Spatial proteomic characterization of HER2-positive breast tumors through neoadjuvant therapy predicts response

Katherine L. McNamara[1,2,3], Jennifer L. Caswell-Jin[1], Rohan Joshi[4], Zhicheng Ma[2], Eran Kotler[3], Gregory R. Bean[4], Michelle Kriner[5], Zoey Zhou[5], Margaret Hoang[5], Joseph Beechem[5], Dennis J. Slamon[6], Sara A. Hurvitz[6], Christina Curtis[1,2,3 *]

Affiliations
1 Department of Medicine, Division of Oncology, Stanford University School of Medicine, Stanford, California, USA
2 Stanford Cancer Institute, Stanford University School of Medicine, Stanford, California, USA
3 Department of Genetics, Stanford University School of Medicine, Stanford, California, USA
4 Department of Pathology, Stanford University School of Medicine, Stanford, California, USA
5 NanoString Technologies, Seattle, Washington USA
6 David Geffen School of Medicine, University of California Los Angeles, Los Angeles, California, USA

* Corresponding Author:
Christina Curtis, PhD
Stanford University School of Medicine
265 Campus Drive, Lorry Lokey Building Suite G2120C
Stanford, CA 94305
(650) 498-9943
Email: cncurtis@stanford.edu

Abstract

Addition of HER2-targeted agents to neoadjuvant chemotherapy has dramatically improved pathological complete response (pCR) rates in early stage Human epidermal growth factor receptor 2 (HER2)-positive breast cancer. Still, up to 50% of patients have residual disease following treatment and biomarkers predictive of response are urgently needed. We performed spatial proteomic characterization using NanoString GeoMX Digital Spatial Profiling (DSP) of 122 samples from 57 HER2-positive breast tumors from the neoadjuvant TRIO-US B07 clinical trial (discovery cohort, n=28; validation cohort, n=29) sampled pre-treatment, after 14-21 days of HER2-targeted therapy and at surgery. *In situ* quantification of 40 tumor and immune proteins across multiple pancytokeratin-enriched regions per sample revealed that treatment results in decreased HER2 signaling and increased immune infiltration after several weeks of therapy. These changes were more dramatic in tumors that ultimately undergo pCR, and a classifier trained to predict pCR using on-treatment and pre-treatment DSP protein levels had a cross-validation mean Area Under the Receiver Operating Characteristics (AUROC) of 0.733 in the discovery cohort and validated in an independent cohort (AUROC = 0.725). Thus, we demonstrate robust stratification of sensitive tumors early during neoadjuvant HER2-targeted therapy using a multiplex spatial proteomic biomarker with implications for tailoring subsequent therapy.

Human epidermal growth factor receptor 2 (HER2)-positive breast cancer accounts for 15–30% of invasive breast cancers and is associated with an aggressive phenotype [1]. While the addition of HER2-targeted agents to neoadjuvant chemotherapy has dramatically improved pathological complete response (pCR) rates in early stage HER2-positive breast cancer, 40-50% of patients have residual disease after treatment [2,3]. Conversely, HER2 inhibition with two targeted agents and without chemotherapy can result in pCR, suggesting that it may be possible to eliminate chemotherapy in a subset of patients [4-7]. Given the heterogeneity in response to HER2-targeted therapy [2,8], identification of biomarkers of response beyond HER2 and estrogen receptor (ER) status is needed.

Bulk gene expression profiling of pre-treatment samples has identified tumor characteristics (HER2-enriched intrinsic subtype, HER2 expression levels, and ESR1 expression levels [7,9-12]), and microenvironmental characteristics (increased immune infiltration [10,12-15]) that associate with response to HER2-targeted therapy in the neoadjuvant setting. Because tumor cells are profiled simultaneously with both co-localized and distant stroma and immune cells, bulk expression profiling is an imperfect tool for analyzing tumor and microenvironmental changes across treatment. In particular, it is difficult to assign observed changes to specific geographic or phenotypic cell populations within the complex tumor ecosystem, where malignant tumor cells interact with fibroblasts, endothelial cells, and immune cells. Moreover, immune cells can be further divided into those that infiltrate the tumor core and those that are excluded [16]. As of yet, how the tumor and immune microenvironment change during therapy remains poorly understood, necessitating multiplexed *in situ* profiling of longitudinal tissue samples.

We used the GeoMx™ Digital Spatial Profiling (DSP, NanoString) technology to assay archival tissue from an initial discovery set of 28 patients with HER2-positive breast cancer enrolled on the neoadjuvant TRIO-US B07 clinical trial [17], whose tumors were sampled pre-treatment, after 14-21 days of HER2-targeted therapy, consisting of lapatinib, trastuzumab, or both (on-treatment), and at the time of surgery after completion of combination chemotherapy with HER2-targeted therapy (post-treatment). We subsequently validated our results in an independent validation set of 29 patients from the B07 clinical trial. Importantly, the neoadjuvant setting allows for early assessment of treatment response and pCR is a strong surrogate for long-term survival in HER2-positive disease [18-20]. DSP enables geographic and phenotypic selection of tissue regions for multiplex proteomic characterization of cancer signaling pathways and the tumor-colocalized immune microenvironment [21,22]. In particular, we characterized spatial heterogeneity in untreated breast tumors as well as changes in cancer signaling pathways and microenvironmental composition in matched on-treatment biopsies and post-treatment surgical samples by profiling 40 tumor and immune proteins across multiple pancytokeratin (panCK)-enriched regions per sample. On-treatment protein expression changed dramatically in tumors that went on to achieve a pCR and a classifier based on these data robustly predicted treatment response in the validation cohort. This new spatial-proteomic biomarker outperformed established predictors such as PAM50 subtype [7,11,23] as well as classifiers based on transcriptomic data in this cohort, suggesting new avenues to personalize therapy in early-stage HER2-positive breast cancer.

Results

Spatial proteomic analysis of untreated HER2-positive breast tumors

Participants in the TRIO-US B07 clinical trial (NCT00769470 in early-stage HER2-positive breast cancer) received one cycle of neoadjuvant HER2-targeted therapy, including either trastuzumab, lapatinib, or both agents, followed by six cycles of the assigned HER2-targeted therapy plus docetaxel and carboplatin given every three weeks [17]. Core biopsies were obtained pre-treatment and

on-treatment after 14-21 days of HER2-targeted therapy, and surgical resection specimens were obtained post-treatment (**Figure 1a**). Initially, we defined a discovery cohort of 28 patients for whom FFPE samples were available from all three timepoints (pre-treatment, on-treatment, and at surgery). The cohort was balanced for both pCR and ER status (**Supplementary Figures 1,2**) and was used for all exploratory analyses. We additionally defined a validation cohort of 29 patients from the TRIO-US B07 cohort with matched pre and on-treatment FFPE samples for evaluation of model performance.

DSP enables multiplex proteomic profiling of formalin-fixed paraffin-embedded (FFPE) tissue sections (**Supplementary Figure 3**), where regions of interest (ROIs) can be selected based on both geographic and phenotypic characteristics. We employed a panCK enrichment strategy to profile cancer cells and colocalized immune cells across an average of four regions per tissue specimen (**Figure 1b**). Using CD45, panCK, and dsDNA immunofluorescent markers for visualization, we selected spatially separated regions (**Supplementary Figure 4**) and a mask governing the UV illumination for protein quantitation was generated based on panCK immunofluorescence. In total, 40 tumor and immune proteins were profiled in the discovery cohort (**Figure 1b**). We further utilized paired pre and on-treatment bulk gene expression data from the same patients to infer PAM50 subtype and enable comparisons with the spatially resolved DSP data [17].

In untreated tumors, the correlation amongst immune markers was striking, suggesting the coordinated action of multiple immune cell subpopulations (**Figure 1c**). HER2 pathway members and other downstream cancer signaling markers were also highly correlated, while the correlation between tumor and immune markers was minimal for most marker pairs. Inter- and intra-tumor variability at the proteomic level was evident prior to treatment, including for HER2 and the pan-leukocyte marker CD45 (**Figure 1d**). Averaging all ROIs per patient to derive a composite score per marker, we found that baseline HER2 levels were similar among tumors that achieved a pCR versus those that did not (mean pCR cases: 14.50, mean non-pCR cases: 15.00), as were CD45 levels (mean pCR: 9.90, mean non-pCR: 9.64). Using a linear mixed-effects model with blocking by patient (Methods), we further found that individual DSP protein markers, including HER2 and CD45, did not significantly differ in pCR versus non-pCR cases prior to treatment (unadjusted $p > 0.10$ for all markers).

Decreased cancer signaling and increased immune infiltration after short-term HER2-targeted therapy

We used DSP to investigate treatment-related changes in both breast tumor and immune markers during short-term HER2-targeted therapy by profiling on-treatment (after a single cycle of HER2-targeted therapy alone) biopsies in the discovery cohort. The protein markers that were most associated with pCR at the on-treatment timepoint were CD45 (unadjusted $p=0.0024$) and CD56, a natural killer (NK) cell marker (unadjusted $p=0.0055$) (**Supplementary Figure 5**). We quantified the fold change in protein levels on-treatment relative to pre-treatment using a linear mixed-effects model with blocking by patient and visualized the significance (false discovery adjusted p-value) of all markers relative to their fold change in volcano plots. These analyses revealed a dramatic reduction in HER2 and Ki67, accompanied by other downstream pathway members, including pAKT, AKT, pERK, S6, and pS6, with the phosphorylated proteins decreasing comparatively more (**Figure 2a**). Immune markers – including CD45 and CD8, a marker of cytotoxic T-cells – exhibited the greatest increase in expression with treatment. Of note, increased expression of CD8+ T-cells was similarly observed in the TRIO-US B07 transcriptomic data through cell-type deconvolution [17], however given the lack of a control arm undergoing repeated biopsy without intervening treatment, it is impossible to be certain whether the immune changes observed were related to HER2-targeted therapy or repeated biopsy. More generally, the on- versus pre-treatment bulk transcriptome data mirrored the

changes seen at the protein level, but the fold changes were attenuated (**Supplementary Figure 6**). For example, using genes that correspond with the DSP protein markers, we found that the expression of HER2, AKT, Ki67, and breast cancer-associated keratin genes (KRT7, KRT18, and KRT19) [24] decreased significantly with treatment, while immune markers increased (**Supplementary Figure 7**). Despite the use of different analytes, measurements, and tissue sections, the DSP protein and bulk RNA datasets consistently showed decreased HER2 signaling and breast cancer-associated markers, accompanied by increased immune cell infiltration during neoadjuvant treatment. Given that lapatinib was associated with lower pCR rates in the TRIO-US B07 trial and more generally [25], we additionally assessed on-treatment changes in the trastuzumab-treated cases (arms 1 and 3, n=23) and observed similar patterns as in the full cohort (**Supplementary Figure 8**).

We next examined how treatment-associated changes differed based on tumor sensitivity to HER2-targeted therapy, stratifying tumors based on achievement of pCR following neoadjuvant therapy (**Figure 2b**). In the pCR cases, a multitude of immune markers increased with treatment, while, in the non-pCR cases, no significant treatment-associated immune changes were observed and the reduction in Ki67 and HER2 signaling was modest. These patterns can also be visualized via pairwise comparisons of protein marker correlations, which revealed a stronger negative correlation between the immune marker cluster and the cancer cell marker cluster in tumors that went on to achieve a pCR (mean fold change across all markers in pCR cases: -0.231, non-pCR cases: -0.075, two-sided Wilcoxon rank sum test $p < 2.2e-16$) (**Figure 2c**).

Since both ER status [5,23] and HER2-enriched subtype have been associated with response to neoadjuvant therapy [7,11,15], we explored how protein marker expression changed with these covariates. ER-negative tumors exhibited more significant changes on-treatment (relative to pre-treatment) compared to ER-positive tumors (mean absolute fold change ER-negative cases: 0.59, mean ER-positive cases: 0.36, two-sided Wilcoxon rank sum test $p=0.0045$, **Figure 2d**). However, when tumors were stratified by outcome, pCR cases exhibited more significant changes than non-pCR cases regardless of ER status (**Supplementary Figure 9**) and ER status was not predictive of pCR in this cohort ($p=0.47$). Similarly, tumors classified as HER2-enriched prior to treatment exhibited significant changes in tumor and immune markers in the on-treatment biopsy relative to other subtypes (**Figure 2e, Supplementary Figure 10**). For example, while CD8+ T-cells increased significantly with treatment in HER2-enriched cases, they decreased slightly in other cases. As in the full TRIO-US B07 transcriptomic cohort [17], HER2-enriched subtype was not predictive of pCR ($p=0.87$).

In order to assess the utility of multi-region sampling, we measured changes on- versus pre-treatment using a single randomly selected region per tissue sample averaged across 100 simulations (**Supplementary Figure 11**). Consistent with the findings based on all tumor regions, CD45 and CD8 showed the greatest increase on-treatment, while HER2 and pS6 decreased most in the single region analysis. While the magnitude of marker fold change with treatment was greater for pCR cases than non-pCR cases (mean absolute fold change across all markers in pCR cases: 0.87 versus non-pCR cases: 0.33, two-sided Wilcoxon rank sum test $p= 1.02e-07$), individual markers did not increase significantly with treatment in the single region analysis, reflecting increased variance.

We also examined treatment-associated changes in patients with residual tumor cells present at the time of surgery (non-pCR cases) to elucidate the biology associated with combined HER2-targeted therapy and chemotherapy. While the non-pCR cases showed limited changes at the on-treatment timepoint, by the time of surgery there was a substantial decrease in the HER2 and downstream AKT signaling pathway, and a concomitant increase in immune markers in panCK-enriched regions (**Supplementary Figure 12**). Notably, HER2 decreased more significantly than its downstream pathway members, which may reflect compensatory pathway activation contributing to resistance [6,26].

While some immune markers increased significantly in non-pCR cases at surgery (n=8), the fold change was diminished relative to pCR cases sampled on treatment (mean fold change non-pCR post-treatment: 0.30, mean fold change pCR on-treatment: 0.85, two-sided Wilcoxon rank sum test p=0.0021, **Figure 2b**). Amongst the immune markers that increased at surgery in the non-pCR cases, CD56 was most significant and potentially related to the role of NK cells in identifying and killing chemotherapy-stressed tumor cells [27]. NK-cells were similarly found to increase at time of surgery in the TRIO-US B07 bulk expression data [17].

Increased heterogeneity of tumor and immune markers during HER2-targeted therapy

Given that tumor heterogeneity is a defining feature of HER2-positive breast cancer, we next evaluated the extent to which HER2 protein expression varied within different regions of a breast tumor biopsy through neoadjuvant treatment and between patients. As shown for two exemplary cases (**Figure 3a**), HER2 protein levels across geographically disparate regions within each tissue sample exhibited relatively consistent HER2 protein levels prior to treatment in the majority of cases (**Figure 3b**). Far greater heterogeneity in HER2 protein expression was observed on treatment both between regions and between patients (**Figure 3c**). Such regional heterogeneity may reflect pharmacokinetic differences due to vasculature, tissue architecture, immune infiltration, or the biopsy itself, underscoring the importance of profiling multiple regions per sample on-treatment.

We also investigated regional heterogeneity across both tumor and immune protein markers during treatment. For each marker and at each timepoint, regional heterogeneity across the cohort was computed as the within-patient mean squared error based on ANOVA (Methods). Across all markers, DSP protein heterogeneity increased significantly on-treatment relative to pre-treatment (**Figure 3d**), similar to that noted for HER2. These changes were widespread, with heterogeneity being higher for all tumor and immune markers on-treatment compared to pre-treatment. The probes with the greatest heterogeneity included both tumor (HER2, pS6) and immune (CD3, CD8) markers. Amongst tumors that failed to achieve a pCR, we evaluated heterogeneity throughout the course of neoadjuvant therapy. Heterogeneity amongst tumor markers was not significantly different on-treatment and pre-treatment (two-sided Wilcoxon rank sum test p=0.52), but increased at surgery (post-treatment), whereas immune marker heterogeneity increased on treatment with a subsequent decrease at surgery (**Supplementary Figure 13**). Tumors that achieved a pCR exhibited higher protein heterogeneity amongst tumor markers (including HER2) on-treatment, whereas those that did not exhibited higher heterogeneity across immune markers (**Figure 3e**). Higher immune marker heterogeneity on-treatment in the non-pCR cases may reflect a less consistent immune response with some regions experiencing a greater immune influx than others. We did not observe higher pre-treatment HER2 heterogeneity in the non-pCR cases compared to pCR cases, as recently suggested based on FISH [28]; and comparable regional heterogeneity amongst tumor markers was noted in pCR and non-pCR cases (**Supplementary Figure 14**).

We further analyzed the DSP data to investigate the composition of immune cells in panCK-enriched regions (used in all other analyses) relative to the surrounding panCK-negative regions designed to capture the neighboring microenvironment (**Supplementary Figure 15**). Prior to treatment, both T cell (CD3, CD4, CD8) and macrophage (CD68) markers were more prevalent in the surrounding microenvironment, while CD56-positive NK cells and immunosuppressive markers (e.g. VCTN1, PD-L1, IDO) were higher in panCK-enriched regions (**Figure 3f**). These findings are consistent with T cell exclusion, where IDO and PD-L1 are thought to impair intratumor proliferation of effector T cells [29]. A similar immune profile was observed during HER2-targeted therapy alone. However, post-treatment, in the non-pCR cases with tumor remaining, most immune markers were more prevalent in the panCK-enriched regions, compared to the neighboring microenvironment, including CD8 and CD68. Both prior to treatment and on-treatment, immune cell localization was similar in patients that

achieved a pCR and those that did not (**Supplementary Figure 16**), as well as for ER-positive versus ER-negative cases (**Supplementary Figure 17**).

