Molecular Oncology

FEBS PRESS science publishing by scientists

Intratumor heterogeneity defines treatment-resistant HER2+ breast tumors

Inga H. Rye[1], Anne Trinh[2,†], Anna B. Sætersdal[3], Daniel Nebdal[1], Ole Christian Lingjærde[1,4], Vanessa Almendro[5,‡], Kornelia Polyak[5], Anne-Lise Børresen-Dale[1,6], Åslaug Helland[1,3,6], Florian Markowetz[2] and Hege G. Russnes[1,7]

1 Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Norway
2 Cancer Research UK, Cambridge Institute, University of Cambridge, UK
3 Department of Oncology, Oslo University Hospital, Norway
4 Biomedical Informatics Division, Department of Computer Science, University of Oslo, Norway
5 Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA
6 Department of Clinical Medicine, University of Oslo, Norway
7 Department of Pathology, Oslo University Hospital, Norway

Keywords
breast cancer; HER2; heterogeneity; *in situ* analysis; outcome; therapy response

Correspondence
Hege G. Russnes, Department of Cancer Genetics, Institute for Cancer Research, Oslo University Hospital Radiumhospitalet, Postboks 4953 Nydalen, 0424 Oslo, Norway
E-mail: hege.russnes@rr-research.no

Present addresses
[†]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA
[‡]Vertex Pharmaceuticals, Boston, MA, USA

Inga H. Rye and Anne Trinh contributed equally to this article
Florian Markowetz and Hege G. Russnes shared last author

(Received 3 April 2018, revised 6 July 2018, accepted 26 July 2018, available online 21 September 2018)

doi:10.1002/1878-0261.12375

Targeted therapy for patients with HER2-positive (HER2+) breast cancer has improved overall survival, but many patients still suffer relapse and death from the disease. Intratumor heterogeneity of both estrogen receptor (ER) and HER2 expression has been proposed to play a key role in treatment failure, but little work has been done to comprehensively study this heterogeneity at the single-cell level. In this study, we explored the clinical impact of intratumor heterogeneity of ER protein expression, HER2 protein expression, and *HER2* gene copy number alterations. Using combined immunofluorescence and *in situ* hybridization on tissue sections followed by a validated computational approach, we analyzed more than 13 000 single tumor cells across 37 HER2+ breast tumors. The samples were taken both before and after neoadjuvant chemotherapy plus HER2-targeted treatment, enabling us to study tumor evolution as well. We found that intratumor heterogeneity for *HER2* copy number varied substantially between patient samples. Highly heterogeneous tumors were associated with significantly shorter disease-free survival and fewer long-term survivors. Patients for which *HER2* characteristics did not change during treatment had a significantly worse outcome. This work shows the impact of intratumor heterogeneity in molecular diagnostics for treatment selection in HER2+ breast cancer patients and the power of computational scoring methods to evaluate *in situ* molecular markers in tissue biopsies.

Abbreviations
BAC, bacterial artificial chromosome; CAP, College of American Pathologists; Cent, centromere; CN, copy number; CR, complete response; ER, estrogen receptor; FEC, fluorouracil, epirubicin, and cyclophosphamide; FFPE, formalin-fixated paraffin-embedded; FISH, fluorescence *in situ* hybridization; *HER2*amp, amplified *HER2* copy number; *HER2*gain, gain *HER2* copy number; HER2, human epidermal growth factor receptor 2; *HER2*norm, normal *HER2* copy number; IntClust, integrated cluster; K-L, Kullback–Leibler divergence score; pCR, pathological complete response; PgR, progesterone receptor; RECIST, The Response Evaluation Criteria In Solid Tumors; SI, Shannon index.

1. Introduction

Breast cancer is divided into several distinct subtypes, and the expression level of estrogen receptor (ER), progesterone receptor (PgR), and human epidermal growth factor receptor 2 (HER2) is fundamental for treatment decision and prognosis of the disease. The HER2-positive (HER2+) tumors account for 15–20% of all breast cancers and are characterized by either overexpression of HER2 protein and/or increased copy number of the *HER2* gene. With the introduction of HER2-targeted therapy, such as trastuzumab and lapatinib, the overall survival for both early- and late-stage disease has increased (Baselga *et al.*, 2012; Cortazar *et al.*, 2014; Gianni *et al.*, 2010; Guarneri and Conte, 2004; Viani *et al.*, 2007).

Breast cancer was one of the first solid cancer types where comprehensive molecular profiling revealed robust molecular subtypes (Curtis *et al.*, 2012; Perou *et al.*, 2000), and HER2+ tumors are found within several subtypes. By PAM50 classification, HER2+ tumors are mainly found in the HER2-enriched but also in the luminal B and luminal A subtypes (Parker *et al.*, 2009). Similarly, in the 10 integrated cluster (IntClust) subtypes, the HER2+ tumors dominate group 5 but are also found within other subtypes (Curtis *et al.*, 2012). The notion that HER2+ tumors do not represent a separate subtype but a wider biological spectrum was strengthened by a recent study identifying four different subtypes of HER2+ breast carcinomas based on gene expression signatures (Ferrari *et al.*, 2016).

Pathologists have noticed the presence of cell-to-cell variation in HER2+ tumors since the introduction of biomarkers into diagnostic routine. In early-stage HER2+ breast cancer, neither the average level of HER2 protein expression nor the average level of *HER2* gene amplification across a tumor seem to have an impact on therapy response (Wolff *et al.*, 2013; Zabaglo *et al.*, 2013). However, as reflected by the comprehensive College of American Pathologists (CAP) guidelines, some HER2+ tumors display intratumor variation in *HER2* copy number (*HER2* CN) levels. The ASCO/CAP guidelines from 2013 state that breast cancers with aggregations of *HER2*-amplified cells (with *HER2*/CEP17 ratio > 2.0 or more than six *HER2* copies per cell) in more than 10% of the tumor must be quantified and reported separately (Wolff *et al.*, 2013). The clinical challenge of such a definition has been addressed for HER2 equivocal cases (Bartlett *et al.*, 2011; Lewis *et al.*, 2005), but the clinical impact of intratumor heterogeneity within nonequivocal HER2+ tumors is less studied (Arena *et al.*, 2013; Gulbahce *et al.*, 2016). The regional variation of *HER2* gene amplification has been studied to some extent (Lee *et al.*, 2014; Seol *et al.*, 2012), and heterogeneity of *HER2* CN even in tumors classified as nonamplified was recently described (Buckley *et al.*, 2016), but there are very few studies addressing this at the single-cell level estimating multiple biomarkers from a high number of cells.

To investigate and quantify the heterogeneity of HER2+ carcinomas by using single cell investigation, we performed detailed *in situ* analyses on samples from a Norwegian observational study (RA-HER2), comprised of 37 HER2+ patients treated in a neoadjuvant setting with trastuzumab and chemotherapy where both response data and clinical follow-up were available. For objective assessment of the molecular *in situ* markers, we used GOIFISH, a software for image analysis developed to objectively score both immunofluorescence and FISH signals from numerous individual tumors cells (Trinh *et al.*, 2014). With this quantitative approach, we examined 103 images and more than 13 000 cells showing the clinical impact of different types of genomic and phenotypic intratumor heterogeneity in HER2+ breast cancer.

2. Materials and methods

2.1. Patient samples

Breast cancer patients diagnosed with HER2+ tumors between 2004 and 2010 who qualified for neoadjuvant treatment according to the national guidelines were included in this prospective observational trial. Informed and written consent was obtained from all patients, and the study was approved by the Regional Ethical Committee (South-east of Norway, no. S-06495b). The study methodologies conformed to the standards set by the Declaration of Helsinki.

The clinical characteristics are shown in Table S1. All 37 patients received combinatorial neoadjuvant treatment of four cycles of fluorouracil, epirubicin, and cyclophosphamide (FEC) followed by four cycles of taxanes in combination with the HER2-targeted monoclonal antibody trastuzumab. The average neoadjuvant treatment period was 6 months (range 3–10 months). The Response Evaluation Criteria In Solid Tumors (RECIST) (Nishino *et al.*, 2010) was used to score the effect of the neoadjuvant treatment, with pathological complete response (pCR) defined as no invasive tumor cells in primary tumor region or lymph nodes after neoadjuvant treatment. Non-pCR

was defined as the presence of residual invasive tumor cells in primary tumor region or lymph nodes (Table S1). After neoadjuvant treatment, 12 patients had pathological complete response (pCR), and among the 25 patients with noncomplete pathological response (non-pCR), a variation in tumor reduction from almost complete response to no reduction in tumor size was observed (Table S1).

Formalin-fixated paraffin-embedded (FFPE) tumor tissue from the 37 patients was collected from several hospitals throughout Norway. FFPE core needle biopsies from the time of diagnosis and FFPE surgical biopsies after neoadjuvant treatment were available for analysis. In addition, FFPE tissue biopsies from later distant metastases were available for three patients.

2.2. IFISH analyses

The FISH probes for *HER2* were made from the BAC clones RP11-94L15 and RP11-909L6, and FISH probes for centromere 17 (cent17) were made from BAC clones RP11-170N19 and RP11-909L10. The BAC probes were isolated according to the instructions from the manufacturer and labeled with fluorescent UTPS by nick translation. Primary antibody recognizing estrogen receptor (clone 6G11) was detected with secondary antibody IgG-conjugated Alexa Fluor 594. The HER2 (CB11) primary antibody was detected with a secondary biotinylated antibody and visualized using streptavidin-conjugated Alexa Fluor 488 antibody in order to visualize the protein expression of ER and HER2. A detailed IFISH protocol including antibody and BAC catalogue numbers is described in the previous publication (Trinh *et al.*, 2014). The tissue samples were mounted with DAPI counterstain, and areas of interest were photographed with 25 z-stacks in a Zeiss Axiovision M1 microscope. The areas with a high number of tumor cells and with high quality of IFISH staining were selected for photography. The number of biopsies, areas, and tumor cells analyzed per sample are listed in Table S2.

