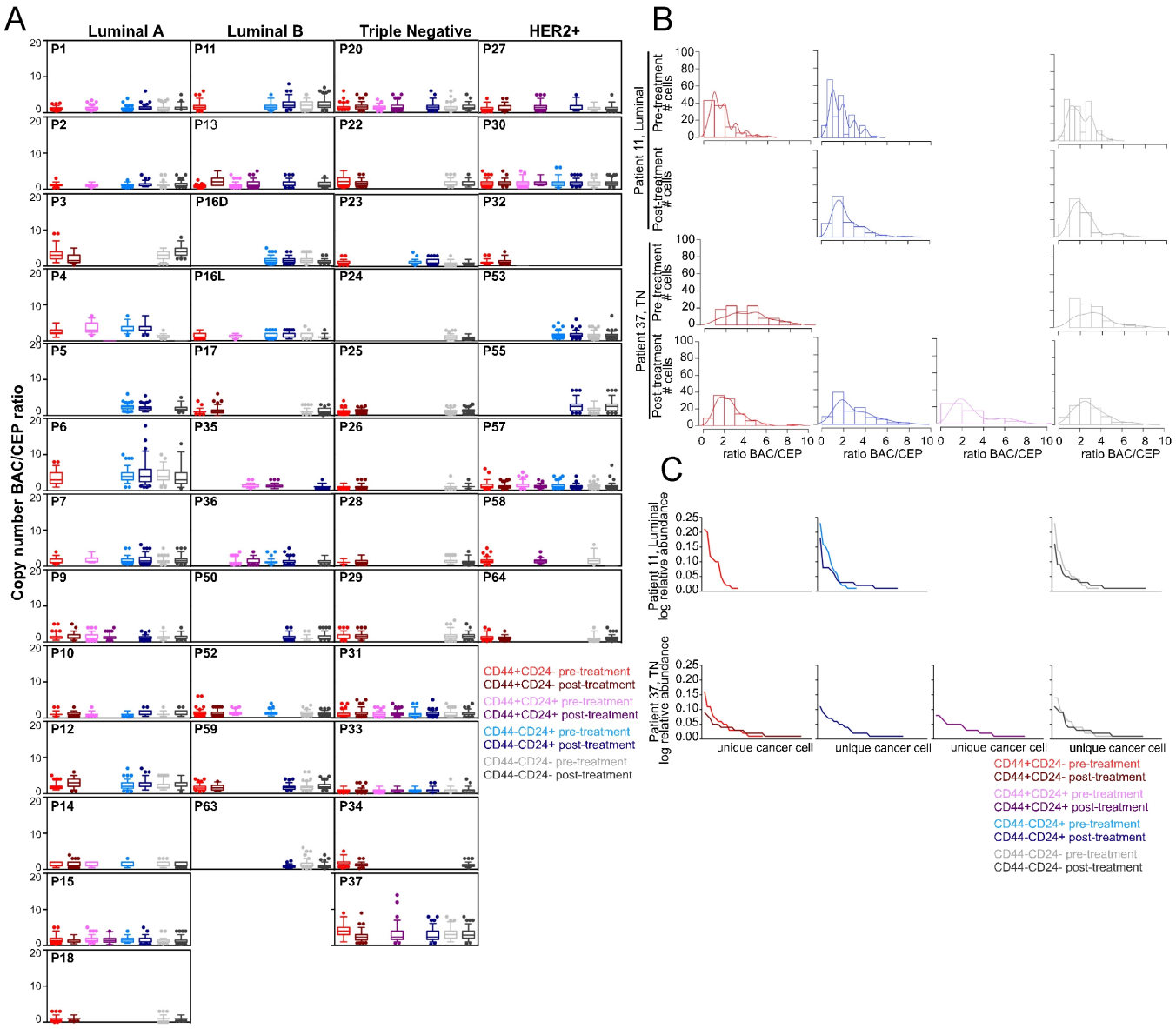

Figure S1. Box plots depicting 8q24 BAC/Chr8 CEP ratios in each cell subpopulation before and after treatment as well as Kernel and Whittaker plots. Related to Figure 1.
(A) The boxes show the 25th to 75th percentile, and the whiskers extend to the 5th and 95th percentiles. Outliers outside the 5th and 95th percentile are shown as colored dots. P1-64 indicates individual patients.
(B) Histograms and kernel density estimates of species abundance depicting the distribution of cells with the indicated copy number ratio.
(C) Whittaker plots (rank-abundance plots) depicting the abundance of unique cancer cells.