The geospatial distribution of tumor and immune cells has been associated with relapse and survival in multiple tumor types [30,31]. Here we investigated the relationship between treatment and the tumor-microenvironment border using perimetric complexity, which is proportional to the perimeter of a region squared, divided by the area of the region [32] (Methods, **Supplementary Figure 18a**). Prior to treatment, no significant difference in perimetric complexity was observed between pCR and non-pCR cases (p=0.299, **Supplementary Figure 18b**). However, perimetric complexity decreased significantly on-treatment relative to pre-treatment (p=1.32e-6, **Supplementary Figure 18c**). These data suggest that treatment may affect the geographic distribution of tumor cells as well as tumor cell content. Indeed, the proliferative marker, Ki67, was highly correlated with perimetric complexity (**Supplementary Figure 19**). Thus, for highly proliferative tumors, the perimeter of the tumor-microenvironment border may be relatively larger, potentially allowing for increased crosstalk with the surrounding microenvironment.

DSP of paired pre- and on-treatment biopsies reveals features associated with pCR

Given the dramatic differences in treatment-associated changes in pCR cases compared to non-pCR cases (**Figure 2b**), we next sought to evaluate whether DSP protein marker status prior to treatment or early during the course of therapy could be used to predict pCR. We used an L2-regularized logistic regression to classify tumors by pCR status based on average DSP protein expression levels across multiple ROIs profiled pre-treatment, on-treatment, or the average marker expression both pre-treatment and on-treatment (denoted “on- + pre-treatment”) and evaluated model performance via nested cross validation within the discovery cohort (Methods). Tumors with data at both timepoints were utilized in this analysis (n=23 cases, **Supplementary Figure 2**). A model based on on-treatment protein expression outperformed that based on pre-treatment protein expression (mean AUROC=0.728 versus 0.614) and performed comparably to a model incorporating both on-treatment and pre-treatment protein expression levels (mean AUROC=0.733) (**Figure 4a**). A classifier trained using both immune and tumor markers outperformed a model using tumor markers alone, highlighting the importance of simultaneous tumor and immune profiling to predict therapy response (**Figure 4b**).

For the DSP protein on- plus pre-treatment classifier, we investigated the importance of multi-region sampling and heterogeneity by extending the model to incorporate both the mean marker expression across all regions and the standard error of the mean (SEM) for each marker between regions (Methods). This analysis was restricted to patients with at least 3 regions profiled at both timepoints (n=16, Methods). We found that utilizing the mean immune values and the SEM for tumor markers outperformed a model based on mean values for both tumor and immune markers (**Supplementary Figure 20**), suggesting that classifiers that capture the heterogeneity amongst tumor markers may improve the prediction of pCR.

We next compared the performance of the DSP protein on- plus pre-treatment classifier with features previously associated with outcome (ER status and PAM50 subtype), where models were again evaluated via cross-validation in the discovery cohort. Of note, a model based on ER status and HER2-enriched PAM50 status performed poorly in this cohort (mean AUROC=0.589) and the addition of these two features to the DSP protein on- plus pre-treatment data set did not improve the AUROC (**Figure 4c**). Given the availability of bulk transcriptomic data for these cases, we also built a model using paired on- and pre-treatment bulk RNA expression data for the 37 markers that overlapped with the DSP protein panel. This model also performed significantly worse than that based on the DSP protein data (**Figure 4d**, p<0.0001 via cross-validation). This is not surprising since amongst the 37 overlapping DSP and bulk RNA expression markers, only 16 were positively correlated pre-treatment

(**Supplementary Figure 21**). Various factors may contribute to the lack of strong correlation between protein and RNA expression levels, including panCK enrichment, RNA transience/degradation [33], and post-translational regulation [34], where protein expression is a more proximal readout of cellular phenotype.

DSP predicts pCR in an independent validation cohort

In light of these promising findings, we further sought to evaluate the performance of the DSP protein on- plus pre-treatment classifier in an independent cohort (n=29) of patients from the TRIO B07 clinical trial (**Supplementary Figure 22a**). As with the discovery cohort, an average of four panCK-positive regions were profiled from each pre- and on-treatment tumor tissue and the same panel of 40 protein antibodies was utilized. The change in markers on-treatment relative to pre-treatment mirrored that observed in the discovery cohort (**Supplementary Figure 22b**): T-cell markers (CD3, CD4, and CD8) increased while the tumor markers HER2 and Ki67 showed the most significant decrease. Similar to the discovery cohort, in the validation cohort, we observed that on-treatment versus pre-treatment protein expression differences were more dramatic in tumors that ultimately underwent pCR (**Supplementary Figure 22c**). AUROC performance of the L2-regularized logistic regression model, trained in the discovery cohort, was evaluated in the validation (test) cohort. The performance of the DSP protein on- plus pre-treatment model in predicting pCR was comparably high in the discovery (assessed via cross-validation, mean AUROC = 0.733) and validation (assessed via train-test, AUROC=0.725) cohorts (**Figure 5a**). In the on- plus pre-treatment classifier, which was trained in the discovery cohort and tested in the validation cohort, the marker with the largest L2-regularized coefficient was on-treatment CD45 protein levels. In general, features with large coefficients included on-treatment markers that represent tumor-infiltrating lymphocyte and macrophage populations (CD45, CD44, CD66B) (**Figure 5b**). On-treatment HER2 protein expression had a negative coefficient in the model, consistent with poor outcome being associated with high HER2 levels during treatment.

Given the widespread use of trastuzumab in current neoadjuvant treatment paradigms [35], we further assessed model performance for the best performing on-+pre-treatment DSP protein model in the trastuzumab-containing cases (arms 1 and 3, n=19). Again, we observed similar model performance and marker coefficients in the full discovery cohort and in the subset of cases in the validation cohort who were treated with trastuzumab (**Supplementary Figure 23**). The validation of these findings in an independent cohort demonstrates the potential of multiplex spatial proteomic profiling to predict which patients will respond early during HER2-targeted therapy, such that subsequent therapy could be tailored accordingly.

Discussion

Bulk genomic and transcriptomic profiling has been a mainstay of cancer biomarker discovery efforts in recent years. However, admixture amongst heterogeneous cellular populations complicates the analysis of such data, issues which are compounded when studying longitudinal samples, where the changing composition and localization of cell populations may reflect the biology of disease progression or mechanisms of treatment response. Indeed, efforts to establish validated biomarkers of response to HER2-targeted therapy based on bulk genomic and transcriptomic profiling have met with limited success to date in other trial cohorts and in TRIO-US B07 [7,9,17]. We reasoned that *in situ* proteomic profiling of the tumor-immune microenvironment through therapy would circumvent the limitations of dissociative techniques and improve our ability to uncover features associated with response to neoadjuvant HER2-targeted therapy. Here we used the DSP technology [22] to

simultaneously profile 40 tumor and immune markers *en bloc* on a single 5μm section of archival tissue from breast tumors sampled before, during, and after neoadjuvant HER2-targeted therapy in the TRIO-US B07 clinical trial. In order to enhance signal while accounting for intra-tumor heterogeneity, we employed a pan-CK masking strategy to enrich for tumor cells and co-localized immune cells across multiple regions per sample.

DSP of longitudinal breast biopsies from this trial cohort uncovered changes associated with therapy, including markedly decreased HER2 and downstream AKT signaling on-treatment, accompanied by increased CD45 and CD8 expression, consistent with infiltrating leukocytes and cytotoxic T-cells, respectively. By the time of surgery, following a full course of neoadjuvant therapy, the tumor-immune composition changed considerably with increased CD56 expression in non-pCR cases, potentially reflecting NK cell-mediated killing of chemotherapy-stressed tumor cells [27]. Changes in both tumor and immune markers on-treatment were more dramatic in tumors that went on to achieve a pCR and, critically, on-treatment and pre-treatment protein expression robustly predicted response in an independent validation cohort (AUROC = 0.725). Whereas on-treatment protein expression levels were similarly predictive of pCR, neither pre-treatment protein expression, established predictive features, nor bulk pre- and on-treatment gene expression data were predictive in this cohort, emphasizing the superiority of this novel multiplexed spatial proteomic biomarker and its potential utility for patient stratification. Our findings thus address a critical unmet clinical need given the considerable emphasis devoted to identifying subsets of the population in which therapy should be escalated, for example by combining HER2-targeted agents, or safely de-escalated, for example through shortening or omission of chemotherapy and its associated toxicities [36,37]. While numerous biomarkers have been considered to help guide personalized targeting of escalated versus de-escalated approaches in early -stage HER2+ breast cancer – including imaging, circulating tumor DNA, and pre-treatment immune scores or intrinsic subtype – there is currently no validated biomarker that can guide patient stratification. The increasing plethora of options for HER2-targeted therapy, including novel highly effective but potentially toxic agents [38,39], combined with great heterogeneity in response make HER2+ breast cancer the ideal setting for the development of optimally personalized therapy over the next decade. The next step towards clinical translation of the multiplex proteomic biomarker described here will involve prospective validation in a neoadjuvant clinical trial that selects therapy based on biomarker status [40]. Of note, the size of this multiplex panel (40 proteins), is on par with that of existing breast cancer diagnostic assays such as Mammaprint (70 genes), PAM50 Prosigna (50 genes) and OncotypeDX (21 genes).

More generally, our results illustrate the feasibility and power of multiplex *in situ* proteomic analysis of archival tissue samples to provide proximal readouts of tumor and immune cell signaling through therapy. Many signaling proteins/phospho-proteins, including those profiled here, are considered protein network bottlenecks and integrate mutational and transcriptional changes [41,42], making this a particularly powerful approach to studying treatment-associated changes. Importantly, DSP antibody panels can now be customized, allowing for inclusion of additional/alternate markers of interest, such as ER or other tumor-specific markers and signaling pathways. This work also illuminates study design considerations, including the value of panCK enrichment of tumor cells (or other markers to enrich for specific cell populations) and multi-region profiling to capture regional tumor heterogeneity and treatment-associated changes, concepts that should be broadly applicable to other epithelial tumor types. Of note, the DSP measurements in this study were based on regional analysis of defined cellular populations comprised of ~300-600 cells. Although single cell resolution was not necessary for the development of the novel biomarker described within (and may indeed complicate the clinical implementation of such approaches), such data can further enable the identification of cell states and cell-cell interactions and is likely to be an area of future research, fueled by advances in the resolution and throughput of spatial profiling technologies.

Online Methods

Cohort selection

The TRIO-US B07 clinical trial was a randomized, multicenter study that included 130 women with stage I-III unilateral, HER2-positive breast cancer [17]. The IRB at the University of California Los Angeles (UCLA) approved the clinical trial TRIO-US B07 (08-10-035). The IRB at Stanford approved the use of the TRIO-US B07 clinical trial specimens for correlative studies in the Curtis Lab (eProtocol #32180). Informed consent was obtained from all participants. This covers consent from patients for their samples to be shared with other researchers. Enrolled patients were randomly assigned to three treatment groups, dictating the type of targeted therapy namely trastuzumab, lapatinib, or trastuzumab and lapatinib in combination. Breast tumor biopsies were obtained prior to treatment and following 14-21 days of the assigned HER2-targeted therapy (without chemotherapy), which was followed by six cycles of the assigned HER2-targeted treatment plus docetaxel and carboplatin given every three weeks and surgery. For each timepoint, core biopsies or surgical tissue sections were obtained and stored as either fresh frozen or FFPE material. In total, 28 cases with FFPE samples available from all three timepoints (pre-treatment, on-treatment, and at surgery) were selected for inclusion in the discovery cohort based on sample availability and quality, with balancing by pCR status and ER status (**Figure 1a, Supplementary Figure 1**). An additional 29 cases with FFPE samples available pre-treatment and on-treatment were selected for the validation cohort in order to assess performance of the classifier (**Supplementary Figure 22**). Of note, the validation cohort was used exclusively to evaluate model performance. All other analyses are based on the discovery cohort. Tumor cellularity was assessed by a board-certified breast pathologist (GRB) using tumor sections stained with hematoxylin and eosin. Samples with cellularity of 0 were omitted from further analysis. For other tissue sections estimated to have a cellularity of 0, tumor cells were identified on the FFPE sections used to perform DSP (distinct from the H&E sections used for pathology review), and these were included in the analysis (**Supplementary Figure 2**). FFPE blocks were sectioned at 5 μm thickness and stored at 4°C for less than three weeks prior to the DSP experiment.

DSP Data Generation and Analysis

Digital Spatial Profiling (DSP, Nanostring FOR RESEARCH USE ONLY. Not for use in diagnostic procedures) was performed as previously described [22]. In brief, tissue slides were stained with a multiplexed panel of protein antibodies contained a photocleavable indexing oligo, enabling subsequent readouts (**Supplementary Figure 3**). Regions of interest (ROIs) were selected on a DSP prototype instrument and illuminated using UV light. Released indexing oligos from each ROI were collected and deposited into designated wells on a microtiter plate, allowing for well indexing of each ROI during nCounter readout (direct protein hybridization). Custom masks were generated using an ImageJ pipeline, as described previously [43]. For each tissue sample, counts for each marker were obtained from an average of four (range 1-7) panCK-enriched (panCK-E) ROIs. Raw protein counts for each marker in each ROI were generated using nCounter [44]. The raw counts were ERCC-normalized (based on the geometric mean of the three positive control markers). Histone H3 was used as a housekeeping marker and ROIs with extreme Histone H3 (more than three standard deviations away from the mean) were filtered (< 1% of ROIs). The geometric mean of two IgG antibodies were used to calculate the background noise and we noted markers with signal to noise ratio <3x (**Supplementary Figure 24**). Immune markers were normalized based on ROI area to measure total density of immune content in the region. Tumor markers were normalized using the housekeeping antibody (Histone H3) in order to capture the status of the cancer signaling pathways on a per cell basis. As further quality control, area normalization factors and housekeeping normalization factors were compared per ROI, and ROIs were filtered with disparate normalization factors across the two methods (this represented 6% of all ROIs). All normalized counts were converted to log2 space for downstream analysis. The analyses carried out in this study are

comparative in nature (e.g. pre-treatment vs on-treatment, pCR vs non-pCR) and are robust to variations in normalization methods.

Bulk mRNA Expression Analysis

RNA was extracted using the RNeasy Mini Kit (Qiagen), quantified by the Nanodrop One Spectrophotometer (ThermoFisher Scientific). RNA samples were labeled with cyanine 5-CTP or cyanine 3-CTP (Perkin Elmer) using the Quick AMP Labeling Kit (Agilent Technologies). Gene Expression Microarray experiments were performed by comparing each baseline sample to samples taken after 14-21 days of HER2-targeted therapy (on-treatment). Each on-treatment sample was compared to the pre-treatment sample from the same patient. Limma[45,46] was used for background correction ("normexp"), within-array normalization ("loess"), between-array normalization, and for averaging over replicate probes. For the downstream analyses, including batch correction and comparisons with the DSP cohort, the normalized counts were converted to log2 space. Combat [47] was used to remove potential batch affects associated with microarray run date. PAM50 status pre-treatment and on-treatment was inferred using AIMS (Absolute Intrinsic Molecular Subtyping), an N-of-1 algorithm that is robust to variations in data set composition [48]. This approach was utilized given the expected preponderance of HER2-enriched cases in this cohort.

Correlation Analyses

The Spearman rank correlation between DSP protein data and bulk RNA data was computed for pre-treatment samples using the average of all DSP ROIs (both panCK-enriched ROIs and surrounding microenvironment-enriched ROIs) per patient. Plots showing the correlation between protein markers (**Figure 1c, 2c**) were overlapped with hierarchical clustering in the form of black squares. The difference between the distribution of correlation values in pCR versus the non-pCR cases was evaluated using a two-sided Wilcoxon two-sample t-test.

Comparative Analyses

For the comparative analyses (e.g. pre-treatment vs on-treatment, pCR vs non-pCR, panCK-enriched vs panCK-negative) of DSP protein data, where multiple regions were sampled per patient, we utilized a linear mixed-effects model with blocking by patient [49]. This model allows for marker levels to be compared in a patient-matched manner while controlling for differences in the number of ROIs profiled per patient. The coefficient of the fixed effect is the change attributable to that variable (x-axis of volcano plots), and the p-value used to calculate false discovery rates (y-axis of volcano plots) is based on the t-value (a measure of the size of the difference relative to the variation in the sample data). False discovery rates (FDR) were computed using the Benjamini & Hochberg procedure [50], and an FDR-adjusted p-value of 0.05 was set as the significance threshold.

Region Subsampling

We sought to investigate the impact of utilizing a single randomly selected region per tissue sample, rather than multiple regions, when assessing on- versus pre-treatment protein expression changes. For these analyses (**Supplementary Figure 11**), we performed 100 iterations in which a single region was selected from each tissue and computed fold changes and corresponding p-values averaged over these 100 experiments. The number of random samplings was chosen empirically by raising the number of iterations beyond the number required to make the resulting output robust to further increases in the number of iterations used (p-value convergence).

L2-regularized Logistic Regression Using Molecular Data

Models and features

L2-logistic regression using liblinear as a solver was used for classification of pCR vs non-pCR cases. Marker values pre-treatment and on-treatment were averaged across all ROIs to derive a composite value for each marker for that timepoint. Five patients were excluded from the models because data

was available only at a single timepoint (**Supplementary Figure 2**). Mean DSP marker expression features were used in models comparing patient timepoints, tumor versus immune markers, DSP protein features versus established predictive features (ER status and PAM50 classification), and DSP protein versus Bulk RNA features (using RNA gene transcripts corresponding to DSP protein markers). To assess heterogeneity, standard error of the mean (SEM) was calculated for marker values across all ROIs for tissues with at least 3 ROI to derive a composite value for each marker for that timepoint. These SEM features were used in combination with mean expression features in models assessing the predictive value of heterogeneity.

Model comparisons and evaluation of performance via internal cross-validation
Model performance was evaluated and models compared using nested cross-validation using the python package sklearn [51]. Data were divided into N folds using stratified sampling ("stratified cross-validation"). The number of folds was chosen based on the number of cases in the non-pCR group (the class with fewer cases) such that the testing data would contain two cases from each class. Each model was trained using N-1 folds and scored using mean AUROC on the remaining fold. This process was iteratively repeated holding out a different fold each time. The L2-penalization weight was chosen using stratified cross-validation within the N-1 training dataset, with the weight associated with highest mean accuracy within this inner cross-validation selected for scoring. This nested cross-validation process was repeated 100 times using randomly generated folds. Model scores were then compared using an unpaired two-sided t-test with Holm-Bonferroni correction for multiple hypotheses. ROC curves were generated by averaging across the ROC curves from the 100 repeats of N-fold cross-validation, with each repeat containing a different random split of folds.