2.3. Analysis by GOIFISH

We previously developed and validated the software GOIFISH (Trinh *et al.*, 2014), an image analysis pipeline designed to objectively recognize cell types, score protein intensities in distinct cellular compartments (nucleus, cytoplasm, and membranes), count and measure FISH spots/areas and intensities, measure nuclear size, and display topological distributions of the cells and the analyzed parameters. GOIFISH estimates are highly concordant with visual scoring at the single-cell level, and optimal intensity thresholds of 300 and 50 following adjustment by background and perinuclear intensity were used to define HER2-positive and ER-positive cells, respectively, from 12-bit images (Trinh *et al.*, 2014). ER+ patients were identified according to the national guidelines with a cutoff level at 1% positive cells (Helsedirektoratet 2014). The *HER2* copy number (*HER2* CN) level was assessed by measuring the total area of the HER2 probe signals within each nucleus. For cluster analyses to study phenotypic heterogeneity, we assigned each cell within a tumor into one of four phenotypic groups (HER2+/ER+, HER2+/ER−, HER2−/ER+, HER2−/ER−) based on the defined thresholds. To address heterogeneity based on genomic changes, we assigned each cell into one of three *HER2* CN categories as previously defined (Trinh *et al.*, 2014): normal (*HER2*norm), gain (*HER2*gain) or amplified (*HER2*amp). *HER2*norm reflected cells with up to three spots (0–63 pixels), *HER2*gain: three to six spots (64–200 pixels), and *HER2*amp: > 6 spots (> 200 pixels). Additionally, we considered the combined effect of both phenotype and genotype and classified each cell into one of twelve groups: HER2+/ER+ *HER2* amp, HER2+/ER+ *HER2* gain, HER2+/ER+ *HER2* norm, HER2+/ER− *HER2* amp, HER2+/ER− *HER2* gain, HER2+/ER− *HER2* norm, HER2−/ER+ *HER2* amp, HER2−/ER+ *HER2* gain, HER2−/ER+ *HER2* norm, HER2−/ER− *HER2* amp, HER2−/ER− *HER2* gain, or HER2−/ER− *HER2* norm.

Five samples were excluded in comparisons between pre- and post-treatment samples: three due to low numbers of tumor cells present after neoadjuvant therapy, and two samples had insufficient IFISH staining due to technical problems [immunofluorescence and genomic (FISH) analyses were performed separately].

2.4. Spatial distribution of HER2 amplification within tumor nuclei

Three spatial patterns of *HER2* FISH signals within individual tumor cell nuclei were identified by visual inspection. Cells demonstrating a tight cluster of multiple signals were called ‘cluster’, cells with distinct and separate signals were called ‘scatter’, and those with both patterns were annotated as ‘mix’. The *HER2* spatial distribution pattern was scored in 100 tumor cells from each biopsy (from both pre- and post-treatment samples) and in the three samples from metastases. These single-cell scores were collapsed to the patient level by (a) computing the frequency of each pattern and (b) using a 70% majority cutoff to describe a class for each patient. If a tumor did not show one

particular dominant pattern, it was considered as 'heterogeneous'. In the pretreatment samples, 10 were dominated by 'cluster' cells, six with 'mix', eight with 'scatter', and 13 samples were 'heterogeneous' with regard to spatial patterns.

2.5. Statistical analyses

The Welch *t*-test was used to determine differences in intensity distributions, and Fisher's exact *t*-test was used to calculate the differences between groups of patients. Survival curves were constructed using the Kaplan–Meier method, using both disease-free survival (i.e., time to metastasis) and overall breast cancer-specific survival as events. Differences in survival between groups of patients were studied by univariate cox regression analyses and expressed as hazards ratios with 95% confidence intervals using continuous variables. The Shannon index (SI) was used as measure for heterogeneity of the defined phenotypic and genomic groups and combined phenotypic and genomic groups (Shannon, 1948), and the mean Shannon index for each cluster group was used to determine the differences in heterogeneity between clusters.

To measure the change in the clonal composition during neoadjuvant therapy, the Kullback–Leibler divergence index (K-L) (Kullback and Leibler, 1951) was used to compare the cell-type distributions before and after treatment. Briefly, this describes the divergence between two populations, such as the phenotypic composition of pre- and post-treatment samples:

$$K\text{-}L = -\sum_{i}^{M} P_i \log \frac{Q_i}{P_i}$$

where P_i is the proportion of cells which belong to group i in the pretreatment group and Q_i is the proportion of cells which belong to group i in the post-treatment samples. M indicates the number of discrete groups considered: four for phenotypic change, three for genomic changes, and 12 for the combined change. A high index signifies different clonal compositions in the samples taken after treatment versus the samples taken before. The median of the Kullback–Leibler index was used to divide the samples into two equal sized groups: one group with samples with a high change of *HER2* CN fractions (K-L high) and one group with samples with low change in fractions (K-L low).

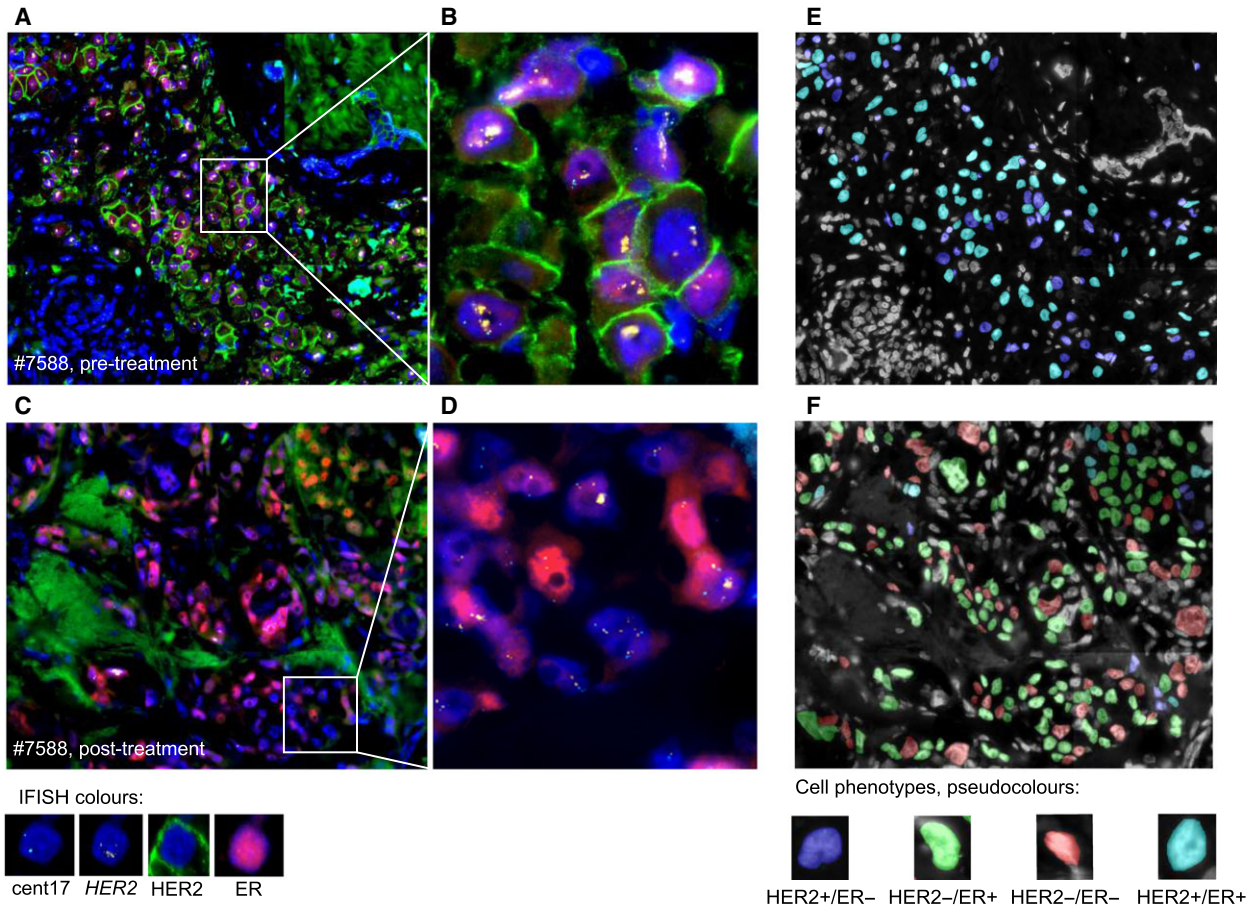
All image analysis was performed in MATLAB (7.12.0 (R2011a), The MathWorks, Natic, MA, USA), and subsequent statistical analyses were performed in R (R Core Team, 2017). All code to reproduce the analyses in this study is available at the following Github Link/ as supplementary information https://github.com/trinhan/HER2heterogeneity.

3. Results

We analyzed more than 13 000 single tumor cells from biopsies taken before treatment (n = 37), after treatment (n = 22) and metastases (n = 3) from 37 HER2-positive (HER2+) breast cancer patients. Single-cell metrics for HER2 and ER expression, *HER2* copy number, and CEP17 copy number were evaluated. This enabled us to evaluate the heterogeneity of the markers both across tumors but also within the individual tumors at different time points, as illustrated in Fig. 1A–D. As an example, images of pre- and post-treatment biopsies from patient 7588 show the protein and FISH staining of the tumor cells. The GOFISH software was used to visualize the spatial distribution of cells with different phenotypic and/or genotypic features, as shown in Fig. 1E,F where each cell is pseudo-colored with regard to HER2 and ER protein expression. Changes in cell populations during therapy are evident; prior to therapy, the tumor had both HER2+/ER+ and HER2+/ER− cells, while in the post-treatment tumor, a new dominant population of HER2−/ER+ cells emerged. The phenotypic change during therapy is further illustrated in Fig. 1G, where each dot represents a tumor cell and the color illustrates the phenotype. Furthermore, there was a substantial reduction of cells with high *HER2* CN after treatment, reflected in Fig. 1G by the size of each dot.

3.1. Intertumor heterogeneity within HER2+ tumors

All images were subjected to the same analyses as for the case shown in Fig. 1, and a substantial variation of marker distribution was seen across the cohort. This is visualized in the compilation of representative images from each of the 37 pretreatment samples shown in Fig. S1. To get a first overview of the cohort, we estimated the mean values of the biomarkers (i.e., measurements from all tumor cells within a sample) and found patients with nonpathological complete response (non-pCR) to have a significantly lower mean copy number of the *HER2* gene compared to patients with pathological complete response (pCR) (Fig. S2A, *t*-test: P = 0.02). No significant difference in mean HER2 and ER protein expression was found. By looking at the same biomarkers and stratifying the patients by disease progression, we found a significantly lower ER expression (P = 0.02) and lower *HER2* CN/cent17



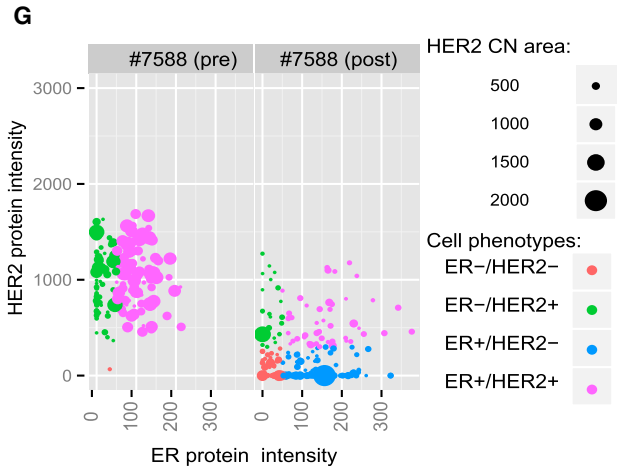


Fig. 1. IFISH images reflecting intratumor heterogeneity before and after treatment. Expression of ER and HER2 protein and copy number of *HER2* gene by IFISH (color code below images) for (A) pretreatment biopsy from patient #7588, (B) magnified image of the outlined area, (C) post-treatment biopsy of patient #7588, and (D) magnified image of the outlined area. Pseudo-colored cell phenotypes of (E) pretreatment biopsy (same area as in A), (F) post-treatment biopsy (same area as in C). (G) Tumor cell heterogeneity before and after treatment for patient #7588, and the scatter plot shows the relationship between ER expression (*X*-axis) and HER2 expression (*Y*-axis) for each of the individual cells. The color reflects the cell phenotype. The size of the dot reflects each cells *HER2* CN level, where a small dot equals fewer copies and a large dot more copies of the *HER2* gene.