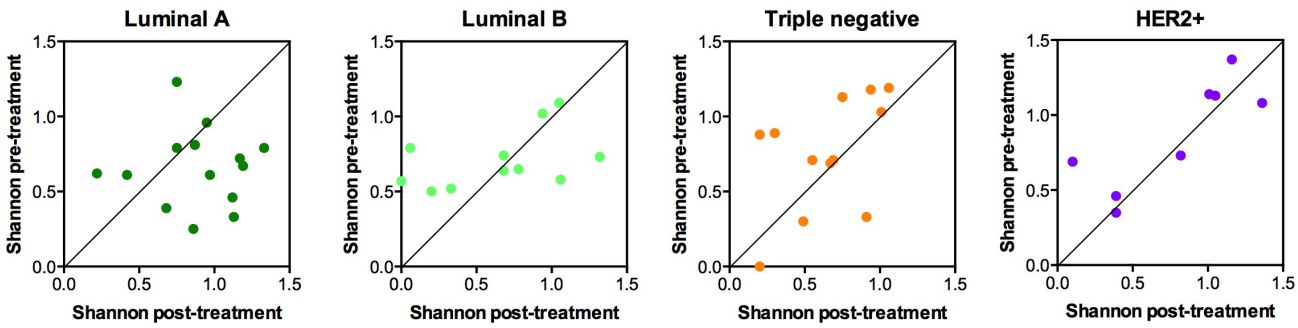

Figure S2. Shannon diversity index based on cellular phenotypes. Related to Figure 1.
The Shannon diversity index for each sample before and after treatment is shown.

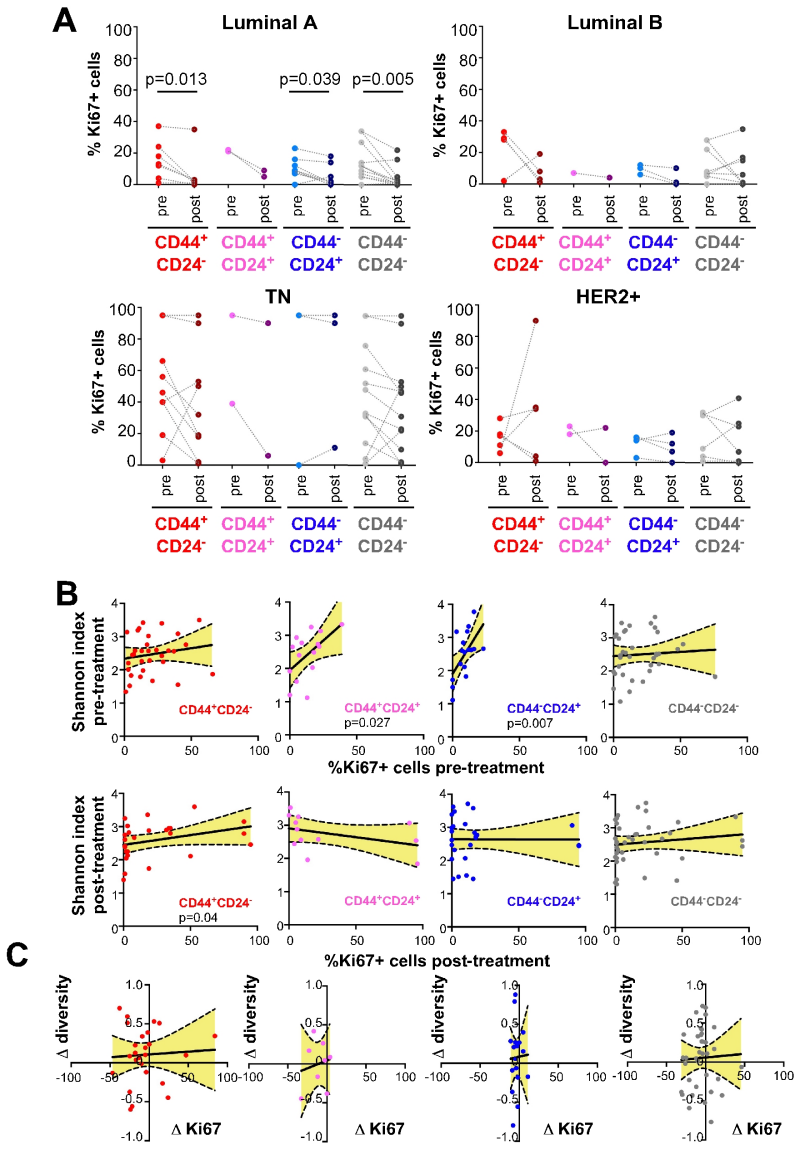

Figure S3. Frequency of Ki67$^+$ cells in the indicated cell subpopulations before and after treatment. Related to Figure 2.

(A) Changes in the frequency of Ki67$^+$ cells within the indicated cell subpopulations in the different tumor subtypes. Dotted line connects values for each cell subpopulation before and after treatment. Significant p values by two-sided Wilcoxon matched-pairs signed rank test are shown. Pre: pre-treatment and post: post-treatment.

(B) Correlation between Shannon index of diversity and proliferation rate (% Ki67$^+$ cells) in the indicated cell subpopulations before and after treatment.

(C) Correlation between changes in Shannon index and fraction of Ki67$^+$ cells (Δ denotes post-minus pre-treatment values). Significance of associations between variables was defined based on Spearman correlation analysis. A 95% confidence interval is indicated in yellow.