Evaluation of model performance in an independent validation cohort
As described above, marker values pre-treatment and on-treatment were averaged across all ROIs to derive a composite value for each marker for that timepoint. Model selection was carried out using cross-validation as described above. The best performing model was selected and trained using the entire discovery cohort. Finally, model performance based on the AUROC was evaluated in the independent validation (test) cohort.

Metrics of Heterogeneity
Marker heterogeneity was calculated as the mean squared error from the analysis of variance done on a linear model with marker values as the dependent variable and patient identity as the independent variable (the data set was subsetted to the particular timepoint or clinical outcome of interest).

Perimetric Complexity
Perimetric complexity was computed for the panCK-enriched binary masks for each ROI using ImageJ[32]. A linear mixed-effects model[49] with blocking by patient was used to the compare the perimetric complexities of all the panCK-enriched regions pre-treatment and on-treatment regions and for cases that achieved a pCR versus those for cases that did not achieve a pCR.

Data Availability

Data associated with the manuscript is available online:
https://github.com/cancersysbio/BreastCancerSpatialProteomics

Code Availability

Code associated with the manuscript is available online:
https://github.com/cancersysbio/BreastCancerSpatialProteomics

Acknowledgments
We thank NanoString for technical support. We thank members of the Curtis lab, especially Zheng Hu and Jose A. Seoane, for feedback on the manuscript. This project was supported by awards from the National Institutes of Health/National Cancer Institute (R01CA182514) and the Breast Cancer Research Foundation to C.C. KLM is supported by F30CA239313-02 from the NIH/NCI. JLC is supported by a Damon Runyon Physician-Scientist Training Award and a Susan G. Komen Postdoctoral Fellowship Award (PDR17481769).

Author Contributions
KLM analyzed the data. RJ and EK contributed to statistical analyses. KLM, ZM, MK, ZZ, MH, JB contributed to data acquisition. GRB performed pathology review. SAH and DJS led the clinical trial and oversaw sample collection. KLM, JC and CC interpreted the data. KLM and CC wrote the manuscript. CC conceived of and supervised the study. All authors read and approved the final manuscript.

Disclosure of Potential Conflicts of Interest
SAH received contracted research and medical writing assistance from Ambrx, Amgen, Arvinas, Bayer, Daiichi-Sankyo, Genentech/Roche, GSK, Immunomedics, Lilly, Macrogenics, Novartis, Pfizer, OBI Pharma, Pieris, PUMA, Radius, Sanofi, Seattle Genetics, Dignitana. MK, ZZ, MH, JB are all employees of NanoString Inc. and declare that there are competing interests. DJS received research funding from Pfizer, Novartis, Syndax, Millenium Pharmaceuticals, Aileron Therapeutics, Bayer, and Genentech, owned stock in Biomarin, Amgen, Seattle Genetics, and Pfizer, served on the Board of Directors for BioMarin, and performed consulting/advisory board work for Eli Lilly, Novartis, Bayer, and Pfizer. CC is a scientific advisor to GRAIL and reports stock options as well as consulting for GRAIL and Genentech. KLM, JLC, ZM, RJ, EK, GB, have no conflicts of interest to report.

References

1. Slamon, D.J.*, et al.* Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235**, 177-182 (1987).
2. de Azambuja, E.*, et al.* Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTTO): survival outcomes of a randomised, open-label, multicentre, phase 3 trial and their association with pathological complete response. *Lancet Oncol* **15**, 1137-1146 (2014).
3. Gianni, L.*, et al.* Neoadjuvant and adjuvant trastuzumab in patients with HER2-positive locally advanced breast cancer (NOAH): follow-up of a randomised controlled superiority trial with a parallel HER2-negative cohort. *Lancet Oncol* **15**, 640-647 (2014).
4. Gianni, L.*, et al.* Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. *Lancet Oncol* **13**, 25-32 (2012).
5. Rimawi, M.F.*, et al.* Multicenter Phase II Study of Neoadjuvant Lapatinib and Trastuzumab With Hormonal Therapy and Without Chemotherapy in Patients With Human Epidermal Growth Factor Receptor 2-Overexpressing Breast Cancer: TBCRC 006. *Journal of Clinical Oncology* **31**, 1726-1731 (2013).
6. Veeraraghavan, J.*, et al.* De-escalation of treatment in HER2-positive breast cancer: Determinants of response and mechanisms of resistance. *Breast* **34 Suppl 1**, S19-S26 (2017).
7. Llombart-Cussac, A.*, et al.* HER2-enriched subtype as a predictor of pathological complete response following trastuzumab and lapatinib without chemotherapy in early-stage HER2-positive breast cancer (PAMELA): an open-label, single-group, multicentre, phase 2 trial. *Lancet Oncology* **18**, 545-554 (2017).
8. Gianni, L.*, et al.* 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 HER2-negative cohort. *Lancet* **375**, 377-384 (2010).
9. Lesurf, R.*, et al.* Genomic characterization of HER2-positive breast cancer and response to neoadjuvant trastuzumab and chemotherapy-results from the ACOSOG Z1041 (Alliance) trial. *Ann Oncol* **28**, 1070-1077 (2017).
10. Fumagalli, D.*, et al.* RNA Sequencing to Predict Response to Neoadjuvant Anti-HER2 Therapy: A Secondary Analysis of the NeoALTTO Randomized Clinical Trial. *JAMA Oncol* **3**, 227-234 (2017).
11. Prat, A.*, et al.* Research-based PAM50 subtype predictor identifies higher responses and improved survival outcomes in HER2-positive breast cancer in the NOAH study. *Clin Cancer Res* **20**, 511-521 (2014).
12. Tanioka, M.*, et al.* Integrated Analysis of RNA and DNA from the Phase III Trial CALGB 40601 Identifies Predictors of Response to Trastuzumab-Based Neoadjuvant Chemotherapy in HER2-Positive Breast Cancer. *Clin Cancer Res* **24**, 5292-5304 (2018).
13. Loi, S.*, et al.* Tumor infiltrating lymphocytes are prognostic in triple negative breast cancer and predictive for trastuzumab benefit in early breast cancer: results from the FinHER trial. *Ann Oncol* **25**, 1544-1550 (2014).
14. Nuciforo, P.*, et al.* A predictive model of pathologic response based on tumor cellularity and tumor-infiltrating lymphocytes (CelTIL) in HER2-positive breast cancer treated with chemo-free dual HER2 blockade. *Annals of Oncology* **29**, 170-177 (2018).
15. Varadan, V.*, et al.* Immune Signatures Following Single Dose Trastuzumab Predict Pathologic Response to PreoperativeTrastuzumab and Chemotherapy in HER2-Positive Early Breast Cancer. *Clin Cancer Res* **22**, 3249-3259 (2016).
16. Binnewies, M.*, et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nature Medicine* **24**, 541-550 (2018).
17. Hurvitz, S., Caswell-Jin JL, et al. Pathologic and molecular responses from a phase II randomized trial of neoadjuvant trastuzumab and/or lapatinib in HER2+ breast cancer (TRIO-US B07). *Submitted*.
18. Huober, J.*, et al.* Survival outcomes of the NeoALTTO study (BIG 1-06): updated results of a randomised multicenter phase III neoadjuvant clinical trial in patients with HER2-positive primary breast cancer. *Eur J Cancer* **118**, 169-177 (2019).
19. Broglio, K.R.*, et al.* Association of Pathologic Complete Response to Neoadjuvant Therapy in HER2-Positive Breast Cancer With Long-Term Outcomes: A Meta-Analysis. *JAMA Oncol* **2**, 751-760 (2016).
20. Cortazar, P.*, et al.* Pathological complete response and long-term clinical benefit in breast cancer: the CTNeoBC pooled analysis. *Lancet* **384**, 164-172 (2014).
21. Toki, M.I.*, et al.* Validation of novel high-plex protein spatial profiling quantitation based on NanoString's Digital Spatial Profiling (DSP) technology with quantitative fluorescence (QIF). *Cancer Research* **77**(2017).
22. Christopher R. Merritt, G.T.O., Sarah E. Church, Kristi Barker, Patrick Danaher, Gary Geiss, Margaret Hoang, Jaemyeong Jung, Yan Liang, Jill McKay-Fleisch, Karen Nguyen, Zach Norgaard, Kristina Sorg, Isaac Sprague, Charles Warren, Sarah Warren, Phillippa J. Webster, Zoey Zhou, Daniel R. Zollinger, Dwayne L. Dunaway, Gordon B. Mills & Joseph M. Beechem Multiplex digital spatial profiling of proteins and RNA in fixed tissue. *Nat Biotechnol* **38**, 586-599 (2020).
23. Untch, M.*, et al.* Neoadjuvant treatment with trastuzumab in HER2-positive breast cancer: results from the GeparQuattro study. *J Clin Oncol* **28**, 2024-2031 (2010).
24. Shao, M.M.*, et al.* Keratin expression in breast cancers. *Virchows Arch* **461**, 313-322 (2012).
25. Valachis, A., Nearchou, A., Lind, P. & Mauri, D. Lapatinib, trastuzumab or the combination added to preoperative chemotherapy for breast cancer: a meta-analysis of randomized evidence. *Breast Cancer Res Treat* **135**, 655-662 (2012).
26. Pohlmann, P.R., Mayer, I.A. & Mernaugh, R. Resistance to Trastuzumab in Breast Cancer. *Clin Cancer Res* **15**, 7479-7491 (2009).
27. Zingoni, A.*, et al.* Natural Killer Cell Response to Chemotherapy-Stressed Cancer Cells: Role in Tumor Immunosurveillance. *Front Immunol* **8**(2017).
28. Metzger, O.*, et al.* HER2 heterogeneity as a predictor of response to neoadjuvant T-DM1 plus pertuzumab: Results from a prospective clinical trial. *Journal of Clinical Oncology* **37**(2019).
29. Joyce, J.A. & Fearon, D.T. T cell exclusion, immune privilege, and the tumor microenvironment. *Science* **348**, 74-80 (2015).

30. AbdulJabbar, K.*, et al.* Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nat Med* **26**, 1054-1062 (2020).
31. Keren, L.*, et al.* A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* **174**, 1373-1387 e1319 (2018).
32. Watson, A.B. Perimetric Complexity of Binary Digital Images: Notes on Calculation and Relation to Visual Complexity. *Mathematica* **14**(2012).
33. Schwanhausser, B.*, et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342 (2011).
34. Truitt, M.L. & Ruggero, D. New frontiers in translational control of the cancer genome. *Nature Reviews Cancer* **16**, 288-304 (2016).
35. Zhang, B. & Hurvitz, S. Long-term outcomes of neoadjuvant treatment of HER2-positive breast cancer. *Clin Adv Hematol Oncol* **14**, 520-530 (2016).
36. Pernas, S. & Tolaney, S.M. HER2-positive breast cancer: new therapeutic frontiers and overcoming resistance. *Ther Adv Med Oncol* **11**, 1758835919833519 (2019).
37. Goutsouliak, K.*, et al.* Towards personalized treatment for early stage HER2-positive breast cancer. *Nat Rev Clin Oncol* (2019).
38. Modi, S.*, et al.* Trastuzumab Deruxtecan in Previously Treated HER2-Positive Breast Cancer. *N Engl J Med* **382**, 610-621 (2020).
39. Murthy, R.K.*, et al.* Tucatinib, Trastuzumab, and Capecitabine for HER2-Positive Metastatic Breast Cancer. *N Engl J Med* **382**, 597-609 (2020).
40. Javier Cortes, G.G., Manuel Ruiz Borrego, Agostina Stradella, Begoña Bermejo, Santiago Escrivá, Lourdes Calvo Martínez, Nuria Ribelles, Noelia Martinez, Cinta Albacar, Aleix Prat, Florence Dalenc, Kerrou Khaldoun, Peter Schmid, Marco Colleoni, Frederik Marmé, Noemia Afonso, Miguel Sampayo-Cordero, José Manuel Pérez-García, Antonio Llombart-Cussac. Chemotherapy (CT) de-escalation using an FDG-PET/CT (F-PET) and pathological response-adapted strategy in HER2[+] early breast cancer (EBC): PHERGain Trial. *Journal of Clinical Oncology* **38**, 503-503 (2020).
41. Sever, R. & Brugge, J.S. Signal transduction in cancer. *Cold Spring Harb Perspect Med* **5**(2015).
42. Yu, H., Kim, P.M., Sprecher, E., Trifonov, V. & Gerstein, M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput Biol* **3**, e59 (2007).
43. Amaria, R.N.*, et al.* Neoadjuvant immune checkpoint blockade in high-risk resectable melanoma. *Nat Med* **24**, 1649-1654 (2018).
44. Malkov, V.A.*, et al.* Multiplexed measurements of gene signatures in different analytes using the Nanostring nCounter Assay System. *BMC Res Notes* **2**, 80 (2009).
45. Ritchie, M.E.*, et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47 (2015).
46. Ritchie, M.E.*, et al.* A comparison of background correction methods for two-colour microarrays. *Bioinformatics* **23**, 2700-2707 (2007).
47. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).
48. Paquet, E.R., Lesurf, R., Tofigh, A., Dumeaux, V. & Hallett, M.T. Detecting gene signature activation in breast cancer in an absolute, single-patient manner. *Breast Cancer Res* **19**, 32 (2017).
49. Bates, D., Machler, M., Bolker, B.M. & Walker, S.C. Fitting Linear Mixed-Effects Models Using lme4. *J Stat Softw* **67**, 1-48 (2015).
50. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc B* **57**, 289-300 (1995).
51. Pedregosa, F.a.V., G. and Gramfort, A. and Michel, V., and Thirion, B.a.G., O. and Blondel, M. and Prettenhofer, P., and Weiss, R.a.D., V. and Vanderplas, J. and Passos, A. and & Cournapeau, D.a.B., M. and Perrot, M. and Duchesnay, E. Scikit-learn: Machine Learning in Python. *J Mach Learn Res* **12**, 2825-2830 (2011).
52. Schindelin, J.*, et al.* Fiji: an open-source platform for biological-image analysis. *Nat Methods* **9**, 676-682 (2012).

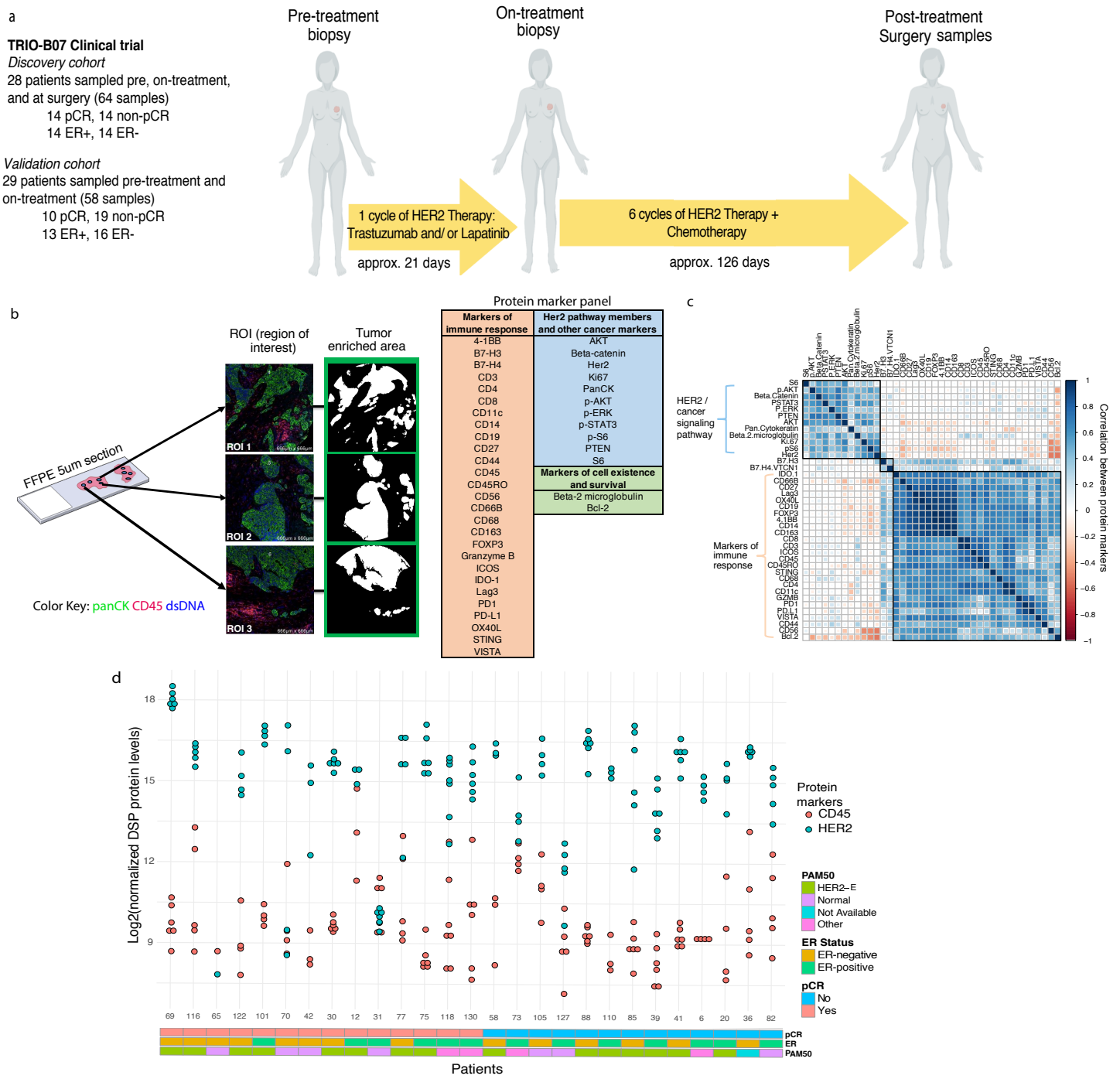


Figure 1. Spatial proteomic analysis of untreated HER2-positive breast tumors. a. Schematic overview of the discovery and validation cohorts analyzed with the GeoMx™ Digital Spatial Profiling (DSP) technology. Patients with invasive HER2+ breast cancer enrolled on the TRIO-US B07 clinical trial were treated with one cycle of the assigned HER2-targeted therapy followed by six cycles of the assigned HER2-targeted treatment plus chemotherapy (docetaxel+carboplatin). Tissue was obtained at three timepoints (pre-treatment, on-treatment, and post-treatment/surgery). b. Multiple regions of interest (ROIs) per tissue sample were selected based on pancytokeratin enrichment (panCK-E) and subject to spatial proteomic profiling of 40 tumor and immune markers. Protein counts were measured within phenotypic regions corresponding to the PanCK-E masks that includes tumor cells and co-localized immune cells and separately for the inverted mask corresponding to panCK-negative regions. c. Pairwise correlation of pre-treatment protein marker expression across all ROIs in the discovery cohort. Black squares indicate probes in the same hierarchical cluster. d. Inter-tumor and intra-tumor variability in HER2 and CD45 protein expression in untreated HER2-positive breast tumors from the discovery cohort, where each point corresponds to an ROI. Clinical characteristics, including pCR status, estrogen receptor (ER) status, and PAM50 subtype (based on gene expression profiling) are indicated.