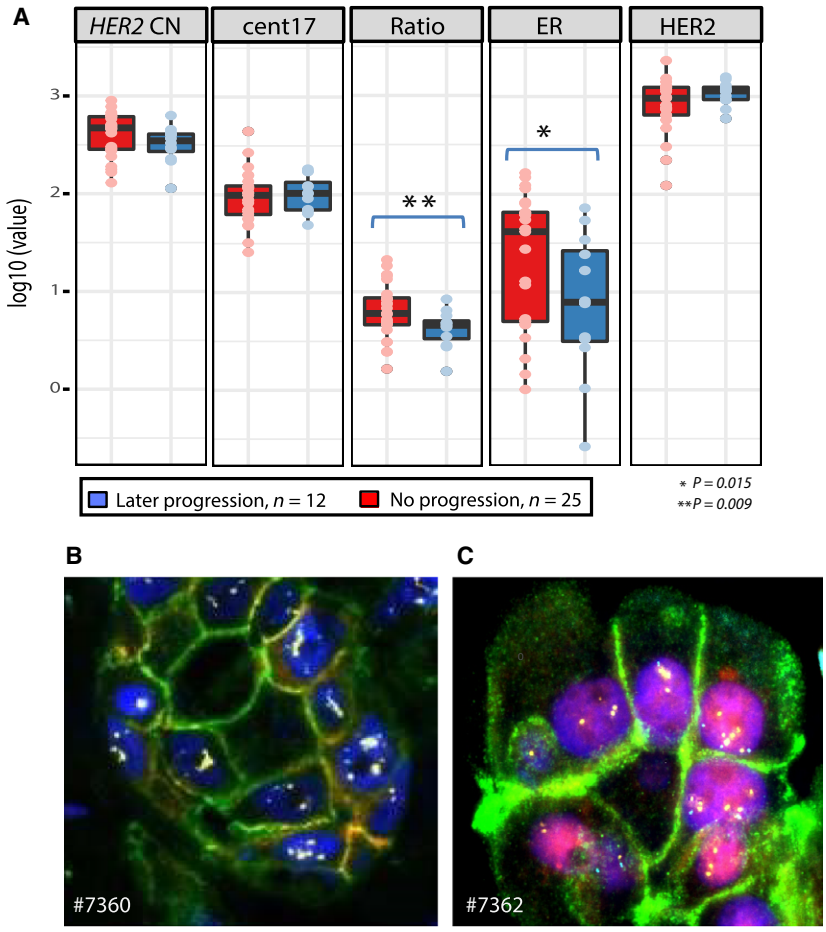


Fig. 2. Biomarker status and later progression of disease. (A) Comparison of GoIFISH measurements (*HER2* copy number (HER2 CN), cent17, ratio (HER2 CN/cent17), ER protein expression, and HER2 protein expression) for all pretreatment biopsies ($n = 37$) stratified by relapse or not after neoadjuvant treatment (Wilcox *t*-test). (B) IFISH image from a patient with later relapse of disease (#7360). The cells were ER−, HER2+ with amplification of *HER2* (same color scheme as in Fig. 1A–D). (C) IFISH image from a patient without later relapse of the disease (#7362). The sample was ER+, HER2+ with gain of *HER2* copies.

CN ratio ($P = 0.009$) in samples from patients with later metastatic disease compared to those without metastasis (Fig. 2A). Figure 2B illustrates the pretreatment cell-type composition in an ER-negative tumor with highly amplified *HER2* CN from a patient which later had progressive disease. The cell composition in an ER-positive tumor with gained *HER2* CN from a patient who has not had progressive disease is shown in Fig. 2C.

Using 1% positive cells as a cutoff level from GOIFISH, we identified 28 patients with ER-positive (ER+) tumors (76%) and nine patients with ER-negative (ER−) tumors (24%). Complete response to neoadjuvant treatment was seen in 7/28 (28%) and 5/9 (55%) patients with ER+ and ER− tumors, respectively. With regard to metastasis, 9/28 (32%) patients with ER+ and 3/9 (33%) patients with ER− tumors developed metastasis (Table S1). Tumors were stratified into four groups based on the percentage of ER+ cells present: ER-negative (< 1%, $n = 9$), low ER (1–10%, $n = 9$), intermediate ER (10–50%, $n = 10$), and high ER (> 50%, $n = 9$). Although not significant, a trend that patients with low or intermediate number of ER+ cells had less local response to treatment was observed, as well as a worse prognosis compared to those with either high-ER or ER-negative tumors (Fig. S2B).

We next sought to determine whether relationship between ER and HER2 protein expression and *HER2* copy number at single-cell level could influence patient outcome. As illustrated by scatterplots in Fig. S3, substantial variation was seen with regard to ER and HER2 protein expression both across tumors and within tumors. In addition, some tumors showed a linear relationship between *HER2* CN and HER2 protein level, but others did not (Fig. S4). In addition, we noticed that the relationship could change during therapy (Figs S3 and S4).

To address the clinical implication of this protein variation, we assigned each cell to one of four categories: HER2+/ER+, HER2+/ER−, HER2−/ER+, or HER2−/ER− (see Materials and methods section). By comparing the fractions of cells with different phenotypes, subsets of tumors with distinct types of phenotypic intratumor heterogeneity were identified.

Hierarchical clustering of the fractions of each cell class revealed three separate groups of tumors. Group P1 contained tumors dominated by HER2+/ER+ cells, while tumors in the cluster group P2 was dominated by HER2+/ER− cells (Fig. 3A and Table S3). IFISH images from two patients representing phenotypic clusters P1 and P2 are shown in Fig. 3B. Patients in cluster group P2 had tumors with negative to intermediate ER expression and were associated with high histological grade (Table S3). They also had a higher frequency of later metastasis, and the Kaplan–Meier curves indicated a worse prognosis, although this was not significant (Fig. 3C, Fig. S5A). Interestingly, P2 was the least heterogenous cluster with a Shannon index (SI) of 0.34, compared to P1 which had SI = 0.66 (Table S4). Cluster group P3 only contained three samples, all dominated by HER2-negative tumor cells. Two of these samples were scored 2+ by IHC (#7619 and #7441); the third sample (#7370) had one HER2-positive and one HER2-negative biopsy prior to therapy.

In contrast to cellular phenotypes, where subpopulations can be dynamic and cells might change

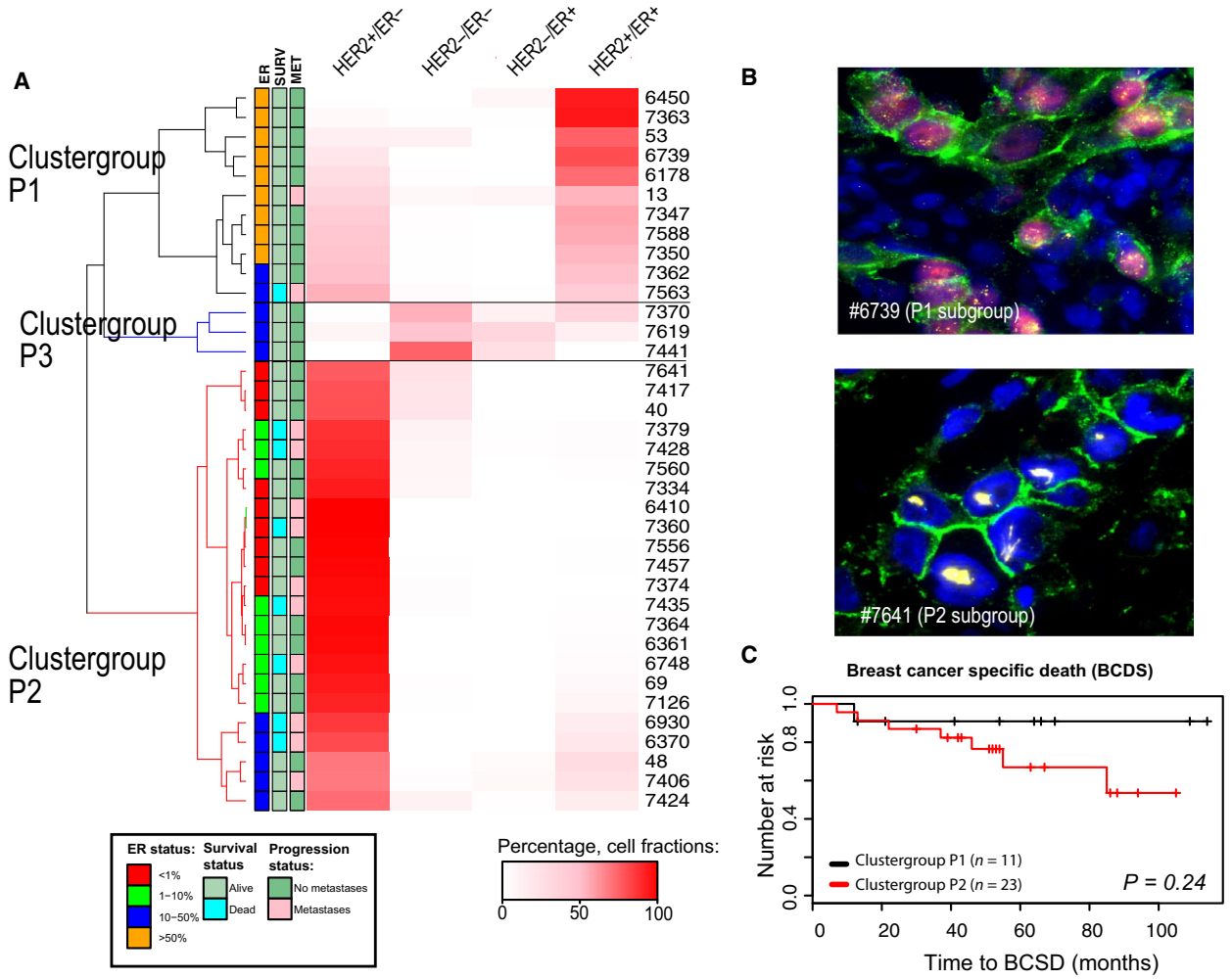


Fig. 3. Identification of subsets of HER2+ breast cancer patients by phenotypic diversity. (A) Unsupervised cluster analysis of the fractions of the phenotypic cell types HER2−/ER−, HER2+/ER−, HER2−/ER+, and HER2+/ER+ in the pretreatment samples ($n = 37$) where the percentage of each cell type (i.e., fraction) is indicated by the color intensity. Two large clusters and one small were identified, where cluster group P1 ($n = 11$) was dominated by HER2+/ER+ cells and cluster group P2 was dominated by HER2+/ER- cells. The smallest cluster group contained three patients whose tumors had mainly HER2− cells. The clinical information for each patient is illustrated by the boxes next to the dendrogram. (B) IFISH image to the left is from pretreatment biopsy from patient #6739 (in cluster group P1) which was dominated by HER2+/ER+ tumor cells. The image to the right is from the pretreatment sample from patient #7641 (cluster group P2) dominated by HER2+/ER− tumor cells. (C) Survival analyses; breast cancer-specific death for the two groups ($P = 0.24$). D) Survival analyses; breast cancer-specific death between patients with different percentage of ER+ cells ($P = 0.14$, log-rank test).