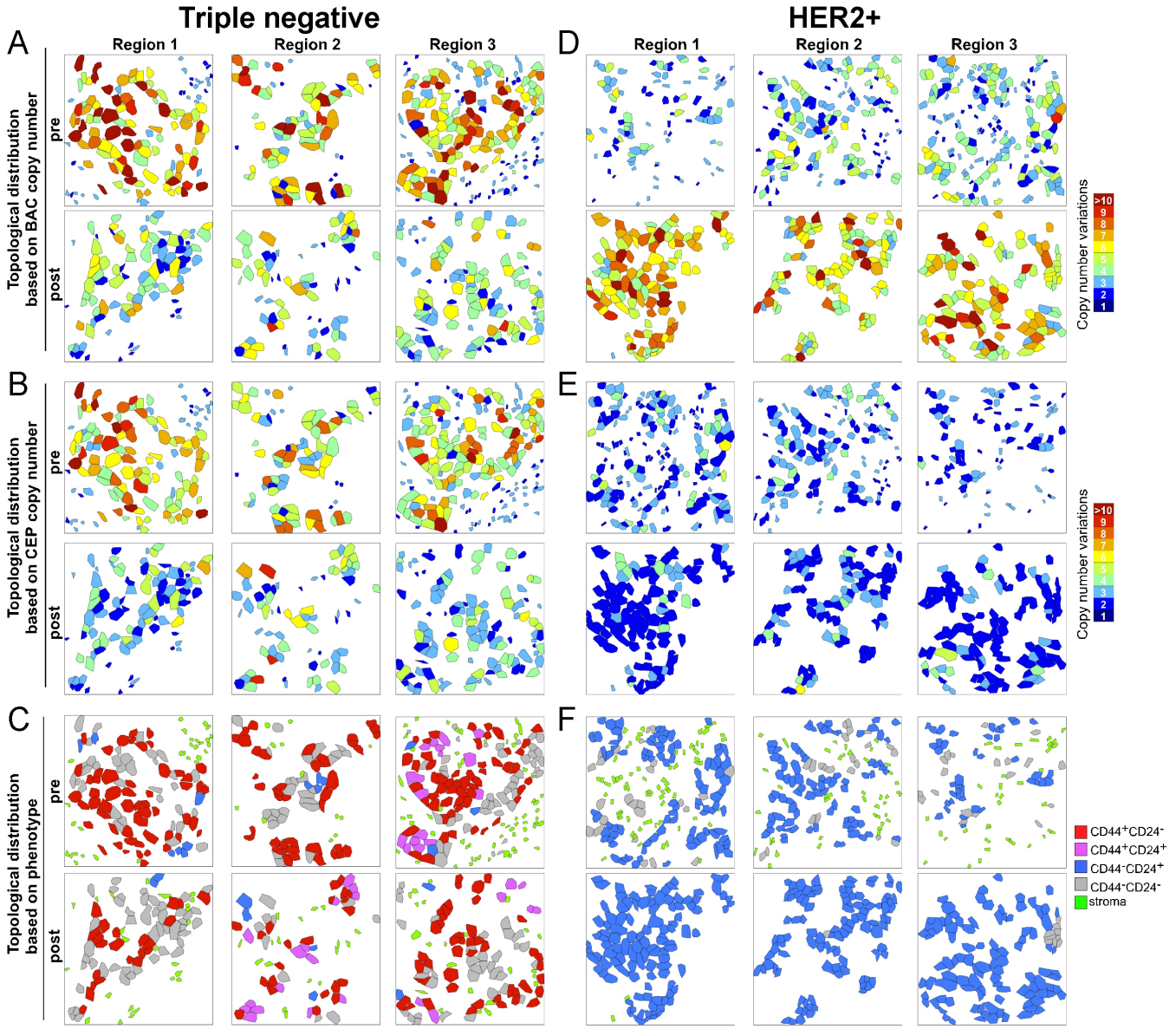

Figure S4. Analysis of the topologic distribution of genetically and phenotypically divergent cells within tumors. Related to Figure 3.

Maps showing topologic differences in the distribution of genetically distinct tumor cells based on copy number for 8q24 BAC, chr8 CEP probes, and cellular phenotypes in three different regions of a triple negative tumor (Patient 20) (A-C) and in a HER2+ tumor (patient 30) (D-E).