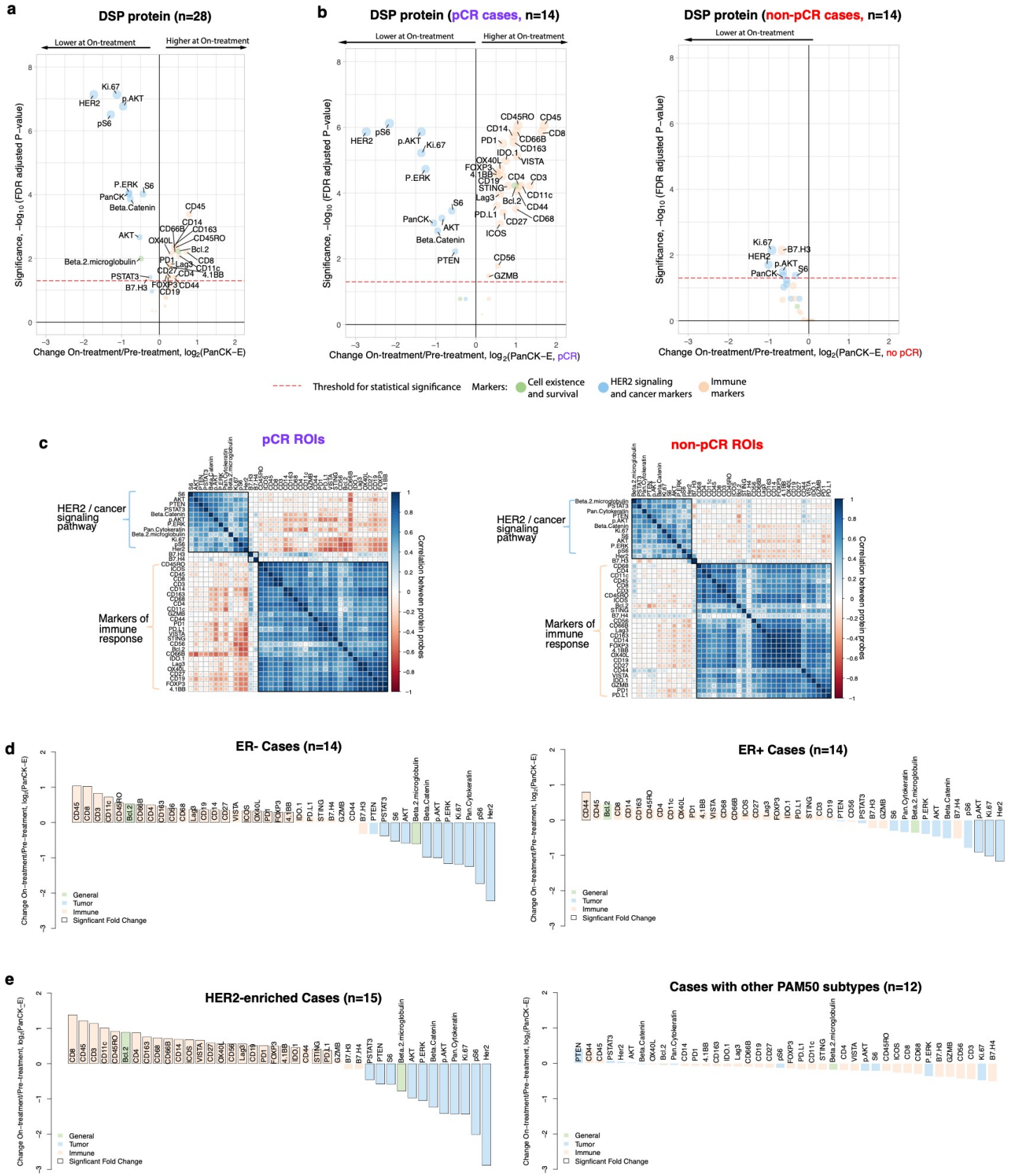
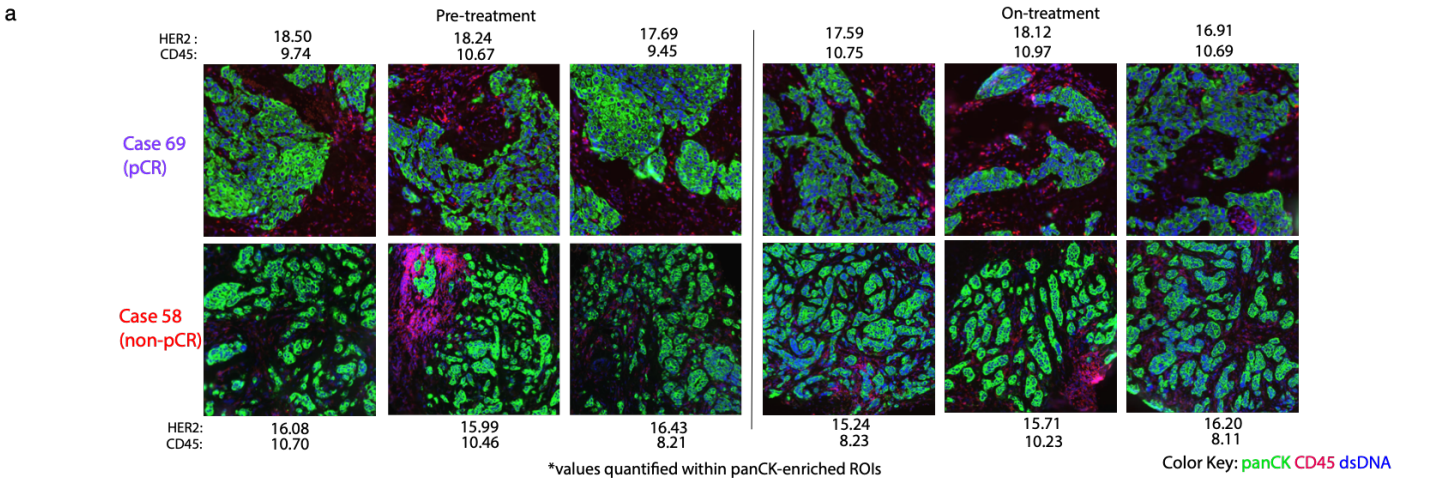
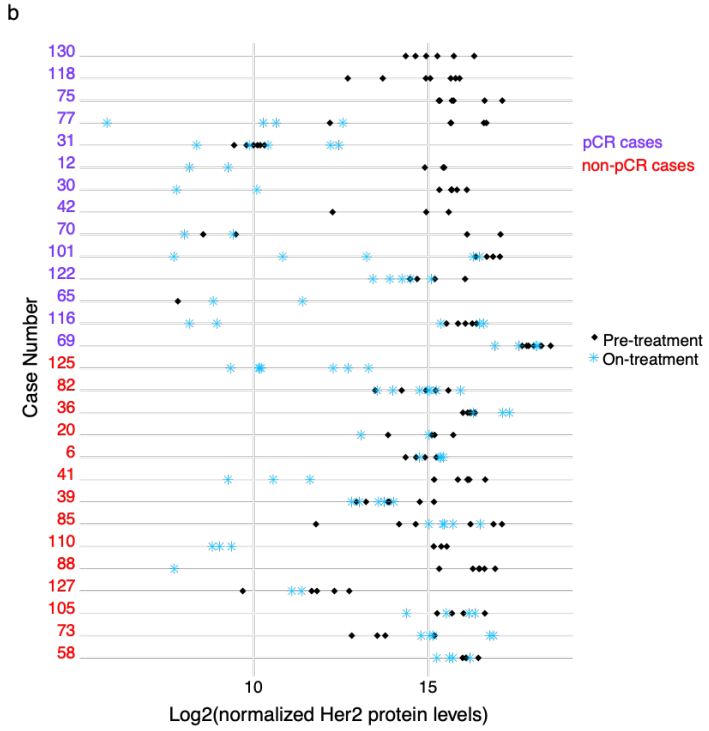


Figure 2. Spatial proteomic analysis reveals changes in cancer signaling and immune infiltration after short-term HER2 targeted therapy. a. Volcano plot demonstrating treatment-associated changes based on comparison of pre-treatment versus on-treatment protein marker expression levels in pancytokeratin-enriched (PanCK-E) regions. Significance, -log10(FDR adjusted p-value), is indicated along the y-axis. b. Volcano plots

demonstrating treatment-associated changes in pCR versus non-pCR cases. c. Pairwise correlation of protein markers in pCR versus non-pCR cases. Black squares demarcate hierarchical clusters. d. Waterfall plots illustrating treatment-associated changes (pre-treatment to on-treatment) in ER+ and ER- cases based on protein expression. e. Waterfall plots illustrating treatment-associated changes in DSP protein expression (pre-treatment to on-treatment) in HER2-enriched and non-HER2-enriched cases (n=7 normal-like, n=2 luminal B, n=2 basal, n=1 luminal A). Analyses were performed in the discovery cohort.

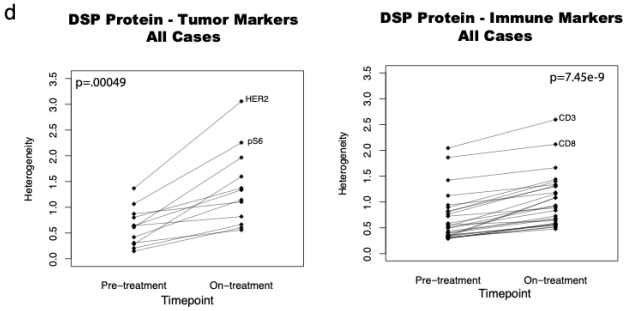




c

DSP HER2 protein heterogeneity

	Pre-treatment	On-treatment
Mean squared error *within* patients	1.37	15.62
Mean squared error *between* patients	3.06	27.19



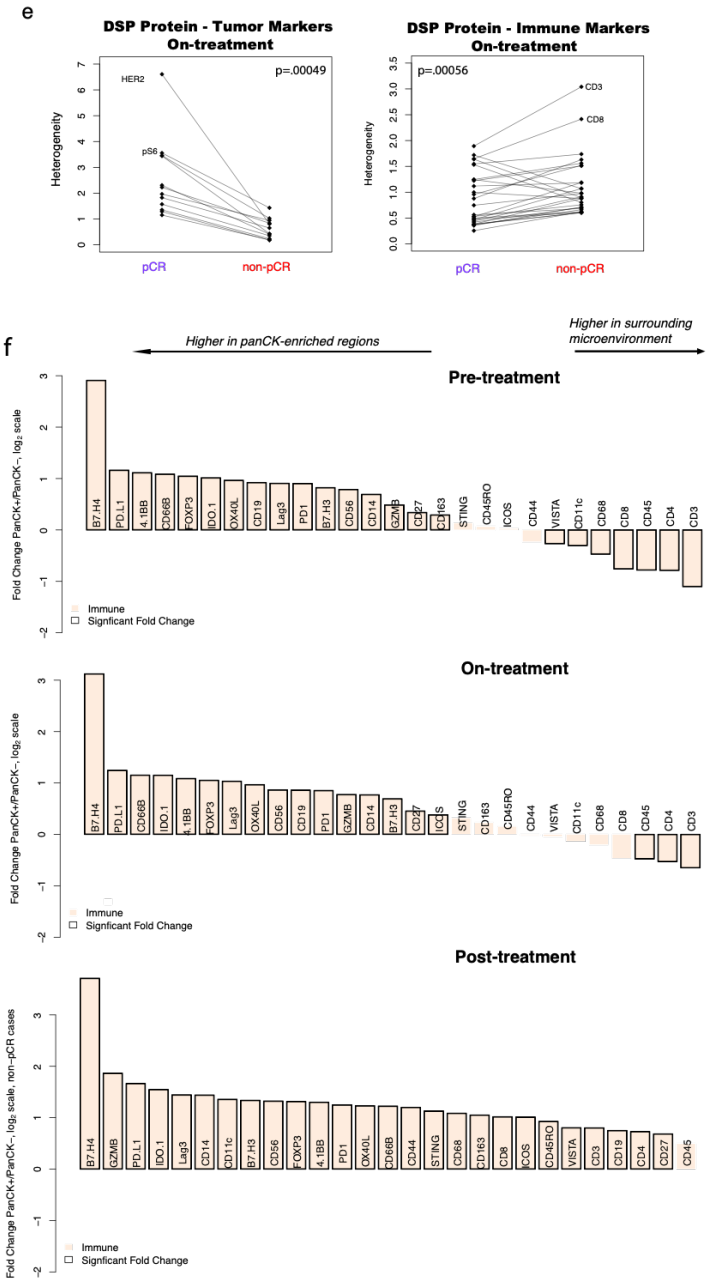
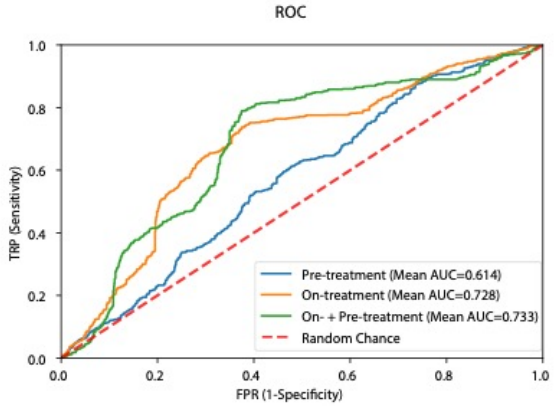


Figure 3. Increased heterogeneity of tumor and immune markers during HER2-targeted therapy and comparison of immune markers in panCK enriched and negative regions. a. Representative images of ROIs from two cases and quantification of HER2 and CD45 protein levels (log2 normalized) in panCK-enriched regions. b. Comparison of DSP HER2 protein levels pre-treatment and on-treatment for all regions profiled per case per timepoint. c. Comparison of the mean squared error in DSP HER2 protein expression pre-treatment versus on-treatment within and between patients. d. Pre-treatment versus on-treatment heterogeneity for each DSP tumor and immune marker. e. On-treatment heterogeneity in DSP protein markers for pCR and non-pCR cases. f. Waterfall plots of DSP protein data reveal differences in immune marker expression between immune-dense panCK-enriched regions and the surrounding panCK-negative regions profiled pre-treatment, on-treatment, and post-treatment. Heterogeneity was calculated as the mean squared error within patients based on analysis of variance in panels c, d, and e. P-values are based on a two-sided paired Wilcoxon signed rank test. Analyses are based on the discovery cohort.

a

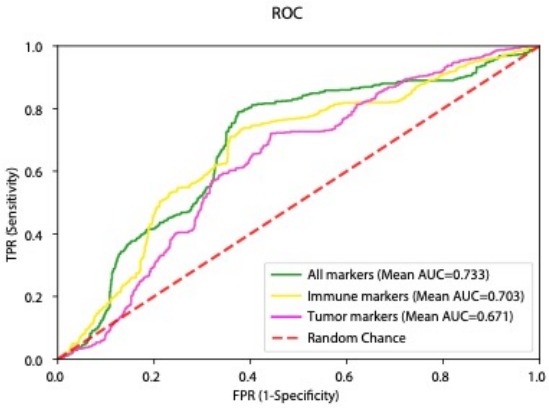
Comparison of timepoints: L2-regularized regression



Model comparisons by timepoint						
Data type	**Markers**	**Timepoint 1**	**Mean AUC 1**	**Timepoint 2**	**Mean AUC 2**	**Adjusted P-value**
DSP protein	All	Pre-treatment	0.614	On-treatment	0.728	<.0001
DSP protein	All	Pre-treatment	0.614	On-treatment + Pre-treatment	0.733	<.0001
DSP protein	All	On-treatment	0.728	On-treatment + Pre-treatment	0.733	0.783

b

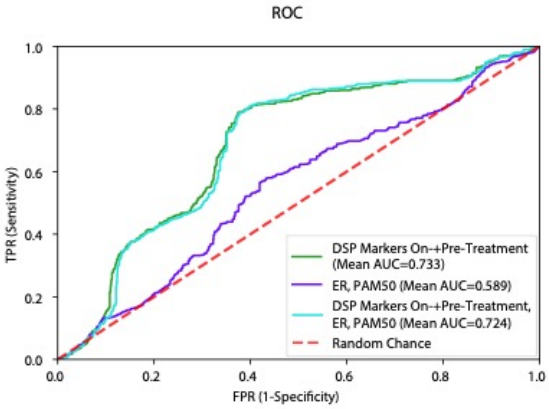
Comparison of markers: L2-regularized regression