expression levels rapidly in response to treatment, *HER2* copy number (CN) will reflect more persistent cellular subclones. We categorized each cell into one of three levels of *HER2* CN (*norm, gain,* and *amp*) and determined the cellular composition of each tumor (see Materials and methods section). We found some tumors to be dominated by cells with similar copy number level, while other tumors had more heterogeneous cellular composition. Hierarchical clustering identified three groups of tumors with different levels of *HER2* genomic heterogeneity (Fig. 4A, Table S4). The most distinct difference between these three groups was the fraction of cells with *HER2* amplification. The smallest group of tumors (cluster group G1, $n = 6$) had overall low-level *HER2* CN with few cells with *HER2*amp and the highest heterogeneity (SI = 1.2). The second largest group (cluster group G2, $n = 13$) had tumors mainly dominated by cells with *HER2*amp and had a low degree of heterogeneity (SI = 0.6). This was in contrast to the third group (cluster group G3, $n = 16$), which had a high fraction of *HER2*amp cells, but also fractions of *HER2*gain and *HER2*norm cells and overall a high degree of heterogeneity (SI = 0.9). A representative image of each cluster group is shown in Fig. 4B. Interestingly, the patients belonging to cluster G3 displaying high intratumor variation but with *HER2*amp dominating were more likely to experience distant metastases (Table S3) and had the highest risk of disease progression (HR: 14.9, *P*: 0.04, Fig. 4C) but not a significantly increased risk of death by breast cancer (Fig. 4D). However, the groups were not distinguished by other clinical parameters; in particular we were not able to find any significant correlation with treatment response measured by tumor reduction (Table S3).

To investigate the impact of combined phenotypic and genomic heterogeneity, we next assigned each cell within a tumor to one of twelve combined phenotypic–genomic (PG) groups (see Materials and methods section). Three separate groups were identified (Fig. S5A), where cluster PG1 ($n = 9$) was comprised of highly heterogeneous tumors containing both ER+ and ER− cells with varying *HER2* CN levels (amp, gain, and norm) (SI = 1.8). Cluster PG2 ($n = 8$) consisted predominantly of tumors with ER+/HER2+ cells with *HER2*amp (SI = 1.28). The largest group, cluster PG3 ($n = 20$), was also dominated by cells with *HER2*amp with predominantly a ER−/HER2+ phenotype, but many tumors had cells with normal levels or gain of *HER2* CN (SI = 0.99). Patients in PG2 had > 50% ER+ cells, all had high HER2 protein expression (3+), and none had later progression of the disease (Table S3). Although not significant, a trend was observed where patients in the PG1 and PG3 groups had a higher risk of progressive disease and breast cancer-related death than patients in group PG2 (Fig. S5B,C).

3.2 The HER2 spatial organization

During visual investigation of the images, we noticed different spatial patterns of *HER2* amplifications within each nucleus. Some cells had a tight cluster of multiple signals, others had fewer signals scattered within the nucleus and some had a combination (Fig. 5A, see Materials and methods section for more details). We named the nuclear spatial patterns 'cluster', 'scatter', and 'mix'. As intratumor heterogeneity with regard to *HER2* CN levels seemed to have prognostic information, we wanted to address whether the observed differences in spatial organization of the HER2 gene was of clinical importance. As shown in the triangle plots in Fig. 5B, we observed intertumor variation where some samples were dominated by one spatial type (samples in the corners of the triangle plot in Fig. 5B), while others had a more heterogeneous distribution, illustrated by being plotted towards the centre of the triangle. A significant difference in the distribution of samples from patients with and without pathological complete response (pCR) was observed; samples from patients with pCR were most frequently of 'cluster' or 'mix' type, while samples from patients with non-pCR were more heterogeneous and dominated the group characterized by the 'scatter' type of distribution (Fisher's exact test, $P = 0.007$, Table S5A). We found an indication for patients with tumors dominated by 'mixed' spatial type not to have disease progression, in contrast to patients with tumors dominated by 'cluster' or with a combination of the three types (Fig. 5C, Fig. S6A, Table S5B). Interestingly, these spatial distributions were also associated with ER status: ER-negative tumors were found to be frequently of 'cluster' or 'mix' spatial type (Fig. S6B, Table S5C) when stratifying the ER-positive samples into negative, low (1–10%), intermediate (10–50%), and high ER (> 50%). The intermediate ER+ tumors were predominantly of the 'scatter' spatial type, while the ER-negative and ER-low tumors ($P = 0.007$) were predominantly of the 'cluster' spatial type (Fig. 5D, Table S5D).

3.3. Predicting disease progression by measurements of clonal shift during therapeutic intervention

As patients with more heterogeneous tumors (reflected both by ER status and by cellular subclones displaying different types of *HER2* CN) had a higher risk of

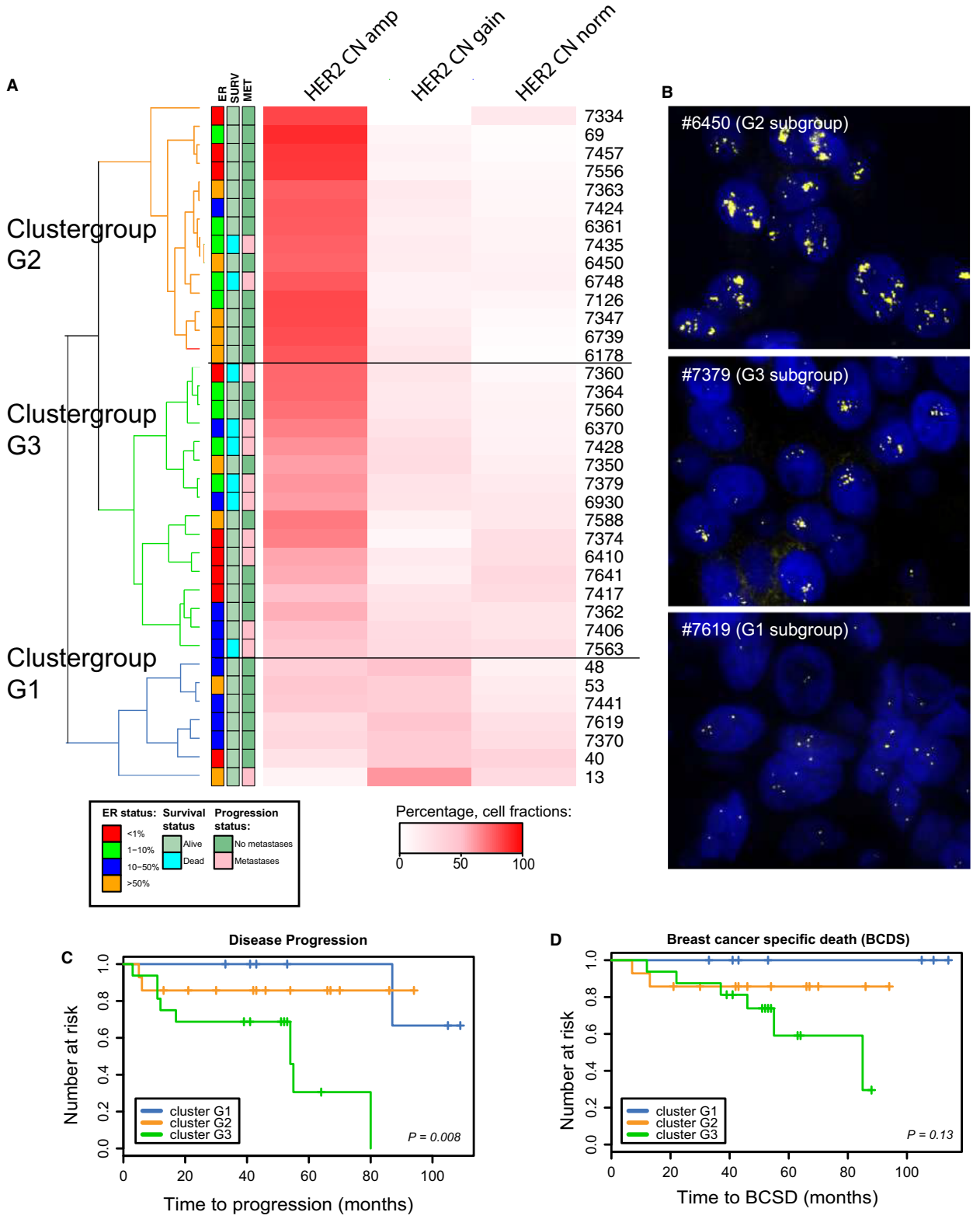

A
ER
SURV
MET
HER2 CN amp
HER2 CN gain
HER2 CN norm
Clustergroup G2
Clustergroup G3
Clustergroup G1
7334
69
7457
7556
7363
7424
6361
7435
6450
6748
7126
7347
6739
6178
7360
7364
7560
6370
7428
7350
7379
6930
7588
7374
6410
7641
7417
7362
7406
7563
48
53
7441
7619
7370
40
13
ER status:
<1%
1–10%
10–50%
>50%
Survival status
Alive
Dead
Progression status:
No metastases
Metastases
Percentage, cell fractions:
0
50
100
B
#6450 (G2 subgroup)
#7379 (G3 subgroup)
#7619 (G1 subgroup)
C
Disease Progression
Number at risk
cluster G1
cluster G2
cluster G3
P = 0.008
Time to progression (months)
D
Breast cancer specific death (BCDS)
Number at risk
cluster G1
cluster G2
cluster G3
P = 0.13
Time to BCSD (months)

Fig. 4. Identification of subsets of HER2+ breast cancer patients by HER2 copy number diversity. (A) Unsupervised clustering based on the fractions of cells with different levels of *HER2* copy number (normal, gain, or amplified). Three clusters (G1–G3) were identified. The clinical information for each patient is illustrated in the boxes next to the dendrogram. (B) FISH (*HER2* CN) images from patient samples representing each of the three cluster groups (G1–G3). The top image is from cluster G2 (patient #6450) and shows a tumor dominated by *HER2* CN amp cell type, the second image is from cluster G3 (patient #7379) and shows a sample with an intermediate fraction of cells with *HER2* CN amp, and the last image is from cluster G1 (#7619) and shows a sample with a high fraction of *HER2 CN* gain and a low fraction of *HER2* CN amp cell types. (C) Survival analyses showed significant differences in risk for progression between the two groups ($P = 0.008$, log-rank test) but not for breast cancer-specific death (D).