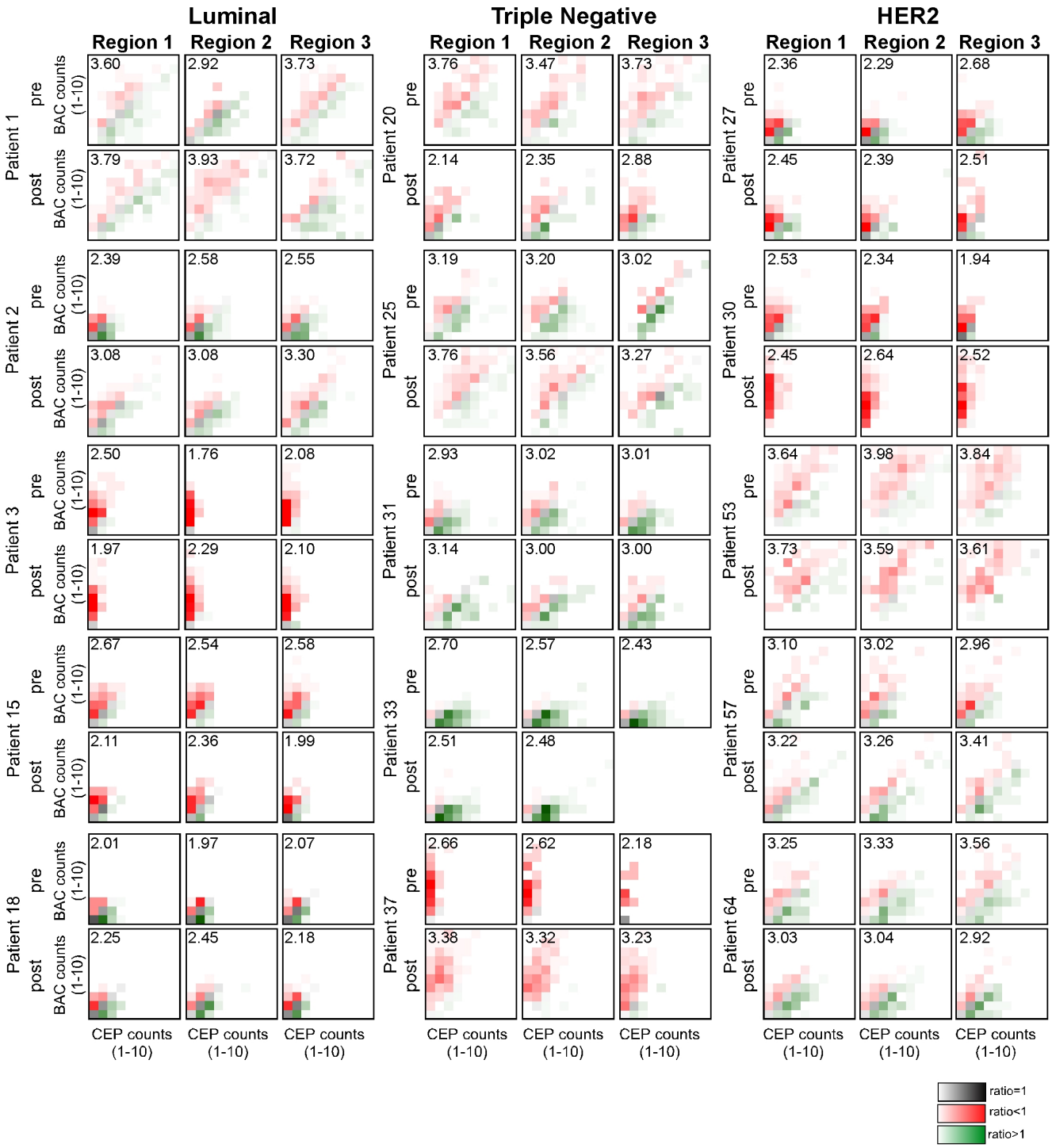

Figure S5. Heatmaps depicting genetic diversity in topologically distinct regions. Related to Figure 3.

Each colored square represents cells with specific 8q24 BAC and Chr8 CEP copy numbers. Red and green color corresponds to relative copy number gain (BAC/CEP>1) and loss (BAC/CEP<1), respectively, whereas grey indicates lack of copy number change (BAC/CEP=1). Intensity of shading corresponds to the frequency of cells with the specific copy numbers. Shannon index calculated based on BAC and CEP unique values are listed in each plot. Pre: pre-treatment and post: post-treatment.