Model comparisons by marker type						
Data type	**Timepoints**	**Marker Set 1**	**Mean AUC 1**	**Marker Set 2**	**Mean AUC 2**	**Adjusted P-value**
DSP protein	On-treatment + Pre-treatment	All Markers	0.733	Tumor Markers	0.671	0.0003
DSP protein	On-treatment + Pre-treatment	All Markers	0.733	Immune Markers	0.703	0.084
DSP protein	On-treatment + Pre-treatment	Immune Markers	0.703	Tumor Markers	0.671	0.084

c

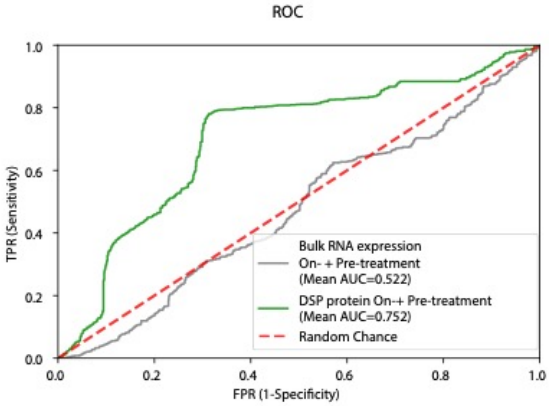
Comparison of data types: L2-regularized regression



Model comparisons by data type				
Data type 1	**Mean AUC 1**	**Data type 2**	**Mean AUC 2**	**Adjusted P-value**
DSP protein On- + Pre-treatment	0.733	ER status, PAM50 status (Pre-treatment)	0.589	<0.0001
DSP protein On- + Pre-treatment	0.733	ER status, PAM50 status (Pre-treatment) + DSP protein On- + Pre-treatment	0.724	0.546

d

Comparison of data types: L2-regularized regression



Model comparisons by data type				
Data type 1	**Mean AUC 1**	**Data type 2**	**Mean AUC 2**	**Adjusted P-value**
DSP protein On- + Pre-treatment	0.752	Bulk RNA expression On- + Pre-treatment	0.522	<0.0001

Figure 4. DSP of pancytokeratin-enriched paired pre- and on-treatment biopsies is associated with pathologic complete response in the discovery cohort and outperforms established markers. AUROC (Area Under Receiver Operating Characteristic) performance of various models were compared using nested cross-validation with Holm-Bonferroni correction for multiple hypotheses in the discovery (training) cohort. For (a-c), Receiver operating characteristic (ROC) curves were generated using cases with DSP panCK-enriched data from both the pre-treatment and on-treatment timepoints (n=23). a. ROC curves and statistical comparison of L2-regularized classifiers trained using DSP protein marker mean values (averaged across ROIs) pre-treatment, on-treatment and the combination of pre-treatment and on-treatment ("On- + Pre-treatment"). b. ROC curves and statistical comparison of DSP protein On- plus Pre-treatment L2-regularized classifiers trained using all marker, tumor marker, and immune marker mean values. Cross-region mean marker values from both the pre-treatment and on-treatment timepoints were used in this analysis. c. ROC curves and statistical comparison of the On- plus Pre-treatment DSP protein L2-regularized classifier to a model trained using ER and PAM50 status. These two models were compared to a model that incorporates On- plus Pre-treatment DSP protein data, ER and PAM50 status. d. ROC and statistical comparison of On- plus Pre-treatment L2-regularized classifiers trained using DSP protein marker mean values versus bulk RNA expression using RNA transcripts corresponding to the DSP protein markers. ROC curves were generated using cases with DSP panCK-enriched data and bulk expression data from both the pre-treatment and on-treatment timepoints (n=21).

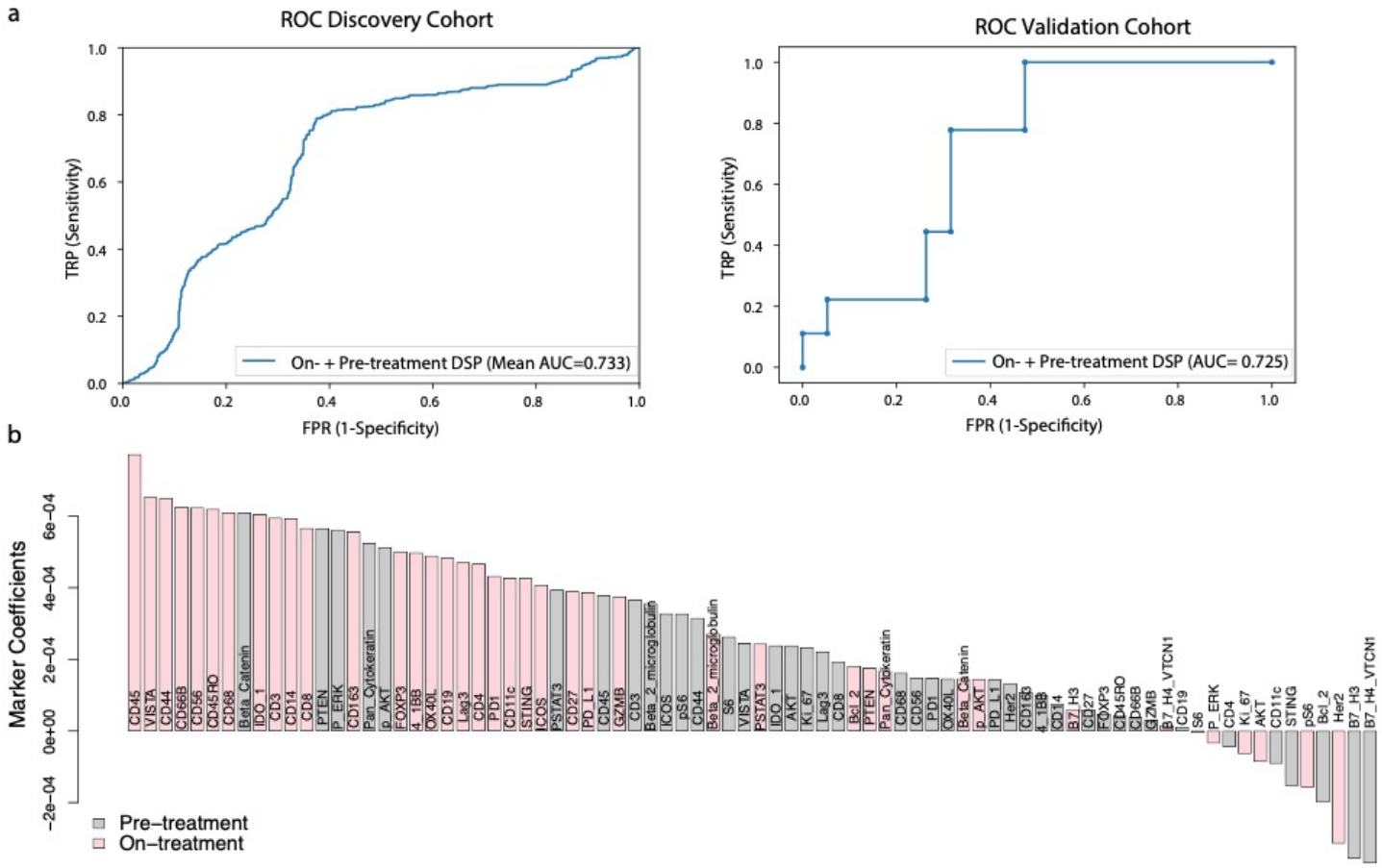


Figure 5. Validation of the DSP-based spatial proteomic biomarker in an independent cohort a. Receiver operating characteristic (ROC) curves for On- plus Pre-treatment DSP protein L2-regularized classifier in the discovery (training) cohort (n=23, assessed via cross-validation) and the validation (test) cohort (n=28, assessed via train-test) using the 40-plex DSP protein marker panel. b. Coefficients for each of the 40 markers in the L2-regularized On- plus Pre-treatment DSP protein model, trained in the discovery cohort, and tested in the validation cohort.

Discovery Cohort Characteristics	n=28 (100%)
Treatment Arm	
Arm 1 (Trastuzumab)	8 (29%)
Arm 2 (Lapatinib)	5 (19%)
Arm 3 (Trastuzumab + Lapatinib)	15 (54%)
pCR Status	
pCR	14 (50%)
non-pCR	14 (50%)
ER Status	
ER+	14 (50%)
ER-	14 (50%)
PAM50 Status: Pre-treatment	
HER2-Enriched	15 (54%)
Normal-like	7 (25%)
Basal	2 (7%)
LuminalA	1 (3.6%)
LuminalB	2 (7%)
No Data	1 (3.6%)

	pCR	non-pCR
Arm 1	4 (14%)	4 (14%)
Arm 2	2 (7.1%)	3 (11%)
Arm 3	8 (29%)	7 (25%)

	pCR	non-pCR
ER+	6 (21%)	8 (29%)
ER-	8 (29%)	6 (21%)

	ER+	ER-
Arm 1	2 (7.1%)	6 (21%)
Arm 2	1 (3.6%)	4 (14%)
Arm 3	11 (39%)	4 (14%)

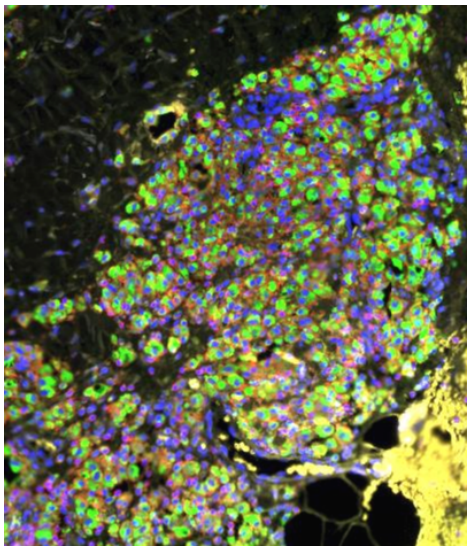
Supplementary Figure 1. Summary of the clinical characteristics of the TRIO-US B07 DSP discovery cohort, including treatment arm, pathologic complete response (pCR), estrogen receptor (ER) status, and PAM50 status inferred based on pre-treatment bulk expression data. Two-way contingency tables compare the distribution of ER status, pCR status, and treatment arm.

a

Case ID	pCR	Arm	ER status	Cellularity Pre-treatment	Cellularity On-treatment
73	0	1	1	50	30
110	0	3	1	30	0
85	0	2	0	30	30
58	0	1	0	20	20
39	0	3	1	20	20
36	0	3	0	20	30
88	0	1	0	15	0
41	0	1	0	10	20
82	0	3	1	10	0
127	0	3	1	5	1
6	0	3	1	5	5
105	0	2	0	2	2
20	0	3	1	2	50
125	0	2	1	0	2
30	1	2	0	80	0
122	1	1	0	30	5
77	1	1	0	30	0
75	1	3	1	30	0
130	1	3	1	30	0
118	1	3	1	20	0
12	1	3	1	20	0
101	1	1	1	20	20
65	1	3	0	10	2
116	1	2	0	10	10
70	1	3	0	10	0
31	1	3	1	5	10
69	1	1	0	0	30
42	1	3	0	0	0

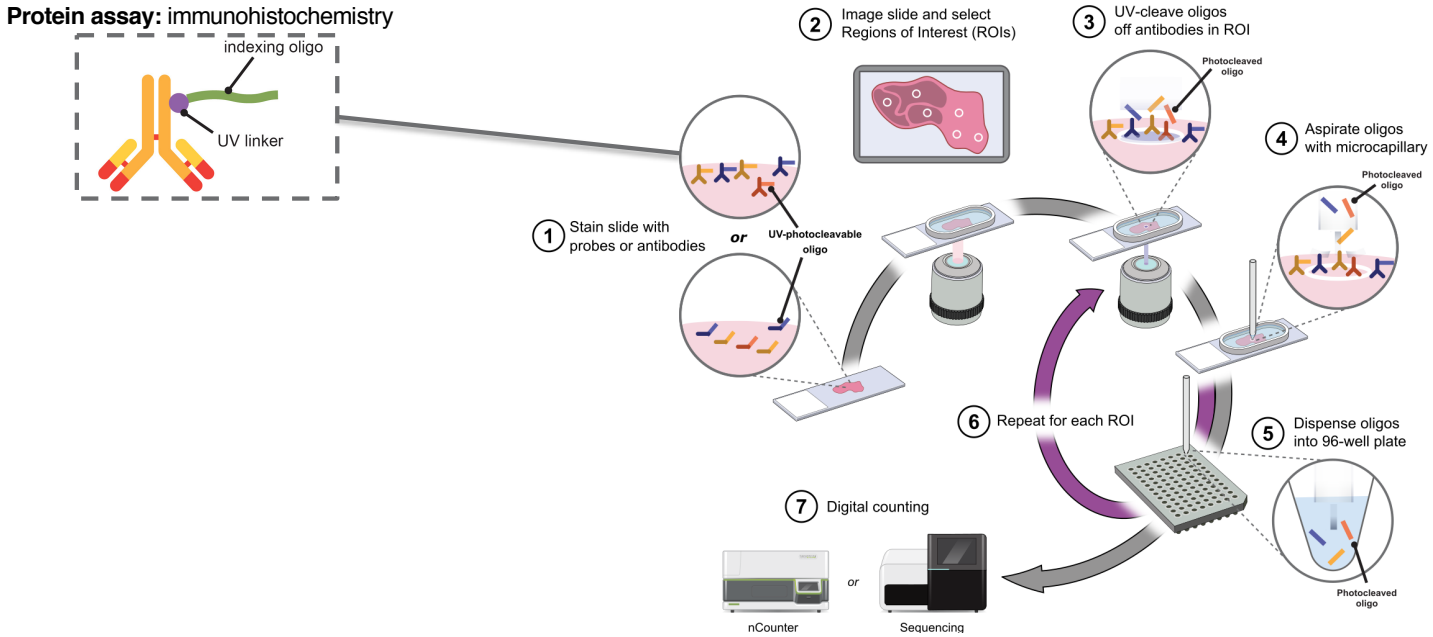
green shading: used in analysis

b

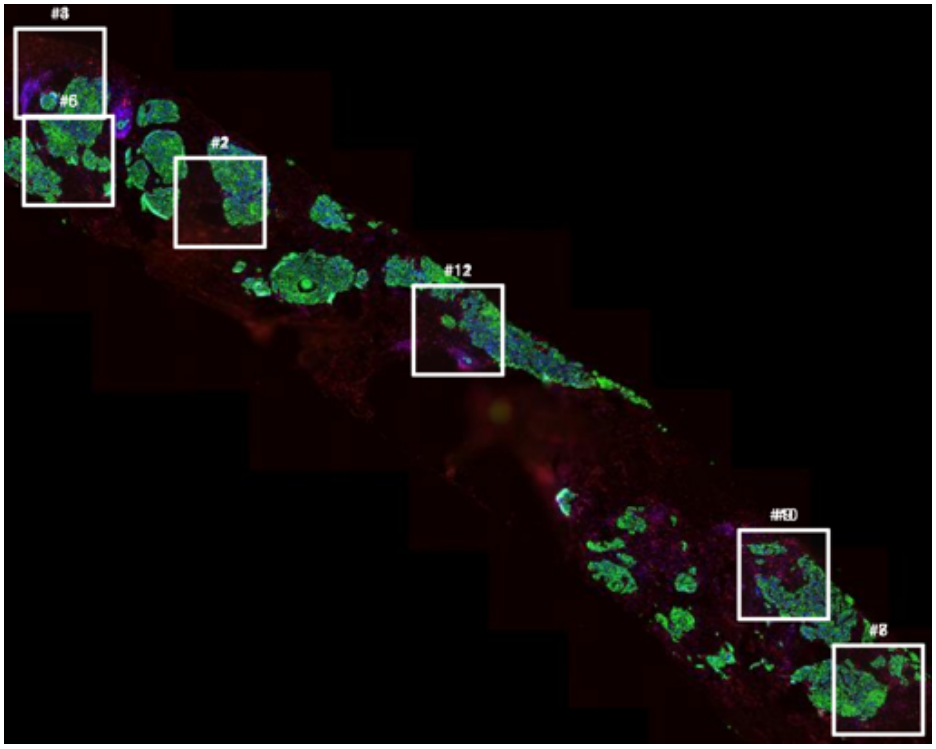


Color Key: panCK CD45 dsDNA

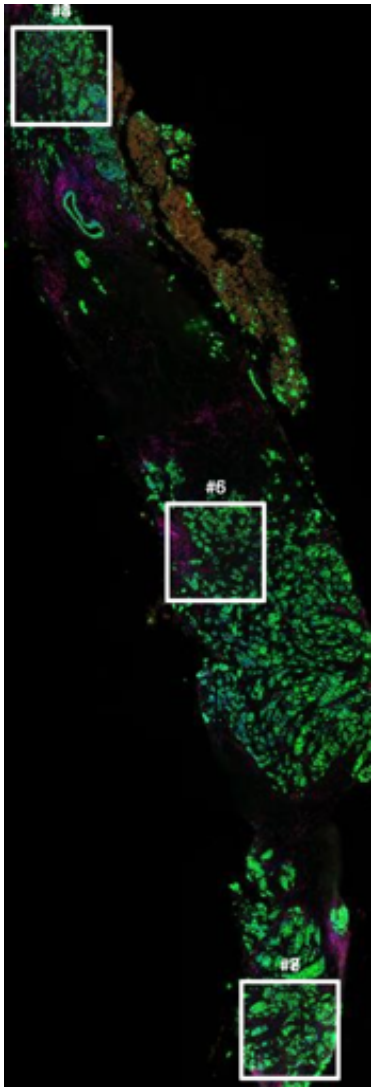
Supplementary Figure 2. a. Pathology-estimated cellularity pre-treatment and on-treatment for the discovery cohort. Samples with green shading indicate those used for subsequent analysis. For the pathologic complete response (pCR) column, 0=non-pCR, 1=pCR. For the estrogen receptor (ER) status column, 0=ER-negative, 1=ER-positive. b. An example region from case 30 sampled on-treatment. While cellularity was estimated to be 0 based on pathology review of a distinct tissue section, tumor regions were identified upon imaging the tissue section used in this analysis.



Supplementary Figure 3. NanoString Digital Spatial Profiler workflow summary. The slide is stained with the mix of protein antibodies. The antibodies have an indexing oligo attached, which is used for subsequent readout. ROIs (regions of interest) are selected and illuminated using UV (ultraviolet) light. The UV light causes the indexing oligos within the ROI to be cleaved off for collection and per-probe quantification.