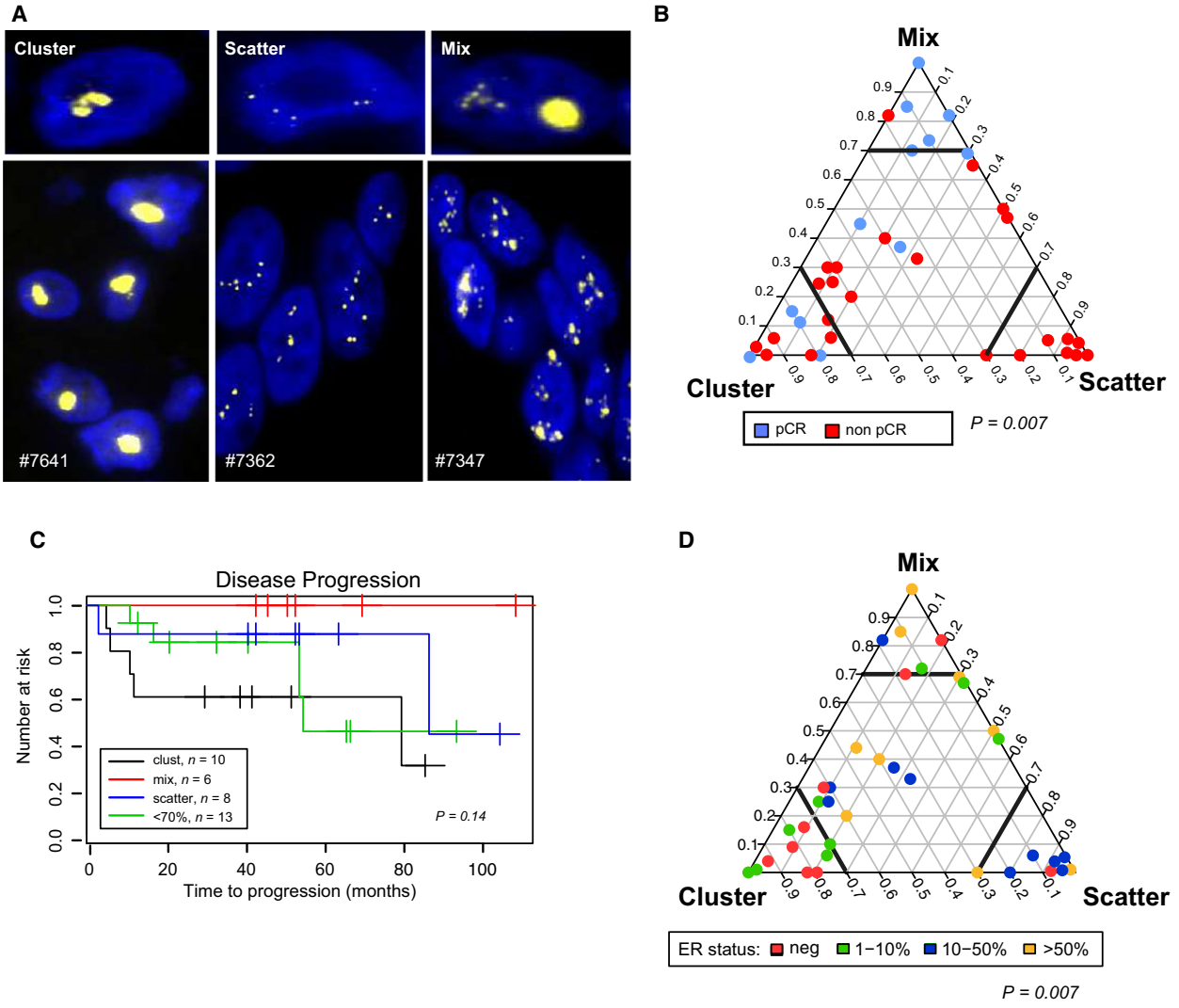


Fig. 5. The spatial organization of the HER2 gene copies within the nuclei. (A) Each cell was categorized as 'cluster', 'scatter', and 'mixed' based on the spatial organization of the *HER2* gene within the nuclei. (B) The spatial organization for the HER2 CN for the pretreatment samples ($n = 37$); in the triangle plot, each corner represents homogenous cell population (100% of cells have one of the spatial patterns). Samples from patients with complete response are colored in blue and from patients with noncomplete response are colored in red (Fisher's exact test, $P = 0.007$). C) Kaplan–Meier curve for time to disease progression for the categorized spatial organization 'cluster', 'mix', 'scatter', and the '< 70%' groups. D) The spatial organization for the pretreatment samples where samples are colored by ER expression level (percentage of positive cells). ER-negative samples are colored in red, ER low (1–10%) colored in green, ER intermediate (10–50%) colored in blue, and ER high (> 50%) colored in yellow (Fisher's exact test, $P = 0.007$).

relapse, we next studied the population dynamics, that is, which cell types did or did not respond to therapy and whether dynamics during therapy can reveal patients with better prognosis or not. We assessed change in tumor composition in 20 patients who did not achieve complete pathological response. To objectively address the dynamics of cell populations during neoadjuvant treatment, we calculated changes in fractions of the predefined cell types (phenotypic and *HER2* CN and the combined phenotypic/*HER2* CN cell types) before and after therapy using the Kullback–Leibler (K-L) divergence index. Figure 6A illustrates the change in *HER2* CN cell types (delta calculated by comparing fractions before and after therapy) sorted according to decreasing K-L index. Patients with low K-L index had a significantly increased risk of breast cancer-related death compared to patients with high K-L index, indicating that patients with smaller changes in subpopulations of cells during treatment actually have worse long-term outcome (Fig. 6B, $P = 0.035$). There was no correlation to any other clinicopathological parameters,

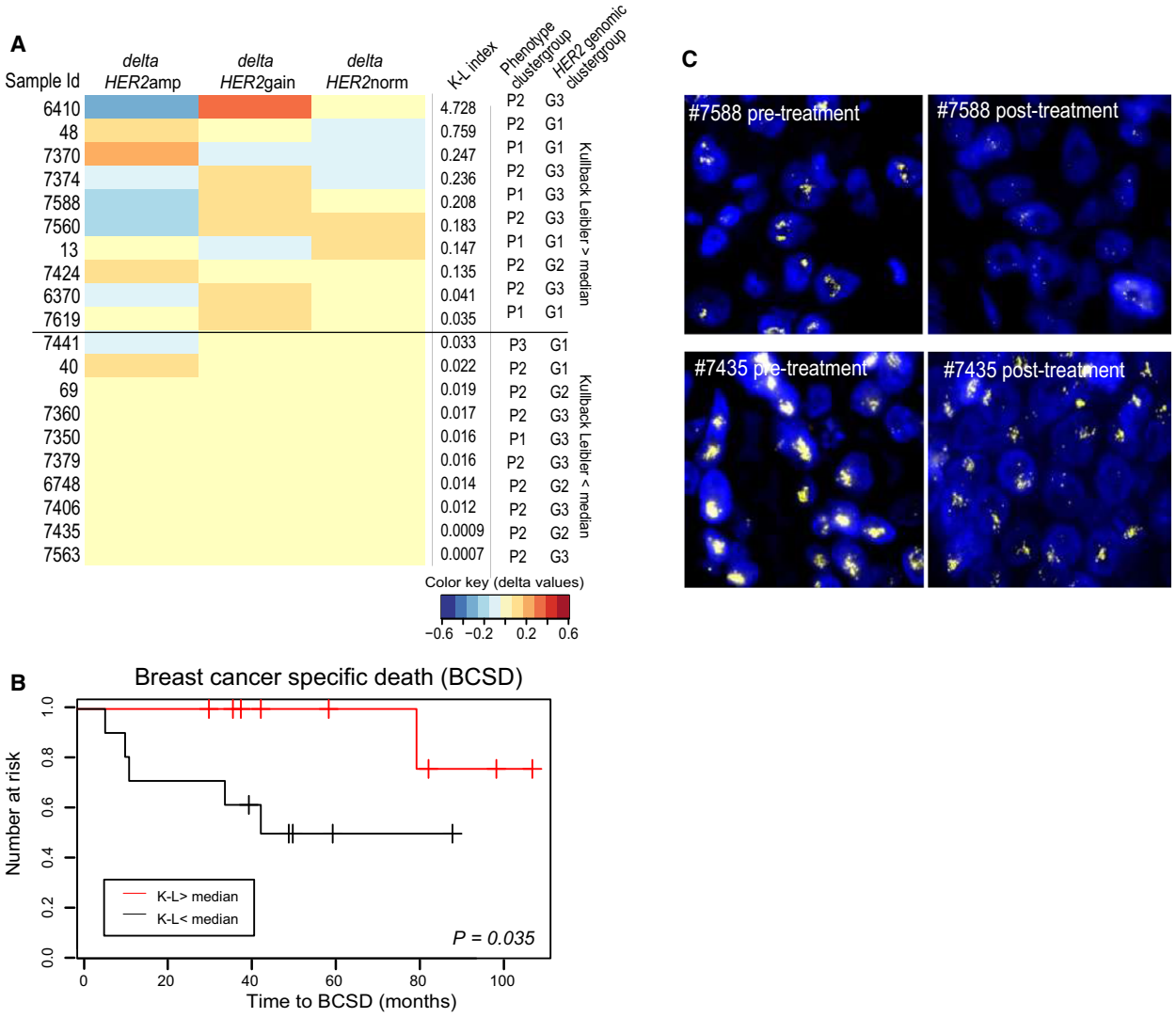


Fig. 6. Tumor evolution during neoadjuvant treatment. (A) The Kullback–Leibler diversity index (K-L index) was calculated reflecting changes in cells with different levels of *HER2* CN during therapy. The samples were sorted from high to low K-L index, and the changes of the different cell typed from pre- to post-treatment are visualized by the delta values. To the right are the K-L index value and the genotypic and phenotypic cluster group for each patient. (B) A significant increase in risk for death of breast cancer was seen for patients with low versus high K-L index ($P = 0.035$, log-rank test). (C) Example images from pre- and post-treatment biopsies from one patient with high K-L index (patient #7588) and from a patient with low K-L index (patient #7435).

including degree of pathological response (Table S6). Figure 6C shows IFISH images (*HER2* CN) from samples taken before and after therapy for two patients. Patient #7588 who did not have progression of the disease showed a decrease in the fractions of cells with *HER2*amp, while patient #7435 who developed progression of the disease did not show any changes in the *HER2* CN cell types during therapy. In contrast, there was neither any association between patient outcomes with phenotypic changes nor with combined phenotypic/*HER2* changes based on the K-L index (Fig. S7A,B).