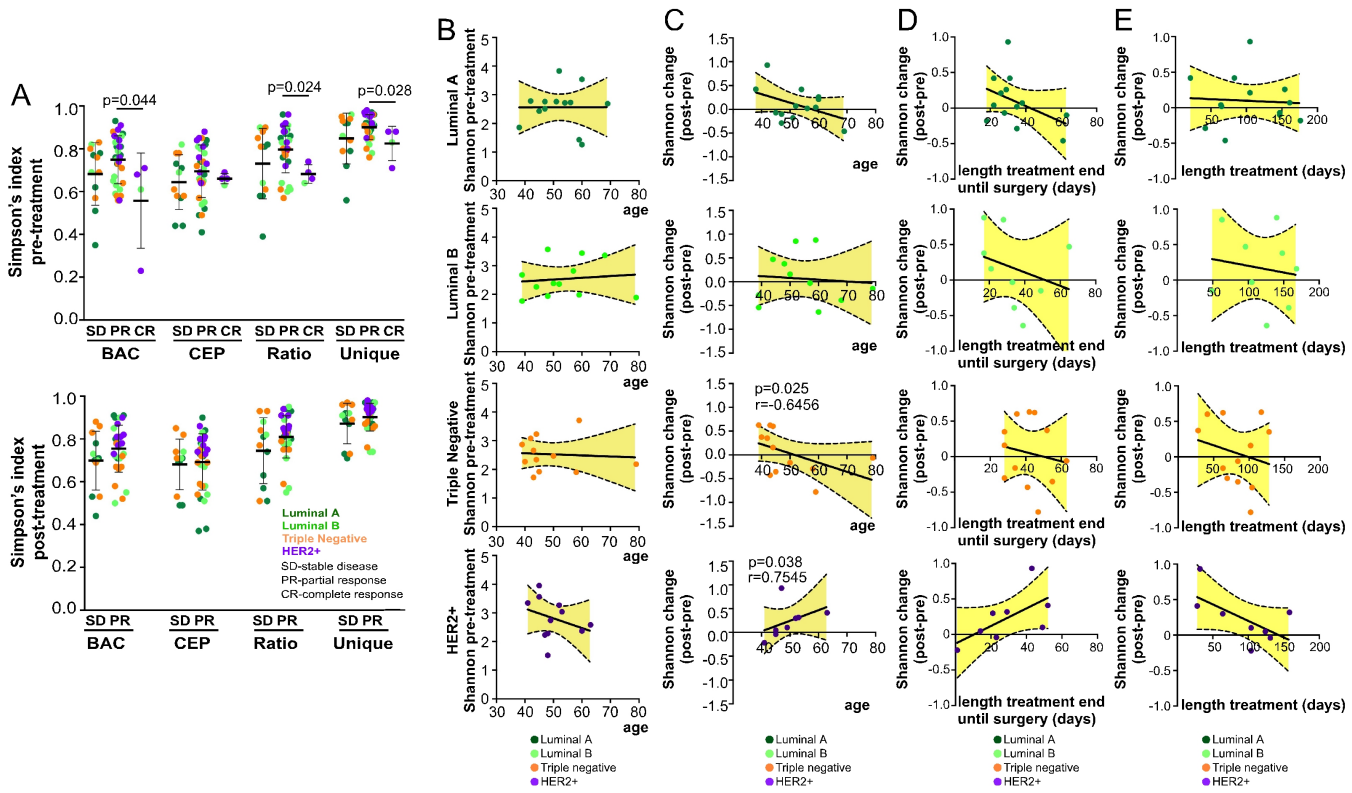

Figure S6. Genetic diversity and clinical outcomes in different tumor subtypes. Related to Figure 7.

(A) Dot plots showing the Simpson's index of diversity before and after treatment in tumors with different response to treatment. The black lines show the mean±SD. Significant p values between groups by Wilcoxon rank sum test are shown.

(B) Correlation between age at diagnosis and Shannon diversity index calculated based on unique 8q24 BAC and Chr8 CEP counts.

(C) Correlation between the age at diagnosis and changes in Shannon index (post-treatment minus pre-treatment). Spearman correlation analysis was used to determine the significance of associations between variables. Significant p values and r values (correlation coefficient) are shown.

(D) Correlation between the length from treatment end until the day of surgery and changes in Shannon index (post-treatment minus pre-treatment). Spearman correlation analysis was used to determine the significance of associations between variables.

(E) Correlation between the full length of treatment and changes in Shannon index (post-treatment minus pre-treatment). Spearman correlation analysis was used to determine the significance of associations between variables. A 95% confidence interval is indicated in yellow.

Supplemental Tables

Table S1. Pathology and clinical information of the samples used in this study. Related to Figure 1. IDC, invasive ductal carcinoma; DCIS, ductal carcinoma in situ; ILC, invasive lobular carcinoma; bx, biopsy; AC, antracyclins; FEC, Fluorouracil Epirubicin Cyclophosphamide; pPR-pathologic partial response; pSD-pathologic stable disease; pCR-pathologic complete response; Tz, Trastuzumab; nd, not determined.

Table S2. Shannon and Simpson's index for each tumor before and after treatment based on BAC, CEP, BAC/CEP and unique counts. Related to Figure 1. Significant p values are highlighted in bold red and italics. NA, not available.

Table S3. Diversity index in each of the three different areas of each tumor used for topology studies. Related to Figure 3. Significant p values are highlighted in bold red and italics. NA, not available.

Table S4. Samples used for simulation of tumor growth. Related to Figure 6. The specific treatment length and time between the core biopsy and the surgery were used as variables to infer tumor evolution.

Table S5. Correlations of the Shannon and Simpson's diversity index with clinical variables. Related to Figure 7. The p values for the correlations between the diversity indices estimated based on BAC, CEP, BAC/CEP and unique values with the different clinical variables for each subtype are shown.

Supplemental Movies

Video S1. Simulation of tumor evolution in a luminal A tumor. Related to Figure 6.
We developed a stochastic computational model of cellular proliferation and death utilizing the tumor topology and Ki67 data obtained from each patient. The video shows a representative example of the simulation in a Luminal A tumor.

Supplemental experimental procedures

Human breast cancer samples

Formalin-fixed paraffin embedded breast tumor samples were collected from patients undergoing neoadjuvant chemotherapy at the Hospital Clinic of Barcelona (Barcelona, Spain), the Royal Marsden Hospital (London, UK), and the Oslo University Hospital (Oslo, Norway) using protocols approved by the Institutional Review Boards of the respective hospitals. Samples were de-identified prior to analysis. Tumor histology and expression of standard biomarkers (ER, PR, and HER2) were evaluated at the time of diagnosis according to ASCO/CAP guidelines (Hammond et al., 2010).