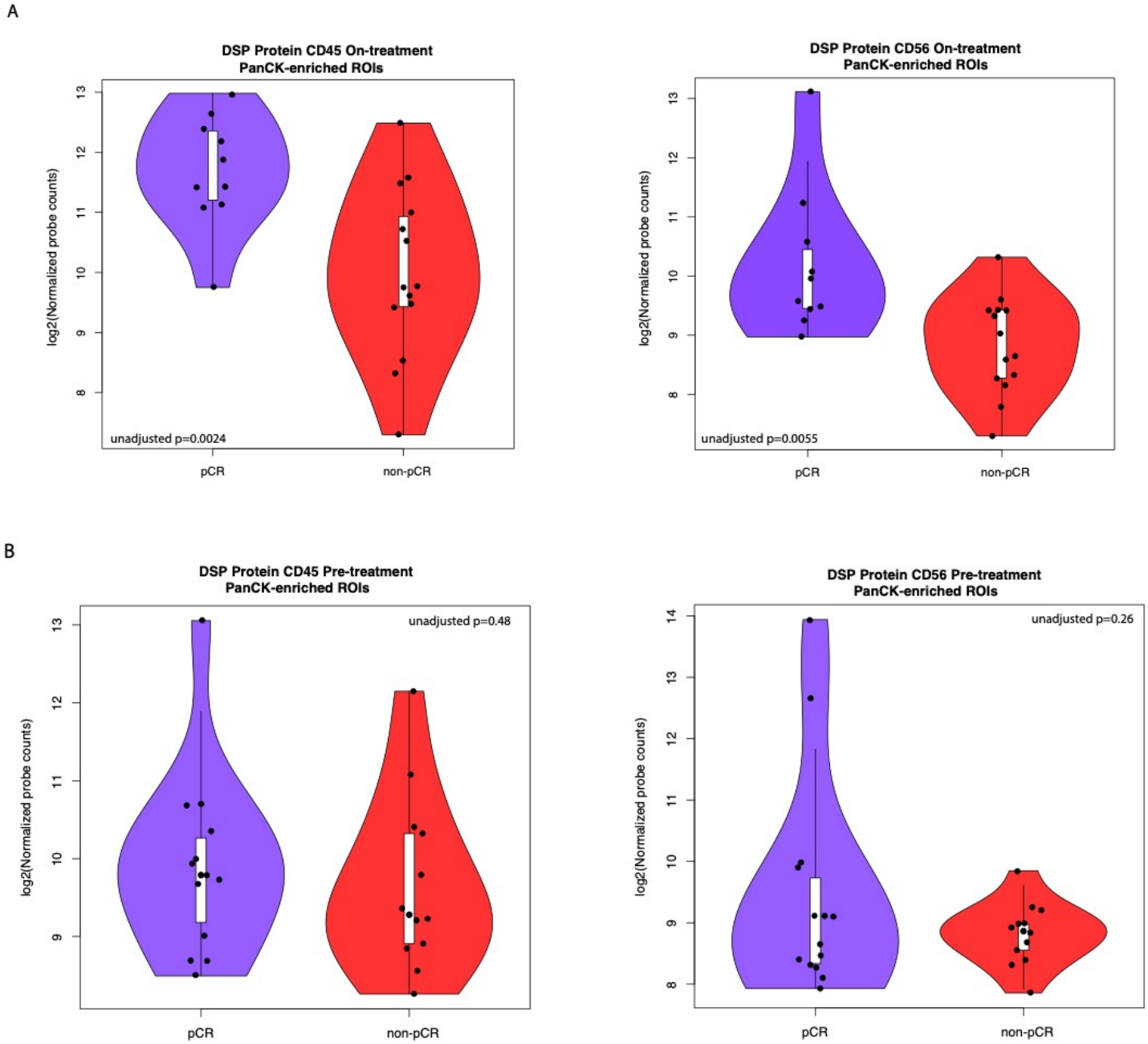


Case 69
pCR

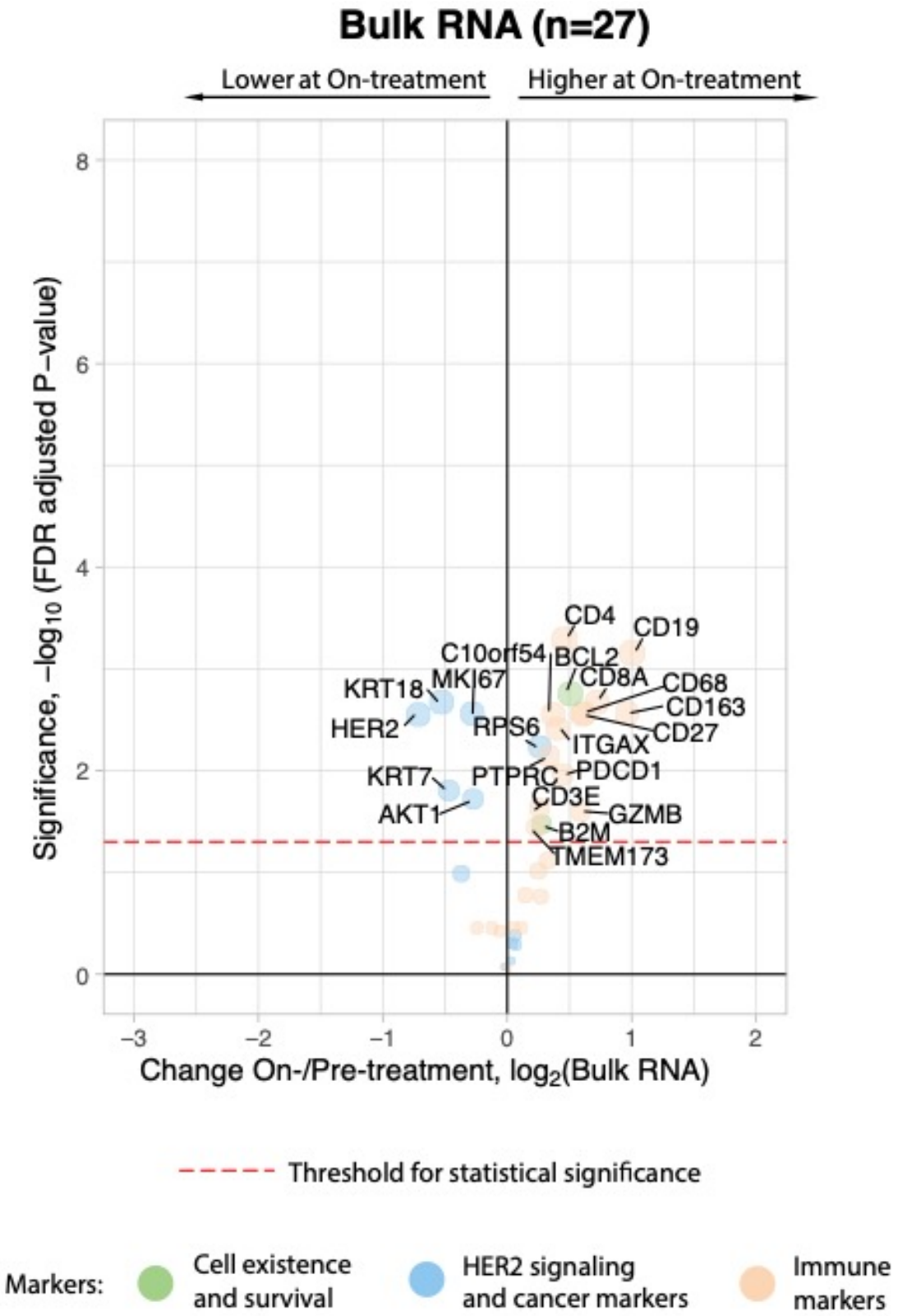


Case 58
non-pCR

Supplementary Figure 4. The location of spatially separated ROIs within tissue specimens for a representative pCR case (69) and a demonstrative non-pCR case (58). An average of 4 ROIs were profiled per tissue (range: 1-7).



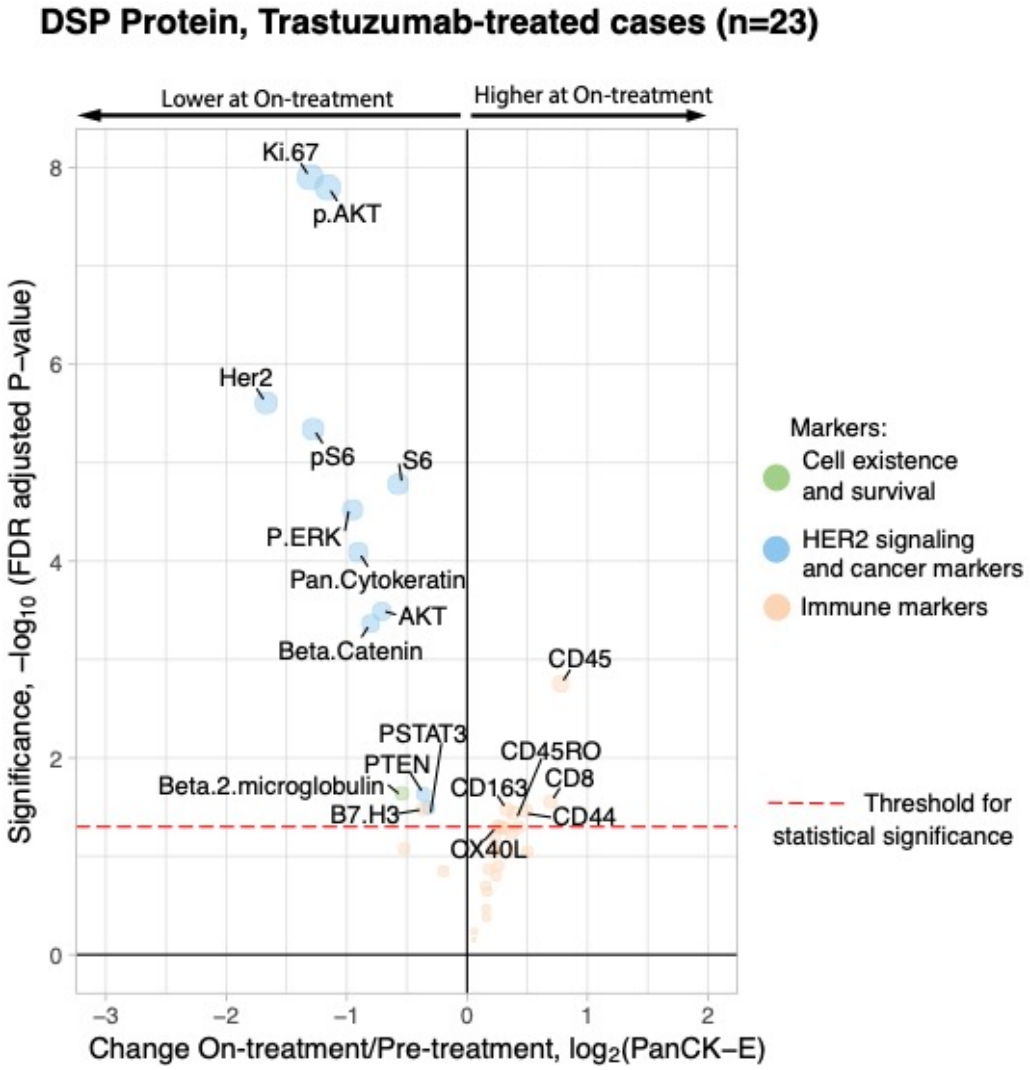
Supplementary Figure 5. a. CD45 values and CD56 values from the Digital Spatial Profiling (DSP) protein data on-treatment in the pCR cases versus the non-pCR cases. Each point represents the average probe values for all panCK-enriched ROIs for that case On-treatment. The p-value was derived using a linear mixed-effect model over the multi-region data with blocking by patient. b. CD45 values and CD56 values from the protein DSP data pre-treatment in the pCR cases versus the non-pCR cases. For each violin plot, the white box represents the interquartile range and the black lines extending from the white box represent 1.5X the interquartile range. Analyses based on the discovery cohort.



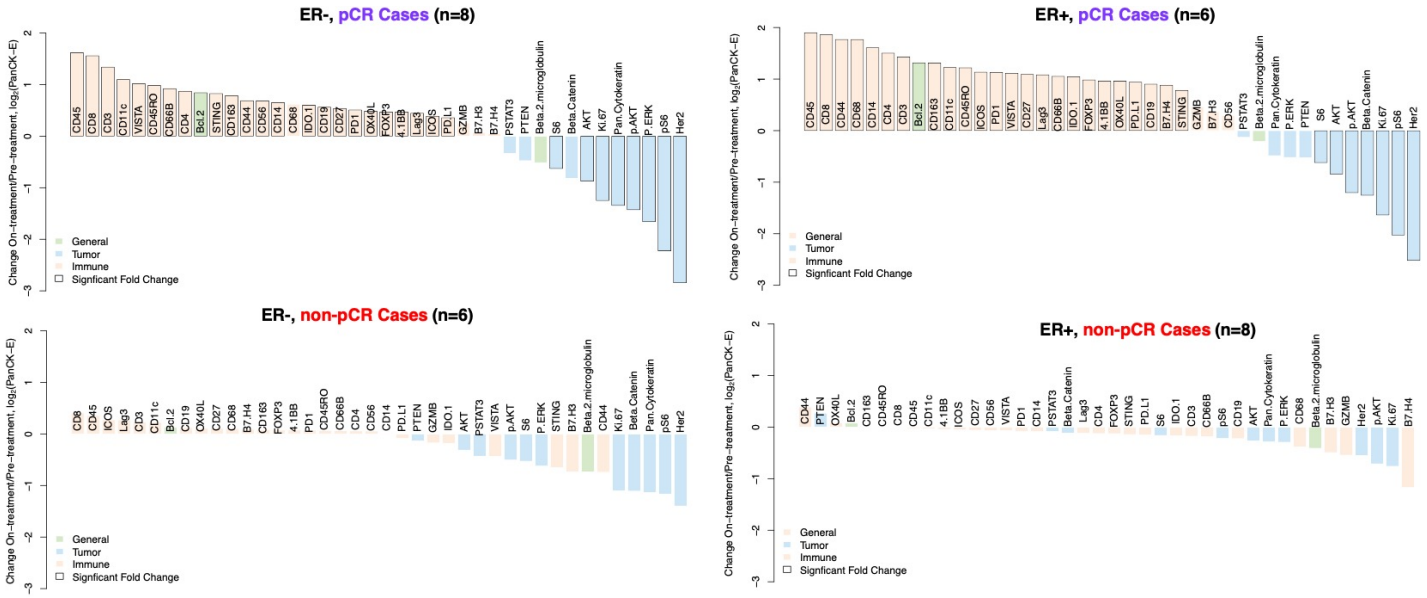
Supplementary Figure 6. Volcano plot demonstrating treatment-associated changes based on comparison of pre-treatment versus on-treatment bulk RNA expression levels. RNA transcripts with corresponding Digital Spatial Profiling (DSP) protein markers were used in this analysis. Significance, -log10(FDR adjusted p-value), is indicated along the y-axis. Analyses based on the discovery cohort.

DSP antibody name	Gene name
Beta.Catenin	CTNNB1
S6*/pS6	RPS6
PTEN	PTEN
P.ERK	MAPK1
Ki.67	MKI67
Beta.2.microglobulin	B2M
AKT*/pAKT	AKT1
PSTAT3	STAT3
Her2	ERBB2
Pan.Cytokeratin	KRT18*, KRT7, KRT19
CD8	CD8A
CD4	CD4
CD68	CD68
CD14	CD14
GZMB	GZMB
CD3	CD3E
CD66B	CEACAM8
VISTA	C10orf54
PD1	PDCD1
PD.L1	CD274
IDO.1	INDO
CD44	CD44
CD56	NCAM1
CD45*/CD45RO	PTPRC
CD19	CD19
STING	TMEM173
Lag3	LAG3
CD11c	ITGAX
ICOS	ICOS
CD27	CD27
CD163	CD163
OX40L	TNFSF4
Bcl.2	BCL2
B7.H3	CD276
B7.H4	VTCN1
4.1BB	TNFRSF9
FOXP3	FOXP3

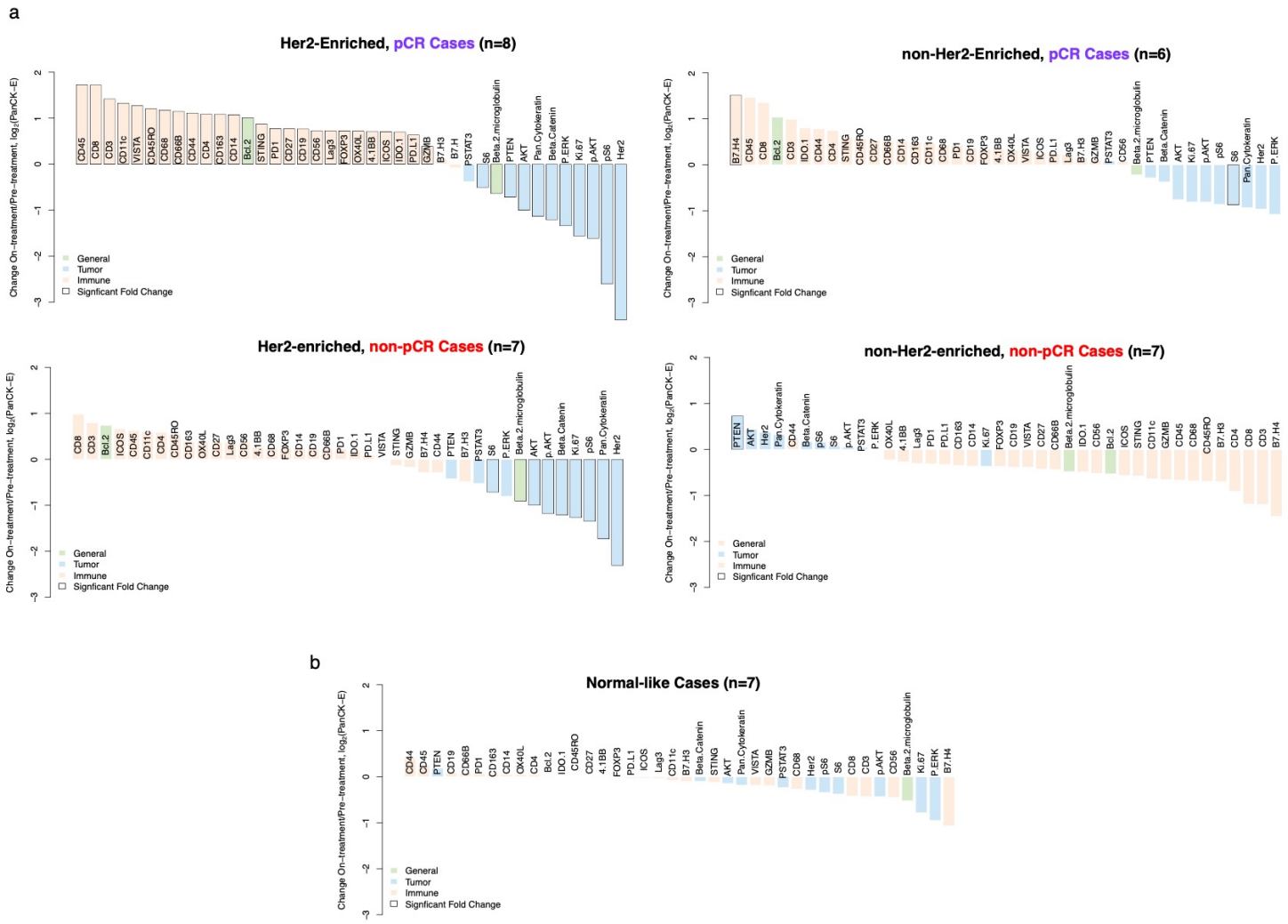
Supplementary Figure 7. Pairing of protein antibodies and gene names used in comparative analyses between DSP and bulk expression data. Genes listed here were used to generate the bulk expression volcano plot shown in Supplementary Figure 6. For direct comparisons yielding the correlation plots shown in Supplementary Figure 21, the marker names indicated by the asterisk were used; specifically, for proteins with multiple form (e.g. AKT/pAKT, CD45/CD45RO), the unmodified form of the protein was used when available and the breast-cancer associated keratin gene with the highest mean expression level was used.