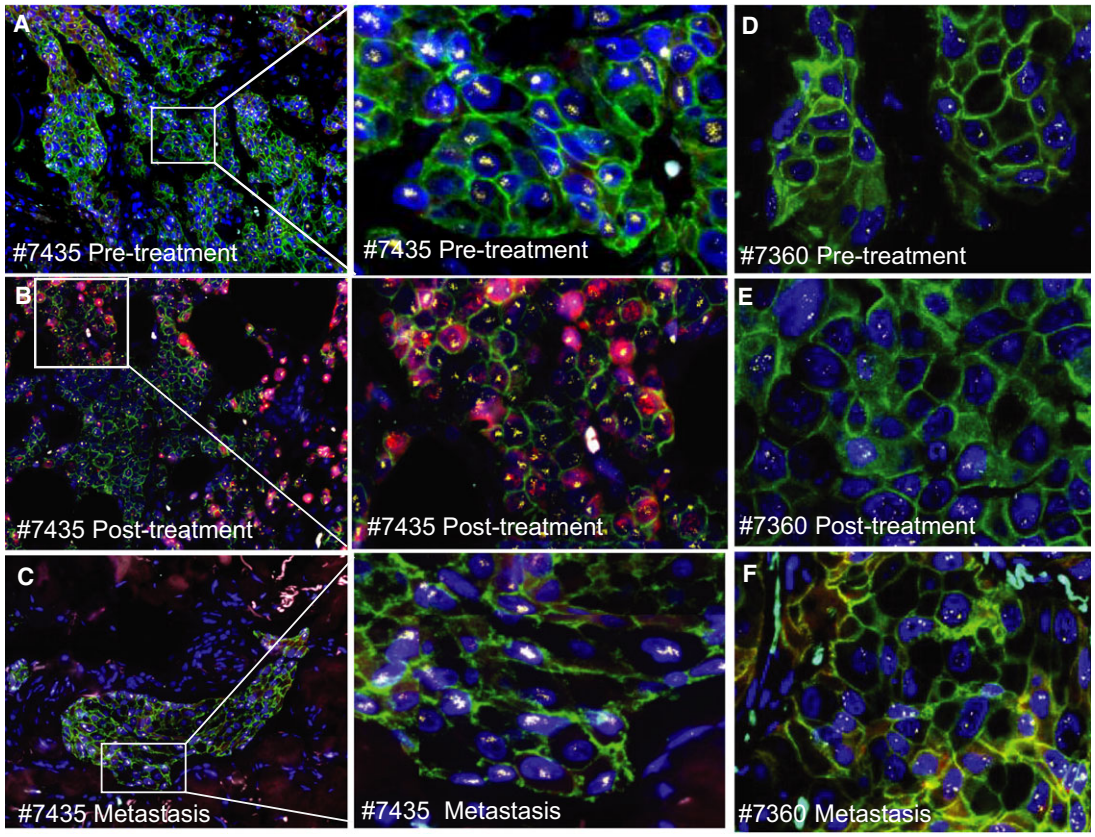
With regard to the individual markers analyzed, we did not observe any significant changes in the global levels of *HER2* and cent17 CN level nor in the HER2 and ER protein intensity in tumors after neoadjuvant treatment (Fig. S7C). In particular, we did not observe a significant difference between patients with a high shift of phenotype or combined phenotypic/*HER2* CN status compared to those with a low shift with regard to survival of disease or treatment outcome.

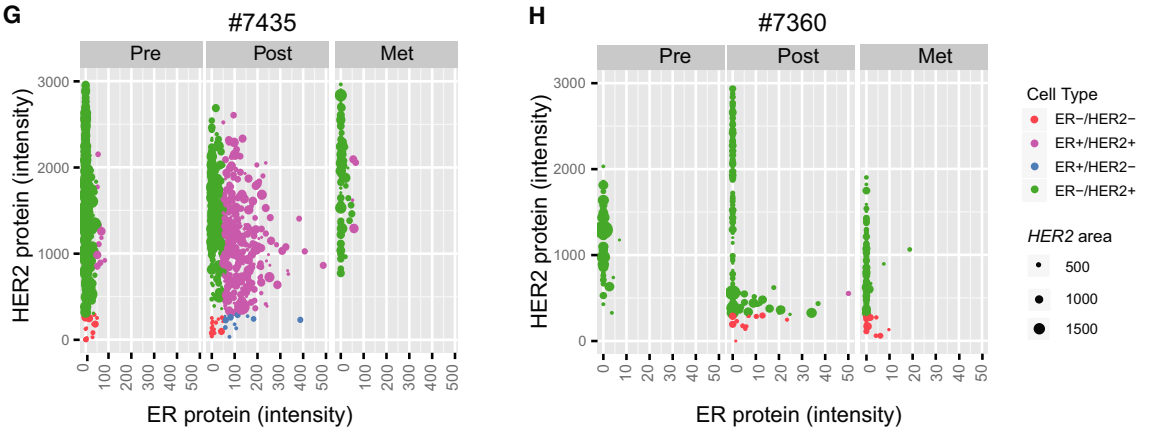
3.4. Diversity in primary tumor versus metastasis

Sampling of tumor metastases was not included in the study protocol, but tissue biopsies from distant metastases were available from three of the patients (two patients with noncomplete response and one patient with complete response to therapy). IFISH images of biopsies from three time points (pre- and post-treatment and later distant metastasis) of two of the patients are shown in Fig. 7A–F. Patient #7435 (Fig. 7A–C) had a primary tumor dominated by HER2+/ER− cells with *HER2* CN amplification. After neoadjuvant treatment, we found an increase in cells with HER2+/ER+ phenotype. Interestingly, the biopsy from a metastasis showed the same cell phenotypes as the pretreatment tumor. There was no evidence of clonal shift as the samples from all three time points were dominated by cells with *HER2* CN amplification (Fig. 7G). In contrast, the tumor from patient #7360 (Fig. 7D–F) had prior to treatment mainly HER2+/ER− cells, but the biopsy after treatment and from the metastasis revealed a small fraction of HER2−/ER− cells. There was only a minimal change in the fraction of cells with *HER2* CN amplification (Fig. 7H). We also investigated the spatial organization of the *HER2* CN at the three time points, and both samples had a more similar spatial pattern for the *HER2* CN for the pretreatment and metastatic lesion in contrast to the post-treatment biopsy, but the changes were only minor (Fig. 7I,J).

4. Discussion

Analysis of tumor samples taken from patients during neoadjuvant treatment is extremely useful for studying the clinical impact of tumor cell diversity. The significance of intratumor heterogeneity for treatment response can be measured by comparing molecular features of tumor cells from pre- and post-treatment biopsies. As *in situ* methods only allow us to measure a small number of markers, we chose the clinically most important biomarkers, namely ER (protein) and HER2 (protein and gene copy number). Even with so few biomarkers, the combined IFISH technique revealed a high diversity both between tumors but also within tumors (i.e., cell-to-cell variation). It is known that tumors classified as HER2+ by immunohistochemistry (i.e., 3+) can have different levels of *HER2* amplification by ISH techniques. Our work supports this observation but also provides a higher resolution as all markers are studied simultaneously in thousands of individual cells. We found remarkable diversity, with regard to the expression of both ER and HER2 proteins as well as for *HER2* CN on a single-cell level (Figs S1, S3, and S4). It was intriguing to find some tumors with a linear correlation between the two proteins and/or between protein and *HER2* CN, while others were not linear. Next, we aimed at investigating whether the type of diversity was truly individual or whether there were patterns of intratumor heterogeneity shared by subsets of tumors. Performing comparative studies of intratumor heterogeneity in sample collections is challenging. Major hurdles are the continuous level of expression/copy number changes per cell, each cell having different combinations of expression/copy number changes, and each tumor having different numbers of cells measured. We therefore chose to categorize the data (by using a threshold for each marker and then assigning each cell into distinct subpopulations) and were thus able to compare the cellular composition across the tumors. Interestingly, when we 'simplified' the complex cellular information this way, we actually found that there are subsets of tumors with similar cell-type composition. Both by classifying each cell into phenotypic and genomic predefined categories and by performing three separate clustering analyses, we found the groups not only having differences in clinical outcome but also in several other interesting features. For instance tumors of patients with a higher risk for disease progression and/or breast cancer-related death have: (a) high expression of HER2 but low or intermediate number of ER+ cells (P2 in Fig. 3), (b) a mixture of cells with different *HER2* CN levels (G3 in


A
#7435 Pre-treatment
#7435 Pre-treatment
D
#7360 Pre-treatment
B
#7435 Post-treatment
#7435 Post-treatment
E
#7360 Post-treatment
C
#7435 Metastasis
#7435 Metastasis
F
#7360 Metastasis


G
#7435
Pre
Post
Met
HER2 protein (intensity)
ER protein (intensity)
H
#7360
Cell Type
ER−/HER2−
ER+/HER2+
ER+/HER2−
ER−/HER2+
HER2 area
500
1000
1500

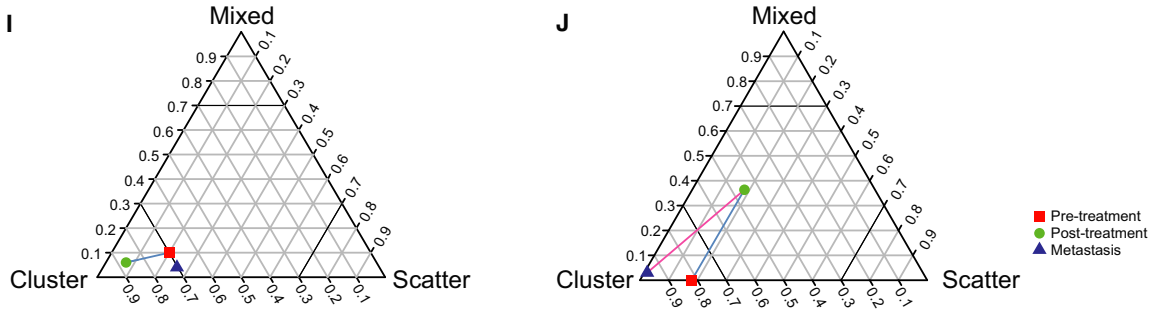

I
Mixed
Cluster
Scatter
J
Pre-treatment
Post-treatment
Metastasis

Fig. 7. Intratumor heterogeneity during disease progression. IFISH images from biopsies from patient #7435 (with a magnified area to the right): (A) pretreatment biopsy, (B) post-treatment biopsy, and (C) biopsy from a metastasis. Equally from patient #7360: (D) pretreatment biopsy, (E) post-treatment biopsy, and (F) biopsy from metastasis (Dapi = blue, HER2 = green, ER = red, *HER2* = yellow, and cent17 = cyan). The phenotype and *HER2* CN level for all tumor cells analyzed from each of the three biopsies are plotted in the diagram (G) patient #7435 and (H) patient #7360 (colored due to their phenotypic cell type and the size of the spot reflect the *HER2* copy number level). Spatial organization of the *HER2* gene visualized in a triangle for the pre-treatment (red square), post-treatment (green circle), and metastatic (blue triangle) sample from patient #7435 (I) and patient #7360 (J).