ImmunoFISH analyses

The detection of the copy number gain for 8q24 and the centromeric region of chromosome 8 was performed using whole sections of formalin fixed paraffin embedded (FFPE) human breast cancer tissue. The tissues were dewaxed in xylene and hydrated in a series of ethanol. After heat-induced antigen retrieval overnight at 70°C in citrate buffer (pH 6), the digestion with pepsin 0.005% was performed in a slide warmer at 37°C for 10-20 minutes depending on the sample. The immunostaining for the next antibodies diluted 1/100 in 5% goat serum was performed at room temperature and sequentially to avoid cross-reaction between them as follows: CD44 (Neomarkers, clone 156-3C11, mouse monoclonal IgG2) for 1 hour, biotin conjugated rabbit anti-Mouse IgG2a (Life Technologies Cat#61-0240) for 30 minutes, CD24 (NeoMarkers, clone SN3b, mouse monoclonal IgM) for 1 hour, pacific blue conjugated-streptavidin (Life Technologies Cat#S-11222) and Alexa Fluor 647 goat anti-mouse IgM (Life Technologies Cat# A-21238). The samples were then fixed in Carnoy's for 10 minutes and dehydrated in a series of ethanol. The BAC probes were labeled with SpectrumOrange (Vysis) using a Nick Translation (Abbot Molecular) according to the manufacturer's recommendations, mixed with the corresponding centromeric probe (CEP probes) (Vysis), diluted in hybridization buffer and applied to each sample. The denaturalization was performed in a slide warmer at 75°C for 7 minutes, and then the samples were incubated in a humid chamber for 20 hours at 37°C. Finally the samples were washed with different stringent SCC buffers, air-dried and protected for long storage with ProLong Gold (Life Technologies). Different immunofluorescence images from multiple areas of each sample were acquired with a Nikon Ti microscope attached to a Yokogawa spinning-disk confocal unit, 60x plan apo objective, and OrcaER camera controlled by Andor iQ software.

Immunofluorescence analysis of cellular phenotypes and proliferation

Multicolor immunofluorescence for CD44, CD24, and Ki67 (DAKO, clone MIB-1, mouse monoclonal IgG1) was performed using whole sections of formalin fixed paraffin embedded breast tumor samples by sequential staining after antigen retrieval in citrate buffer (pH 6): monoclonal (IgG2a) antibody anti-CD44 (1:100 dilution) for one hour; goat anti-mouse IgG2a Alexa555-conjugated (Life Technologies, 1:100 dilution) for 30 minutes; monoclonal (IgM) antibody anti-CD24 (1:100 dilution) and polyclonal antibody anti-Ki67 (Abcam, 1:50 dilution) for one hour; goat anti-rabbit Alexa 488-conjugated (Life Technologies, 1:100 dilution) and Goat anti-mouse IgM Alexa Fluor 647 conjugated for 30 minutes. The samples were washed twice with PBS-Tween 0.05% between each incubation and protected for long-term storage with Vectashield HardSet Mounting Medium with DAPI (Vector laboratories, cat #H-1500). Before image analysis, the samples were stored at -20°C for at least 48 hours. Different immunofluorescence images from multiple areas of each sample were acquired as described

before. The frequency of $Ki67^+$ cells was scored by counting an average of 200 cells of each cell phenotype. The frequency of each cell phenotype (i.e., $CD44^+CD24^-$, $CD44^+CD24^+$, $CD44^-CD24^+$, and $CD44^-CD24^-$ cells) was calculated by counting an average of 300 cells in each sample. The differences between the frequency of $Ki67^+$ and between each cell subpopulation before and after treatment were analyzed by two-sided Wilcoxon matched-pairs signed rank test, applying the method of Pratt when two paired cases had the same value.

Statistical analyses

Genetic diversity was determined essentially as described (Park et al., 2010), but we calculated diversity indices based on copy number counts for (1) 8q24 BAC, (2) chromosome 8 centromeric, (3) ratio of 8q24/CEP counts, and (4) unique 8q24 and BAC count combinations; for example, a cell with 3 BAC and 3 CEP counts (ratio=1) will be considered as a unique cell type and different from a cell with 5 BAC and 5 CEP counts (ratio=1). Statistical differences in pre- and post-treatment BAC and CEP counts were calculated through 100,000 iterations of bootstrapping the BAC and CEP counts from the larger cell population (either pre- or post-treatment) and comparing the mean count of each bootstrap repetition against the mean count of the smaller cell population. Statistical differences in the BAC and CEP counts between adjacent cells pre- and post-treatment were calculated through 100,000 iterations of bootstrapping the absolute difference in BAC and CEP counts of adjacent cells from the larger cell population (either pre- or post-treatment) and comparing the mean absolute difference in counts of each bootstrap repetition against the mean absolute difference in count of the smaller cell population. Correlations between the changes in the frequency of Ki67+ cells and in cell subpopulations were assessed using Spearman's rank-based coefficient. Diversity indices were compared across the three topologically distinct areas of each tumor using the Kruskal-Wallis test. The association between diversity indices and clinical variables was assessed using the Wilcoxon test for categorical clinical variables (such as response) and a permutation test based on Spearman's rank correlation for continuous clinical variables (such as size). Since the number of cases analyzed was limited to 45, we were unable to performed\ multivariate analyses. Statistical differences in pre- and post-treatment BAC and CEP counts were evaluated using the achieved significance level (ASL) method (Efron and Tibshirani, 1993). This amounts to using 100,000 iterations of bootstrapping the BAC and CEP counts from the larger cell population (either pre- or post-treatment) and comparing the mean counts of each bootstrap repetition against the mean count of the smaller cell population. Statistical differences in the BAC and CEP counts between adjacent cells pre- and post-treatment were calculated through 100,000 iterations of bootstrapping the absolute difference in BAC and CEP counts of adjacent cells from the larger cell population (either pre- or post-treatment) and comparing the mean absolute difference in counts of each bootstrap repetition against the mean absolute difference in count of the smaller cell population. Statistical differences in the BAC and CEP counts between adjacent cells pre- and post-treatment were also calculated using ASL, through 100,000 iterations of bootstrapping the absolute difference in BAC and CEP counts of adjacent cells from the larger cell population (either pre- or post-treatment) and comparing the mean absolute difference in counts of each bootstrap repetition against the mean absolute difference in count of the smaller cell population.