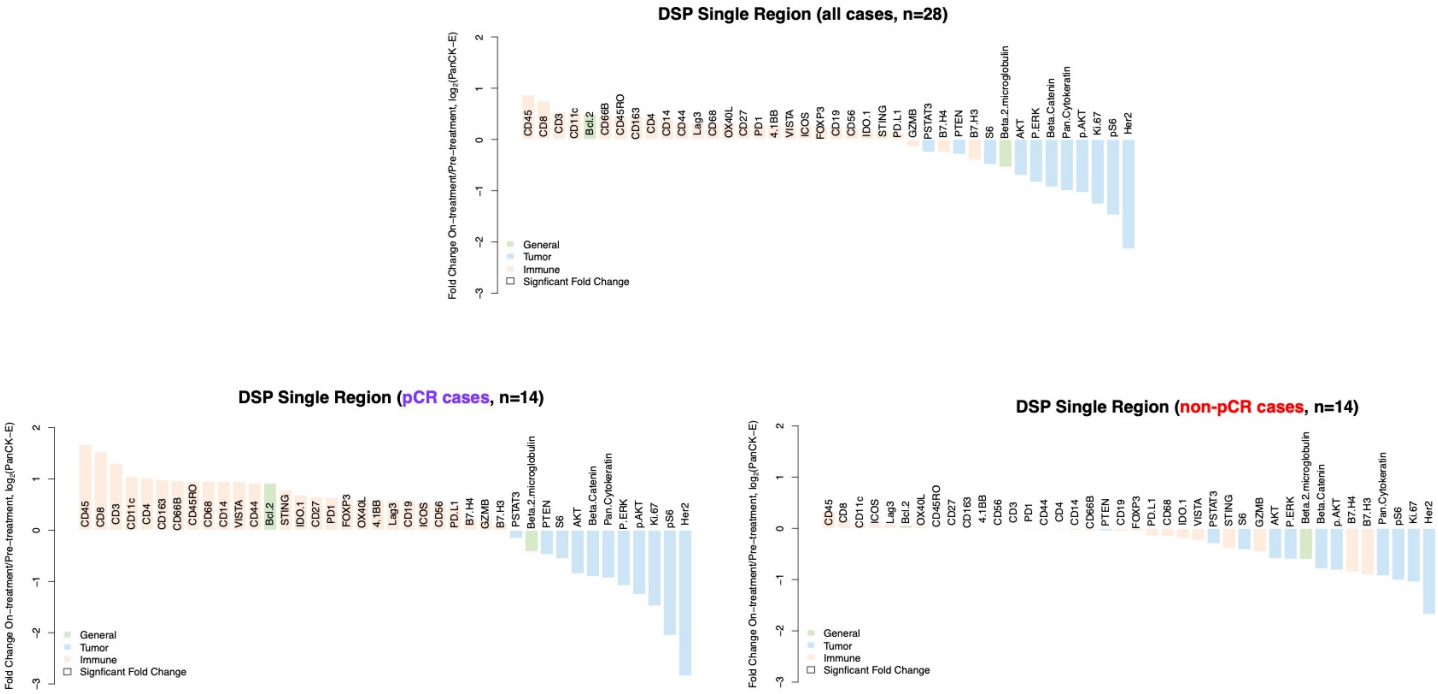
Supplementary Figure 8. Volcano plot demonstrating treatment-associated changes based on comparison of pre-treatment versus on-treatment protein marker expression levels in pancytokeratin-enriched (PanCK-E) regions in the trastuzumab-treated cases (arms 1 and 3, n=23). Significance, -log10(FDR adjusted p-value), is indicated along the y-axis. Analyses based on the discovery cohort.



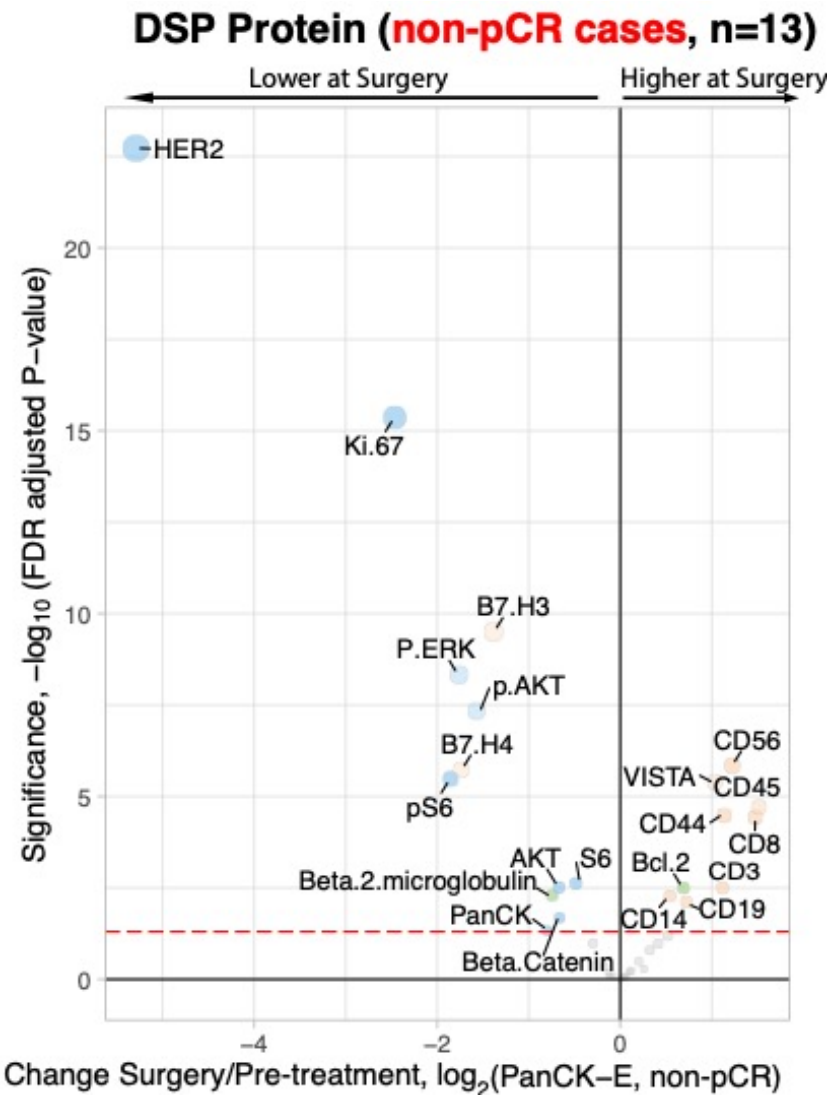
Supplementary Figure 9. Waterfall plots illustrating treatment-associated changes (pre-treatment to on-treatment) based on in pancytokeratin-enriched (PanCK-E) regions from DSP protein expression data. Input data was stratified both by estrogen receptor (ER) status and pathologic complete response (pCR) outcome. Analyses based on the discovery cohort.



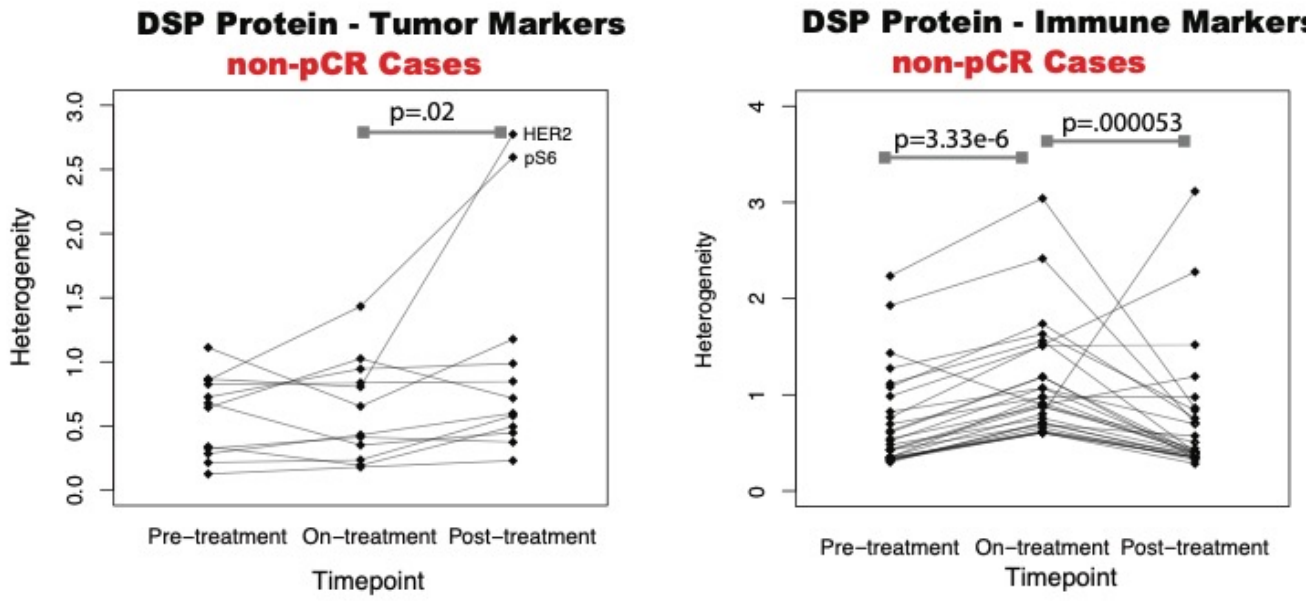
Supplementary Figure 10. a. Waterfall plots illustrating treatment-associated changes (pre-treatment to on-treatment) based on in pancytokeratin-enriched (PanCK-E) regions from the DSP protein expression data. Samples were stratified both by PAM50 status (Her2-Enriched or other) and pathologic complete response (pCR) outcome. b. Waterfall plots illustrating treatment-associated changes (pre-treatment to on-treatment) in the Normal-like cases in the discovery cohort. PAM50 status was determined using bulk expression data from the pre-treatment biopsies. Analyses based on the discovery cohort.



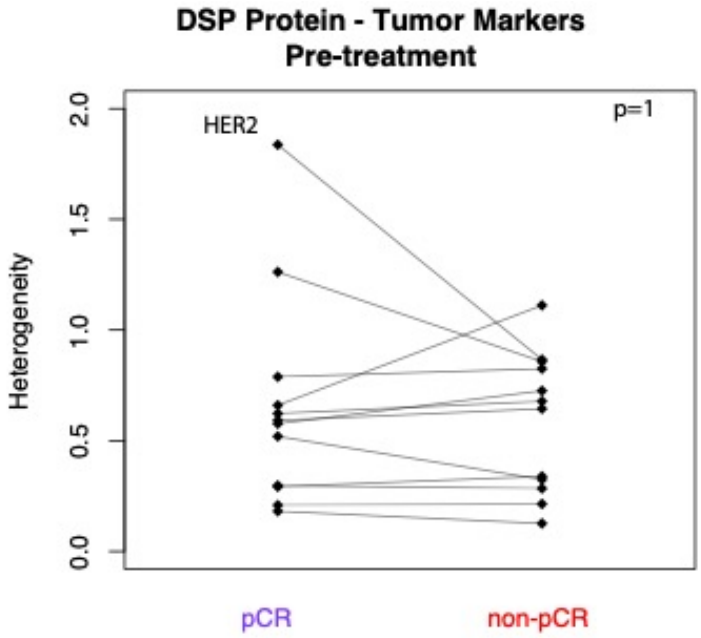
Supplementary Figure 11. Waterfall plots, generated using pancytokeratin-enriched (PanCK-E) regions from DSP protein expression data, illustrating treatment-associated changes (pre-treatment to on-treatment) when only one region is used to profile each sample (averaged across 100 iterations of random samples of a single region per timepoint), rather than the 2-7 regions from each sample used in other analyses. The upper plot is for all patients, and the lower plots are stratified by pathologic complete response (pCR) status. Analyses based on the discovery cohort.

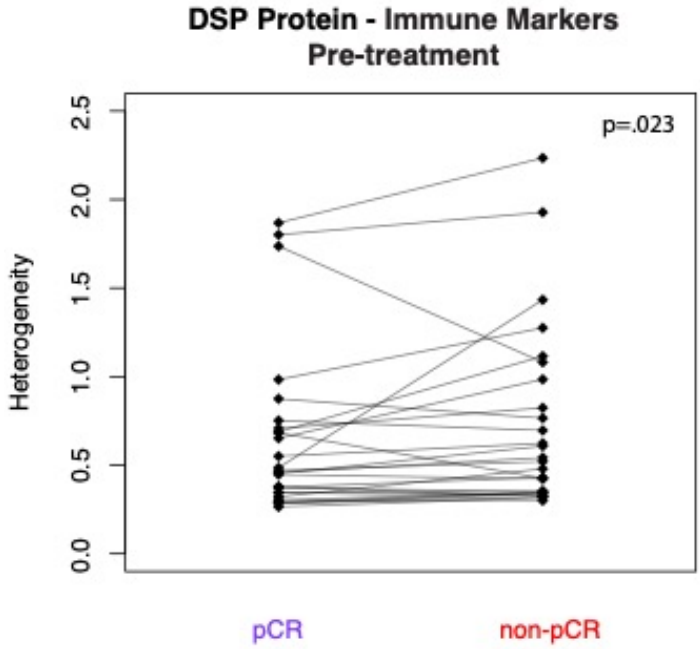


Supplementary Figure 12. Volcano plot demonstrating treatment-associated changes from pre-treatment to surgery in tumors that did not undergo pathologic complete response (pCR) using DSP protein expression levels in pancytokeratin-enriched (PanCK-E) regions. Significance, -log10(FDR adjusted p-value), is indicated along the y-axis. Analyses based on the discovery cohort.

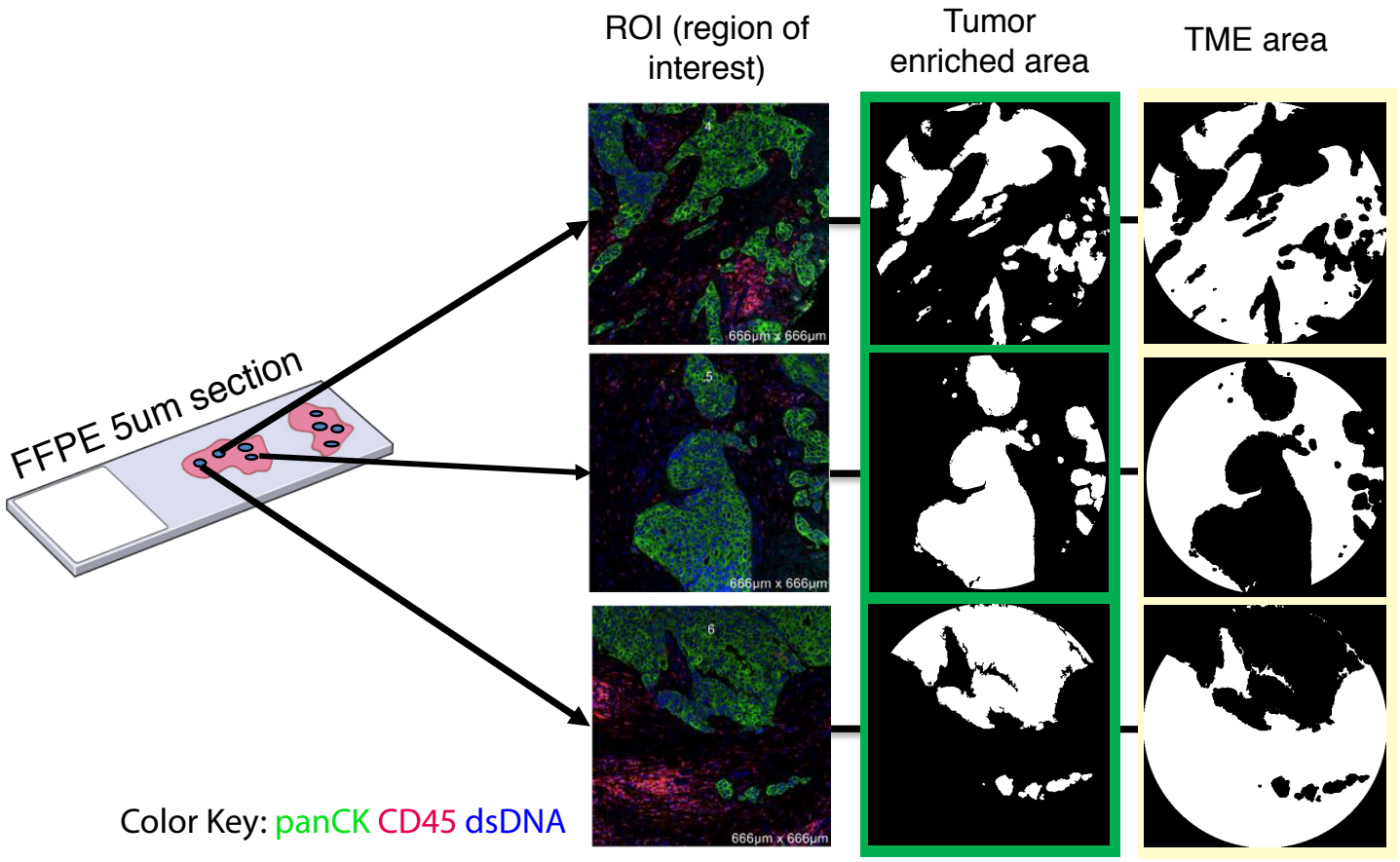


Supplementary Figure 13. Pre-, on-, and post-treatment heterogeneity for each DSP protein marker in non-pCR cases (patients with tumor cells present at surgery). Analyses based on the discovery cohort.