Fig. 4), and (c) a mixture of cells with different *HER2* CN levels with low number of ER+ cells (PG3 in Fig. S5). Combined, these findings indicate that patients with tumors dominated by *HER2*-amplified cells and with homogenous ER expression (either negative or positive) have a good long-term prognosis. It also indicates the importance of addressing not only the heterogeneity of *HER2* CN but also the variation in ER expression in HER2+ breast carcinomas. In the work by Ferrari *et al.* (2016), HER2+ tumors were split into four groups based on gene expression patterns, and the level of ER expression varied between them. Although the study did not address intratumor heterogeneity, it clearly showed that a subgroup of HER2+ carcinomas was composed of ER-negative tumors, one subgroup of highly ER-positive tumors, and two subgroups of tumors with more intermediate ER levels. It will be of interest to see the follow-up studies of this cohort with outcome data as well. In a recent study, approximately 30% of patients with neoadjuvant-treated HER2+ tumors (chemotherapy and HER2-targeted treatment) achieved pathological complete response (pCR), but this fraction was lower for patients with HER2+ and ER+ tumors, but the level of ER positivity was not addressed (Cortazar *et al.*, 2014). In a study by Romond *et al.* (2005), patients with ER+ tumors had a lower response rate to treatment, but this seems to be mainly restricted to those with tumors having < 50% ER-positive tumor cells. These findings are in line with ours; patients with heterogeneous ER expression had a tendency toward a reduced long-term survival (Fig. 3). Carey *et al.* recently published results from the CALGB40601 trial, which also shows that local response varies between ER+ and ER− subtypes of HER2+ breast cancer (Carey *et al.*, 2016). We found no evidence that the HER2 protein intensity level has impact on local response, which is in line with the observation by Zabaglo *et al.* (2013) but contradicts the CALGB 40601 trial which found gene expression levels of both ER and HER2 to be correlated with pCR rates (Carey *et al.,* 2016).

In our study, we find *HER2* CN level to be of clinical importance as the level in pretreatment samples was significantly higher in tumors from responders compared to nonresponders. This is in line with previous studies showing high levels of *HER2* amplification to be associated with pathological complete response (pCR) (Arnould *et al.*, 2007) (Guiu *et al.*, 2010) although *HER2* CN level could not predict long-term disease progression or survival. This is supported by studies of anti-HER2 treatment in adjuvant setting where *HER2* CN level has shown no or negative correlation with disease-free survival (Xu *et al.*, 2016). As mentioned previously, *HER2* CN heterogeneity seems to have an impact on prognosis in our study. We found tumors with heterogeneous composition with regard to *HER2* CN level to have higher risk of relapse and breast cancer-specific death (patients in G3 group in Fig. 4). Some studies indicate the same result in less advanced stage of the disease; in a study of adjuvant-treated HER2+ breast cancer, Seol *et al.* (2012) found regional heterogeneity in *HER2* CN to predict a worse survival. The study by Lee *et al.* (2014) also found patients with both regional and genomic heterogeneity of *HER2* amplification to have decreased disease-free survival, but neither of these two study cohorts had uniform treatment regimens (Seol *et al.*, 2012). Kurozumi *et al.* studied variation in both *HER2* copy number and HER2 protein expression within tumors using a semi-objective analysis (with visual scoring) and found that regional variation of *HER2* CN reflected a worse prognosis particularly in ER-negative disease (Kurozumi *et al.*, 2016). Unfortunately, these patients had not received anti-HER2 therapy, so neither the predictive value nor the impact of dynamics during therapy could be addressed.

One of the most striking findings in our study was the large number of tumors exhibiting intratumor variation with regard to *HER2* CN levels. As copy number alterations are inherited in daughter cells, we believe these populations to reflect true subclones that have undergone different paths of evolution. The cluster analysis based on *HER2* CN levels showed that patients with tumors dominated by cells with amplified *HER2* gene had a significantly better survival compared to the patients with more heterogeneous *HER2* amplification levels (Fig. 4). Patients in the latter

group (cluster G3 in Fig. 4) had tumors with a mixed cellular composition. These patients had a significantly shorter time to progression of the disease and fewer long-term survivors. We suggest that patients belonging to cluster group G3 represent cases similar to those described by Ballard *et al.* (2017) as 'nonclassical' *HER2* FISH results.

Changes in ER and HER2 status are observed for some cases during neoadjuvant treatment, and this change seems to affect protein expression (i.e., phenotype) more than *HER2* copy numbers (Van de Ven *et al.*, 2011). However, studies of genomic and phenotypic intratumor heterogeneity of HER2+ breast carcinomas and their impact on treatment resistance have been scarce. A recent study of HER2+ tumors at the single-cell level found overexpression of *BRF2* and *DSN1* as genomic driver events in HER2-negative cells (Ng *et al.*, 2015). This indicates a presence of subpopulations that can explain treatment resistance. It has also been shown that that important genetic driver events such as *PIK3CA* mutation and *HER2* gene amplification are not always present within the same cell (Janiszewska *et al.*, 2015). As minor subclones might need time to proliferate and progress (by clonal selection), this could explain why we find heterogeneous tumors to have a significantly increased risk for disease progression regardless of the initial local response. When comparing intratumor heterogeneity before and after treatment, we were surprised to find that patients in the group with no changes in the cellular composition had an increased risk for later progression of the disease. One explanation for this finding could be that none of the tumor subclones were affected by the treatment and probably reflecting tumors where *HER2* gene amplification is not the important driver. Another explanation could be treatment resistance due to ligand-independent activation of HER2 (Yarden, 2001) rather than selection of clones proliferating independently of HER2 activity. Interestingly, these tumors do not reflect the situation identified by Ng *et al.* (2015) where a HER2-negative subpopulation could be suspected to explain therapy resistance. Our study was unfortunately not suitable for next-generation sequencing (NGS)-based identification of driver events in resistant subclones and more detailed explorative studies to identify alternative candidate drivers will be needed. Identification of distinct genomic alterations related to the cellular dynamics during treatment might provide clinicians with more therapy options for such patients.

Finally, the cases with samples from three time points showed intriguing results; the pretreatment and metastatic lesion had a more similar spatial pattern for the *HER2* CN in contrast to the post-treatment biopsy (Fig. 7I,J). One of the cases showed a major switch in phenotype (Fig. 7A–C) but had a very low Kullback–Leibler index, reflecting minor influence of treatment on *HER2* CN cell types. The other case had only a minor phenotype change, and the *HER2* CN cell types did not shift enough to be reflected by the Kullback–Leibler index. Although this is just case observation, it reflects breast cancer to be a disease that can evolve along different paths with regard to both phenotypic and genomic/clonal composition.

An important challenge for estimating intratumor heterogeneity is the need for objective measurements of molecular biomarkers. Buckley *et al.* (2016) proposed a simple heterogeneity index for *HER2* CN heterogeneity, but this was based on visual counting of 20 cells (as defined by the CAP guidelines) by an observer. To address potential observer bias and maximize the number of analyzed cells, we estimated heterogeneity by objective assessment of *HER2* CN of more than 13 000 cells using GOIFISH, an image analysis software that can omit artificial staining and specifically characterize tumor cells for further analysis. Still, tissue artifacts such as incomplete tumor cell nuclei due to sectioning can influence the results. We also used cluster analyses of the *fractions* of cell types within a tumor; thus, the presence of some misclassified cells will not influence the results substantially. Finally, the visual categorization of intranuclear spatial distributions of the *HER2 amplicon* also reflected the presence of different types of genomic disruptions and amplification mechanisms, representing a different way of assessing clonal heterogeneity. Here, we analyzed fewer cells per sample (100 cells), but the finding is in line with other studies (by DNA sequencing) showing that *HER2* gene amplifications can be a result of different types of DNA rearrangement mechanisms (Morganella *et al.*, 2016). This cohort does not have tumor material suitable for NGS analyses of this kind, but this is important to address in suitable sample collections.

This study is based on a neoadjuvant observational trial, comprising of HER2+ patients for which matched primary, post-treatment, and, in some cases, metastatic samples were available for analysis. The strength of this cohort lies in the strict inclusion criteria and consistency in terms of treatment regimens, allowing us to make direct comparisons between patient samples and track the cellular dynamics throughout the treatment process. Although this study could benefit from an increased patient sample size and sufficient patient material to conduct DNA sequencing analysis, this observational cohort has nonetheless offered an insight on the wide biological

spectrum within HER2+ breast carcinomas and in particular the negative association between *HER2* CN intratumoral heterogeneity and patient outcome.

5. Conclusion

This is to our knowledge the first study of breast cancer revealing cellular heterogeneity with regard to HER2 expression, *HER2* copy number, and ER expression through analysis of a substantial number of cells from neoadjuvant-treated HER2+ breast cancer patients. HER2+ disease is highly heterogeneous both between and within tumors. The heterogeneity of ER expression as well as *HER2* copy number variation seems to have impact on disease progression and survival. Additionally, tumors with preserved level of *HER2* CN heterogeneity during therapy (i.e., cell-type composition before and after therapy) had a poor prognosis. The study shows the importance of assessing cell-to-cell variation both prior to treatment and during treatment, and consequential population shifts to predict response to therapy. It also shows the importance of having an objective analysis of multiple markers in a high number of cells facilitated by automatized image analysis. The challenge now is not only to validate the clinical impact of molecular subtypes within HER2+ breast cancer patients but also to address the cellular variation within the tumors in more depth.

Acknowledgements

We thank the hospitals in Buskerud, Telemark, Vestfold, Østfold, Bodø, Innlandet, AHUS, Ullevål, and the Norwegian Radiumhospital for retrieving and sending us the tissue blocks from the patients included in this study. KP would like to acknowledge the NIH grants CA195469 and CA197623, and FM and AT would like to acknowledge the support of The University of Cambridge, Cancer Research UK, and C K Hutchison Holdings Limited. Parts of this work were funded by CRUK core grant C14303/A17197 and A19274. IHR and HGR are grateful for funding from Health Region South-East, Norway, and from The Norwegian Cancer Society. IHR, HGR, and A-LB-D are also grateful for support for accommodation in Cambridge, UK, from Radiumhospitalets legater.

Author contributions

HGR, ÅH, A-LB-D, and IHR conceived the study. ABS and ÅH contributed to trial design and clinical data. IHR, VA, KP, and HGR conducted and/or designed all wet-laboratory experiments and tissue analyses. IHR, AT, FM, DN, and OCL performed and/or developed image analyses, bioinformatic analyses, and statistical analyses. HGR, FM, IHR, AT, ÅH, ABS, A-LB-D, KP, VA, DN, and OCL participated in the combined analyses and wrote the manuscript.

References

Arena V, Pennacchia I, Vecchio FM and Carbone A (2013) HER-2 intratumoral heterogeneity. *Mod Pathol* **26**, 607–609.