Topology analysis

For the analysis of the topological distribution of cellular subsets, 3x3 images (corresponding to 71,678 μm^2 area) were obtained using 60x plan apo objective with 5% overlap between areas in order to be able to assemble them into one montage. Three different 3x3 montages were acquired from distant regions in each sample to control for spatial heterogeneity. The loci for the BAC and CEP probes were automatically detected using previously described algorithms

(Itzkovitz et al., 2012). Cellular phenotype was determined manually based on immunofluorescence as described above. Both tumor and stromal cells were analyzed, and both the signals and the coordinates for each cell were recorded; however further analysis was restricted to solely non-stromal tumor cells. For each patient, we determined the distribution of BAC and CEP counts both across all cell phenotype and independently for each phenotype. Statistical differences in the distribution of BAC and CEP counts pre- and post-treatment were then determined through bootstrapping. Additionally, to assess the spatial distribution of different cell phenotypes and the topological genomic diversity we also focused on neighboring cells, which we defined as cells for which the shortest distance between cell boundaries is smaller than 10% of the average cell radius in each analyzed field of view. Significance of the differences between neighboring cells was also determined through bootstrapping. We generated 100,000 bootstrap samples of the absolute difference in BAC and CEP counts between neighbors drawn from the larger of either the pre- or post-treatment cell populations and comparing the mean absolute differences in counts of each bootstrap repetition against the mean absolute difference in counts of the smaller cell population. The fraction of homotypic neighbors was calculated by counting the number of pairs of neighboring cells where both cells were the same phenotype and dividing by the total number of pairs of neighboring cells. To assess the significance of this fraction independently in both pre- and post-treatment samples, we used permutation testing. We determined the fraction of homotypic neighbors in the actual sample and compared this fraction against the fraction of homotypic neighbors in 100,000 randomized ensembles in which the cell phenotypes were randomly shuffled. The null hypothesis of this permutation test is that the pattern of homotypic neighbors seen in the sample falls within the distribution expected under random migration. A two-tailed p-value was used to determine whether the fraction detected fell outside the expected range. However, permutation testing cannot test whether the pattern detected is the result of cell division depositing similar phenotype cells adjacent to one another with no migration. For this, we devised an agent-based model and simulated cell division under the chemotherapy regime undergone by each patient and compared the homotypic neighbor fraction in the post-treatment samples against 100 simulations of cell proliferation generated from the pre-treatment samples. As there are three montages for each patient, each patient simulation had three initialization states: comparison against the post-treatment samples were done on a patient by patient basis by combining together the counts of homotypic pairs of neighboring cells in each set of simulations and dividing by the total number of neighboring cells in all three simulations combined.

Modeling and simulation of tumor cell growth

We designed a two-dimensional agent-based model of cellular proliferation and death within the confines of the region 3x3 montages using the Cancer, Heart, and Soft Tissue Environment (Chaste) software package (Pitt-Francis et al., 2008). Each cell was defined in space by a Cartesian point and an age-dependent radius surrounding it. Each cell also included a temporal component, represented by a stochastic length cell cycle, where the cell cycle consisted of both the proliferating and non-proliferating phases. Since the goal of this study was to investigate the extent of diversity and patterns of topology within a sample, we used cellular phenotypes as the primary parameter governing the length of the cell cycle for each cell, and thus by proxy the size of each cell. Thus, the initialization state for each simulation consisted of the cell coordinates for each cell in the pre-treatment samples, an estimation of the age of each cell based on its area, and the phenotype of each cell.

The average length of the cell cycle for individual cell phenotypes and patients was dependent on the Ki67 values determined for each patient and phenotype in the following manner: cell cycle length = t_{nonG0}/Ki67, where t_{nonG0} is the time spent by cells not in G_0 phase; this parameter was taken to be a constant. We calculated t_{nonG0} as the average cell cycle length

multiplied by the average Ki67 value. In our data, we averaged Ki67 values across all phenotypes and patients and found this average to be 23.24. The average cell cycle length was taken from cell line experiments, where in multiple cell lines the average cell cycle length remained relatively constant centered around 29.29 hours. Use of these values resulted in t_{nonG0} = 6.8 hours. Once t_{nonG0} was determined, we calculated the average cell cycle length for a given patient and cellular phenotype. For each cell, the length of its cell cycle was determined stochastically upon birth and then fixed from that point forward. Once a cell reached the end of its cell cycle, given it did not die previously, it underwent a cell division event. In the initial simulations, the daughter cells resulting from a cell division maintained the same phenotype as the precursor cell; later simulations also tested the effects of stochastically switching the phenotype of daughter cells at various rates.

In simulations, the area of each cell ranged from π/8 to π/4 units^2 and grew linearly following area(t) = π/8* (1 + (t - t_{birth})/cell cycle length). These values were selected to allow for a cell diameter of 1.0 units at maximum. Upon cell division, an axis of division was randomly selected passing through the center of the cell. The area of the mother cell was then evenly split between the two daughter cells. In the space of the simulation, two cells that resided less than the sum of their radii apart from each other (i.e. overlapping cells) exerted a force upon each other that grew linearly with the size of the overlap. As cells increased in size, and since cell divisions conserved area, but not diameter, cells could come in contact and pushed each other away, thus altering neighboring cell patterns during each simulation.

Cell migration was accounted for by allowing cells to move stochastically during each cycle. The range of motion was limited to the equivalent of 125 microns/hour for both $CD44^+CD24^-$ and $CD44^+CD24^+$ cells and 25 microns/hour for both $CD44^-CD24^+$ and $CD44^-CD24^-$ cells. These rates were based on the migration observed of a photobleack 250x250 micron area over the course of 24 hours (Kedrin et al., 2008). These factors were normalized according to the average measured cell diameter of the patient and prescribed the maximum magnitude of the random movement of the cell.

Simulations were run through three phases of proliferation. During each phase, average cell cycle lengths for each cellular phenotype were recalculated from updated Ki67 and apoptotic index data. The first phase consisted of the period of time between the biopsy and the start of chemotherapy. Cell proliferation occurred at the rates determined by the pre-treatment Ki67 data. The probability of cell death per unit time for each phenotype was selected to maintain a roughly constant population size. We chose these values for cell death since rates of apoptosis correlate well with proliferation and the montage of visualized cells did not consist of cells crowded together as would be consistent with high growth rates. During treatment, proliferation rates were lowered by 5% and the rate of cell death was adjusted accordingly. This choice of treatment effect was selected for two reasons: first, by fitting of the number of cells at the end of each simulation to the number of cells seen in the post-treatment samples, and second, due to evidence of a decrease in proliferation with anthracyclines with a corresponding decrease in apoptotic index (Burcombe et al., 2006). The last phase consisted of the period of time between the end of chemotherapy and surgery. Cell proliferation in this phase occurred at the rates determined by the post-treatment Ki67data. Again, the probability of cell death per unit time for each phenotype was selected to maintain a roughly constant population size.

Supplemental references

Burcombe, R., Wilson, G.D., Dowsett, M., Khan, I., Richman, P.I., Daley, F., Detre, S., and Makris, A. (2006). Evaluation of Ki-67 proliferation and apoptotic index before, during and after neoadjuvant chemotherapy for primary breast cancer. Breast Cancer Res *8*, R31.

Efron, B., and Tibshirani, R.J. (1993). An introduction to the bootstrap (London: Chapman & Hall).

Hammond, M.E., Hayes, D.F., Dowsett, M., Allred, D.C., Hagerty, K.L., Badve, S., Fitzgibbons, P.L., Francis, G., Goldstein, N.S., Hayes, M.*, et al.* (2010). American Society of Clinical Oncology/College Of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology *28*, 2784-2795.

Itzkovitz, S., Lyubimova, A., Blat, I.C., Maynard, M., van Es, J., Lees, J., Jacks, T., Clevers, H., and van Oudenaarden, A. (2012). Single-molecule transcript counting of stem-cell markers in the mouse intestine. Nat Cell Biol *14*, 106-114.

Kedrin, D., Gligorijevic, B., Wyckoff, J., Verkhusha, V.V., Condeelis, J., Segall, J.E., and van Rheenen, J. (2008). Intravital imaging of metastatic behavior through a mammary imaging window. Nat Methods *5*, 1019-1021.

Park, S.Y., Gönen, M., Kim, H.J., Michor, F., and Polyak, K. (2010). Cellular and genetic diversity in the progression of in situ human breast carcinomas to an invasive phenotype. The Journal of Clinical Investigation *120*, 636-644.

Pitt-Francis, J., Bernabeu, M.O., Cooper, J., Garny, A., Momtahan, L., Osborne, J., Pathmanathan, P., Rodriguez, B., Whiteley, J.P., and Gavaghan, D.J. (2008). Chaste: using agile programming techniques to develop computational biology software. Philosophical transactions Series A, Mathematical, physical, and engineering sciences *366*, 3111-3136.