Arnould L, Arveux P, Couturier J, Gelly-Marty M, Loustalot C, Ettore F, Sagan C, Antoine M, Penault-Llorca F, Vasseur B *et al.* (2007) Pathologic complete response to trastuzumab-based neoadjuvant therapy is related to the level of HER-2 amplification. *Clin Cancer Res* **13**, 6404–6409.

Ballard M, Jalikis F, Krings G, Schmidt RA, Chen YY, Rendi MH, Dintzis SM, Jensen KC, West RB, Sibley RK *et al.* (2017) 'Non-classical' HER2 FISH results in breast cancer: a multi-institutional study. *Mod Pathol* **30**, 227–235.

Bartlett AI, Starcyznski J, Robson T, Maclellan A, Campbell FM, van de Velde CJ, Hasenburg A, Markopoulos C, Seynaeve C, Rea D *et al.* (2011) Heterogeneous HER2 gene amplification: impact on patient outcome and a clinically relevant definition. *Am J Clin Pathol* **136**, 266–274.

Baselga J, Bradbury I, Eidtmann H, Di Cosimo S, de Azambuja E, Aura C, Gómez H, Dinh P, Fauria K, Van Dooren V *et al.* (2012) Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTTO): a randomised, open-label, multicentre, phase 3 trial. *Lancet* **379**, 633–640.

Buckley NE, Forde C, McArt DG, Boyle DP, Mullan PB, James JA, Maxwell P, McQuaid S and Salto-Tellez M (2016) Quantification of HER2 heterogeneity in breast cancer – implications for identification of sub-dominant clones for personalised treatment. *Sci Rep* **6**, 23383.

Carey LA, Berry DA, Cirrincione CT, Barry WT, Pitcher BN, Harris LN, Ollila DW, Krop IE, Henry NL, Weckstein DJ *et al.* (2016) Molecular heterogeneity and response to neoadjuvant human epidermal growth factor receptor 2 targeting in CALGB 40601, a randomized phase III trial of paclitaxel plus trastuzumab with or without lapatinib. *J Clin Oncol* **34**, 542–549.

Cortazar P, Zhang L, Untch M, Mehta K, Costantino JP, Wolmark N, Bonnefoi H, Cameron D, Gianni L, Valagussa P *et al.* (2014) Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet* **384**, 164–172.

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y *et al.* (2012) The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352.

Ferrari A, Vincent-Salomon A, Pivot X, Sertier AS, Thomas E, Tonon L, Boyault S, Mulugeta E, Treilleux I, MacGrogan G *et al.* (2016) A whole-genome sequence and transcriptome perspective on HER2-positive breast cancers. *Nat Commun* **7**, 12222.

Gianni L, Eiermann W, Semiglazov V, Manikhas A, Lluch A, Tjulandin S, Zambetti M, Vazquez F, Byakhow M, Lichinitser M *et al.* (2010) Neoadjuvant chemotherapy with trastuzumab followed by adjuvant trastuzumab versus neoadjuvant chemotherapy alone, in patients with HER2-positive locally advanced breast cancer (the NOAH Trial): a randomised controlled superiority trial with a parallel HER. *Lancet* **375**, 377–384.

Guarneri V and Conte PF (2004) The curability of breast cancer and the treatment of advanced disease. *Eur J Nucl Med Mol Imaging* **31**, S149–S161.

Guiu S, Gauthier M, Coudert B, Bonnetain F, Favier L, Ladoire S, Tixier H, Guiu B, Penault-Llorca F, Ettore F *et al.* (2010) Pathological complete response and survival according to the level of HER-2 amplification after trastuzumab-based neoadjuvant therapy for breast cancer. *Br J Cancer* **103**, 1335–1342.

Gulbahce HE, Factor R, Geiersbach K and Downs-Kelly E (2016) HER2 immunohistochemistry-guided targeted fluorescence in situ hybridizationanalysis does not help identify intratumoral heterogeneity in breast cancer. *Arch Pathol Lab Med* **140**, 741.

Helsedirektoratet (2014) *Nasjonalt Handlingsprogram Med Retningslinjer for Diagnostikk, Behandling Og Oppfølging Av Pasienter Med Brystkreft*. Helsedirektoratet.

Janiszewska M, Liu L, Almendro V, Kuang Y, Paweletz C, Sakr RA, Weigelt B, Hanker AB, Chandarlapaty S, King TA *et al.* (2015) In situ single-cell analysis identifies heterogeneity for PIK3CA mutation and HER2 amplification in HER2-positive breast cancer. *Nat Genet* **47**, 1212–1219.

Kullback S and Leibler RA (1951) On information and sufficiency. *Ann Math Stat* **22**, 79–86.

Kurozumi S, Padilla M, Kurosumi M, Matsumoto H, Inoue K, Horiguchi J, Takeyoshi I, Oyama T, Ranger-Moore J, Allred DC *et al.* (2016) HER2 intratumoral heterogeneity analyses by concurrent HER2 gene and protein assessment for the prognosis of HER2 negative invasive breast cancer patients. *Breast Cancer Res Treat* **158**, 99–111.

Lee HJ, Seo AN, Kim EJ, Jang MH, Suh KJ, Ryu HS, Kim YJ, Kim JH, Im SA, Gong G *et al.* (2014) HER2 heterogeneity affects trastuzumab responses and survival in patients with her2-positive metastatic breast cancer. *Am J Clin Pathol* **142**, 755–766.

Lewis JT, Ketterling RP, Halling KC, Reynolds C, Jenkins RB and Visscher DW (2005) Analysis of intratumoral heterogeneity and amplification status in breast carcinomas with equivocal (2+) HER-2 immunostaining. *Am J Clin Pathol* **124**, 273–281.

Morganella S, Alexandrov LB, Glodzik D, Zou X, Davies H, Staaf J, Sieuwerts AM, Brinkman AB, Martin S, Ramakrishna M *et al.* (2016) The topography of mutational processes in breast cancer genomes. *Nat Commun* **7**, 11383.

Ng CK, Martelotto LG, Gauthier A, Wen HC, Piscuoglio S, Lim RS, Cowell CF, Wilkerson PM, Wai P, Rodrigues DN *et al.* (2015) Intra-tumor genetic heterogeneity and alternative driver genetic alterations in breast cancers with heterogeneous HER2 gene amplification. *Genome Biol* **16**, 1–21.

Nishino M, Jagannathan JP, Ramaiya NH and Van Den Abbeele AD (2010) Revised RECIST guideline version 1.1: what oncologists want to know and what radiologists need to know. *Am J Roentgenol* **195**, 281–289.

Parker JS, Mullins M, Cheang MC, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z *et al.* (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* **27**, 1160–1167.

Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA *et al.* (2000) Molecular portraits of human breast tumours. *Nature* **406**, 747–752.

R Core Team (2017) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. Retrieved from http://www.R-project.org/

Romond EH, Perez EA, Bryant J, Suman VJ, Geyer CE Jr, Davidson NE, Tan-Chiu E, Martino S, Paik S, Kaufman PA *et al.* (2005) Trastuzumab plus adjuvant chemotherapy for operable HER2-positive breast cancer. *N Engl J Med* **353**, 1673–1684.

Seol H, Lee HJ, Choi Y, Lee HE, Kim YJ, Kim JH, Kang E, Kim SW and Park SY (2012) Intratumoral heterogeneity of HER2 gene amplification in breast cancer: its clinicopathological significance. *Mod Pathol* **25**, 938–948.

Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* **27**, 379–423.

Trinh A, Rye IH, Almendro V, Helland A, Russnes HG and Markowetz F (2014) GoIFISH: a system for the quantification of single cell heterogeneity from IFISH images. *Genome Biol* **15**, 442.

Van de Ven S, Smit VTHBM, Dekker TJA, Nortier JWR and Kroep JR (2011) Discordances in ER, PR and HER2 receptors after neoadjuvant chemotherapy in breast cancer. *Cancer Treat Rev* **37**, 422–430.

Viani GA, Afonso SL, Stefano EJ, De Fendi LI and Soares FV (2007) Adjuvant trastuzumab in the treatment of Her-2-positive early breast cancer: a meta-analysis of published randomized trials. *BMC Cancer* **7**, 153.

Wolff AC, Hammond ME, Hicks DG, Dowsett M, McShane LM, Allison KH, Allred DC, Bartlett JM, Bilous M, Fitzgibbons P *et al.* (2013) Recommendations for human epidermal growth factor receptor 2 testing in breast cancer: American Society of Clinical Oncology/College of American Pathologists clinical practice guideline update. *J Clin Oncol* **31**, 3997–4013.

Xu QQ, Pan B, Wang CJ, Zhou YD, Mao F, Lin Y, Guan JH, Shen SJ, Zhang XH, Xu YL *et al.* (2016) HER2 amplification level is not a prognostic factor for HER2- positive breast cancer with trastuzumab-based adjuvant treatment: a systematic review and meta-analysis table 1: correlations between HER2 levels and response to trastuzumab or survival. *Oncotarget* **7**, 63571–63582.

Yarden Y (2001) The EGFR family and its ligands in human cancer. Signalling mechanisms and therapeutic opportunities. *Eur J Cancer* **37**(Suppl 4), S3–S8.

Zabaglo L, Stoss O, Rüschoff J, Zielinski D, Salter J, Arfi M, Bradbury I, Dafni U, Piccart-Gebhart M, Procter M *et al.* (2013) HER2 staining intensity in HER2-positive disease: relationship with FISH amplification and clinical outcome in the HERA trial of adjuvant trastuzumab. *Ann Oncol* **24**, 2761–2766.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Images from all pre-treatment biopsies.

Fig. S2. (A) Analyses of the pre-treatment samples (n = 37) stratified by treatment response (pCR, n = 12 and non-pCR, n = 25), (B) Risk of relapse and breast cancer related death with regard to the level of ER+ cells in pre-treatment biopsies.

Fig. S3. Distribution of ER and HER2 expression in each cell from all tumors separately (at two time-points for non-pCR samples).

Fig. S4. Distribution of *HER2* copy number and HER2 expression in each cell from all tumors separately (at two time-points for non-pCR samples).

Fig. S5. (A) Cluster analyses of the 12 combined phenotype/*HER2* CN cell types on pre-treatment samples (B) Risk of breast cancer specific death in the three cluster groups and C) Risk for disease progression for patients in the three cluster groups.

Fig. S6. (A) *HER2* spatial organization in metastatic and non-metastatic samples and (B) *HER2* spatial organization in ER positive and ER negative samples.

Fig. S7. Survival analyses of patients with regard to (A) change in phenotypic heterogeneity and (B) phenotypic and genomic heterogeneity. (C) Comparison of marker assessment before and after therapy.

Table S1. Clinical and pathology data for all patients.

Table S2. Overview of biopsies, images and number of cells included in the study.

Table S3. Clinico-pathological demographics for the different cluster groups.

Table S4. Shannon index calculated for each case.

Table S5. Relationship between the different HER2 spatial patterns and (A) response to therapy, (B) metastasis, (C) ER status and (D) ER percentage.

Table S6. Clinico-pathological demographics for Kullback Leibler groups.