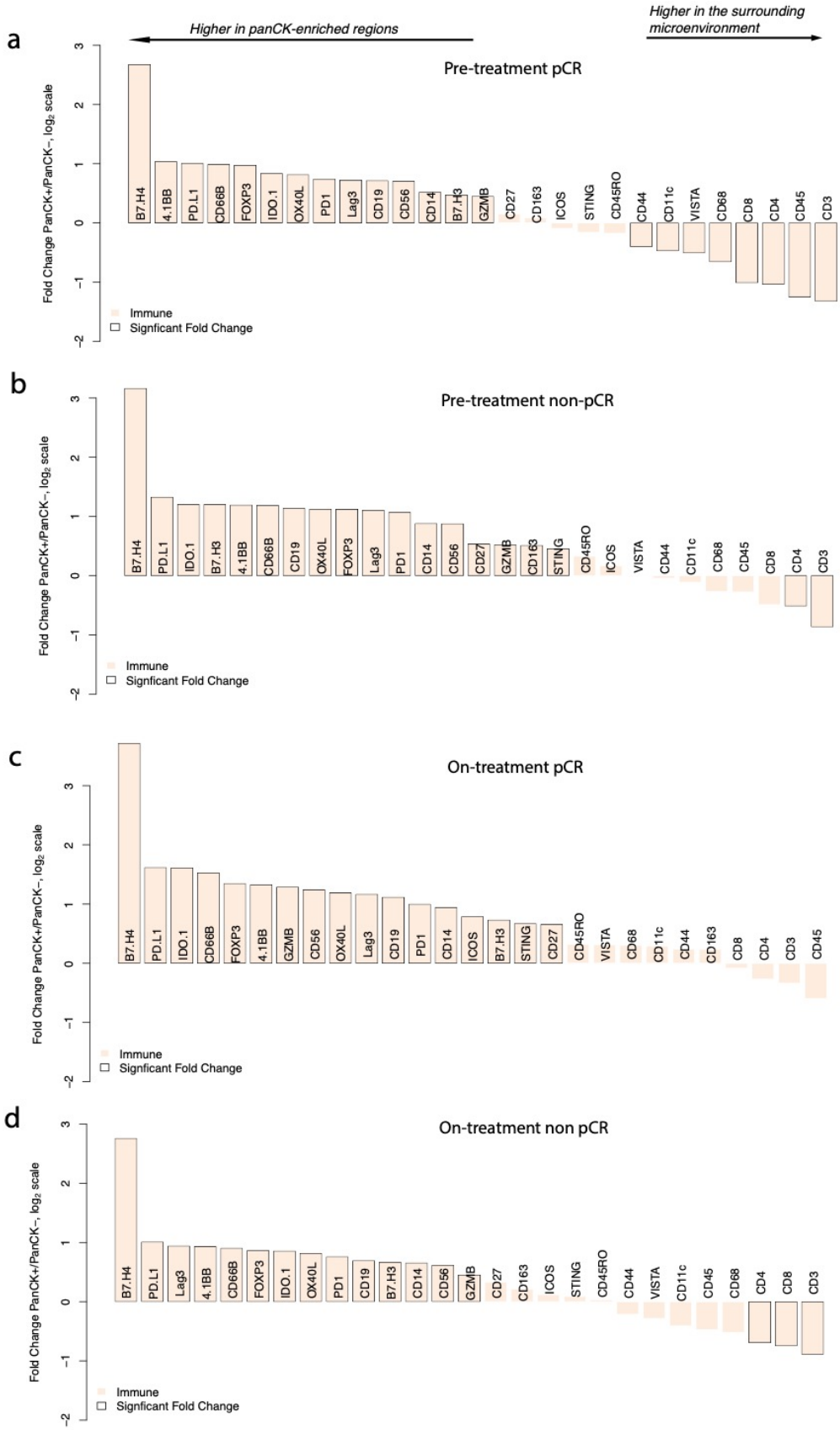




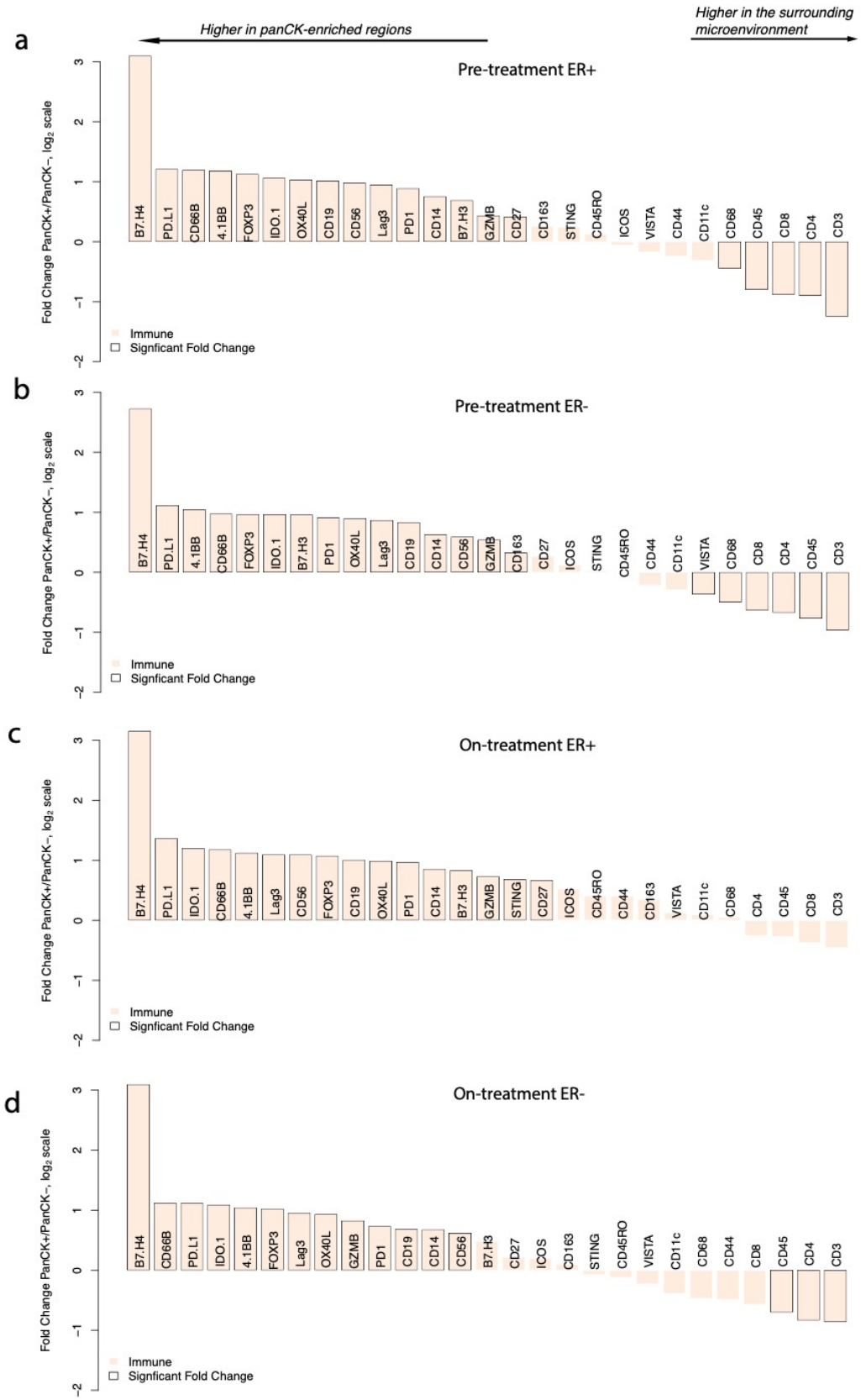
Supplementary Figure 14. Pre-treatment heterogeneity in DSP protein marker expression in pCR and non-pCR cases. Heterogeneity was calculated as the mean squared error within patients based on analysis of variance. P-values are based on a two-sided Wilcoxon matched-pair signed rank test. Analyses based on the discovery cohort.



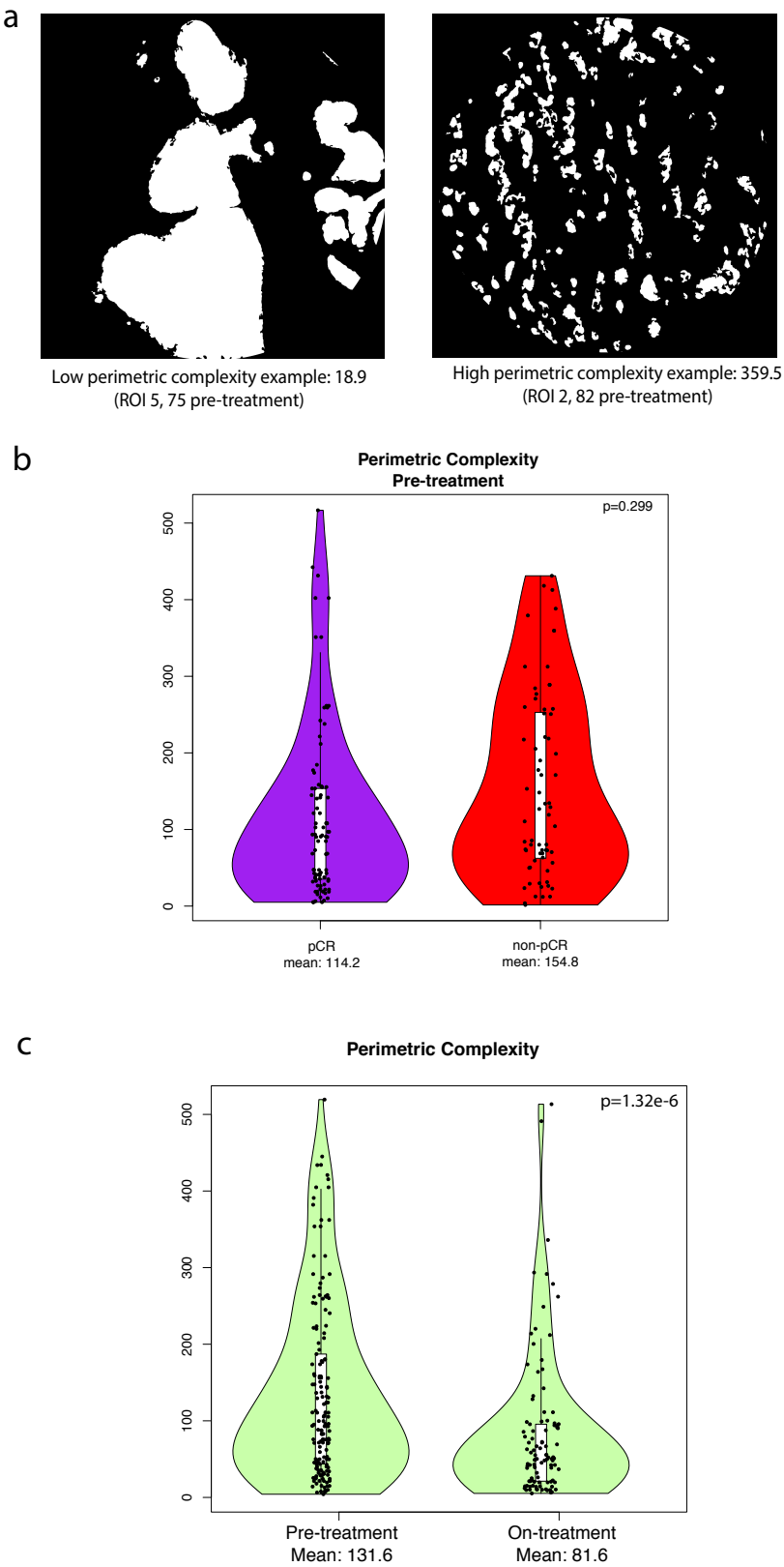
Supplementary Figure 15. Digital Spatial Profiling (DSP) was performed on multiple regions of interest (ROIs) per tissue sample. Protein counts were measured within phenotypic regions corresponding to the PanCK-enriched (tumor-enriched) masks that include tumor cells and co-localized immune cells and separately for the inverted mask corresponding to panCK-negative (tumor microenvironment, TME) regions.



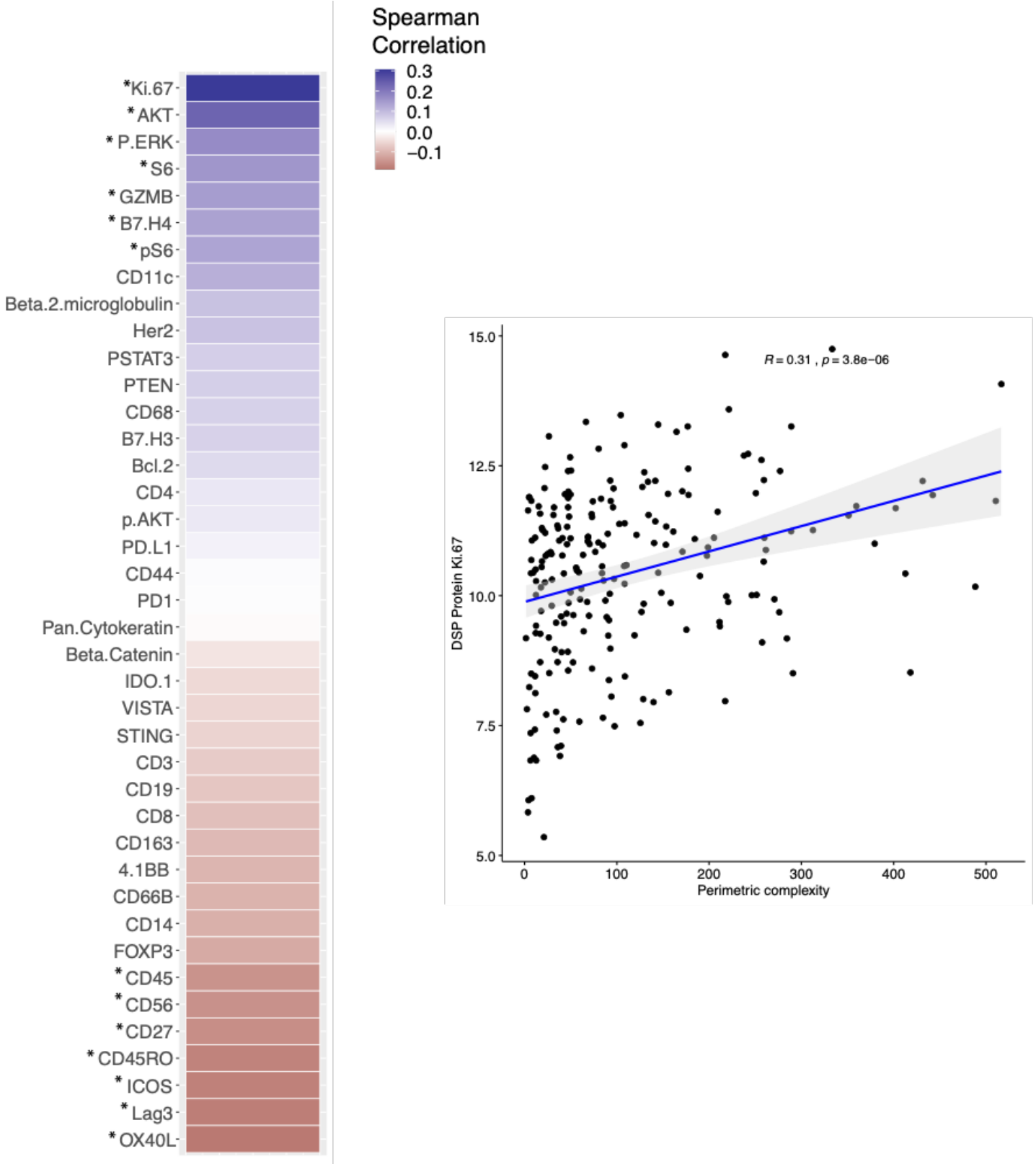
Supplementary Figure 16. Waterfall plots, generated using the DSP protein data, comparing immune marker expression between the panCK-enriched regions and the surrounding panCK-negative regions pre-treatment and on-treatment, in pCR (n=14) and non-pCR cases (n=14). Pre-treatment, the correlation between immune marker fold-change values in the pCR and non-pCR cases was 0.98 indicating similar immune distribution across the panCK-enriched regions and surrounding microenvironment regardless of pCR outcome and this correlation remained high on-treatment (0.95). Analyses based on the discovery cohort.



Supplementary Figure 17. Waterfall plots generated using DSP protein data, comparing immune marker expression between the panCK-enriched regions and the surrounding panCK-negative regions pre-treatment and on-treatment, in ER-positive cases (n=14), and ER-negative cases (n=14). Analyses based on the discovery cohort.



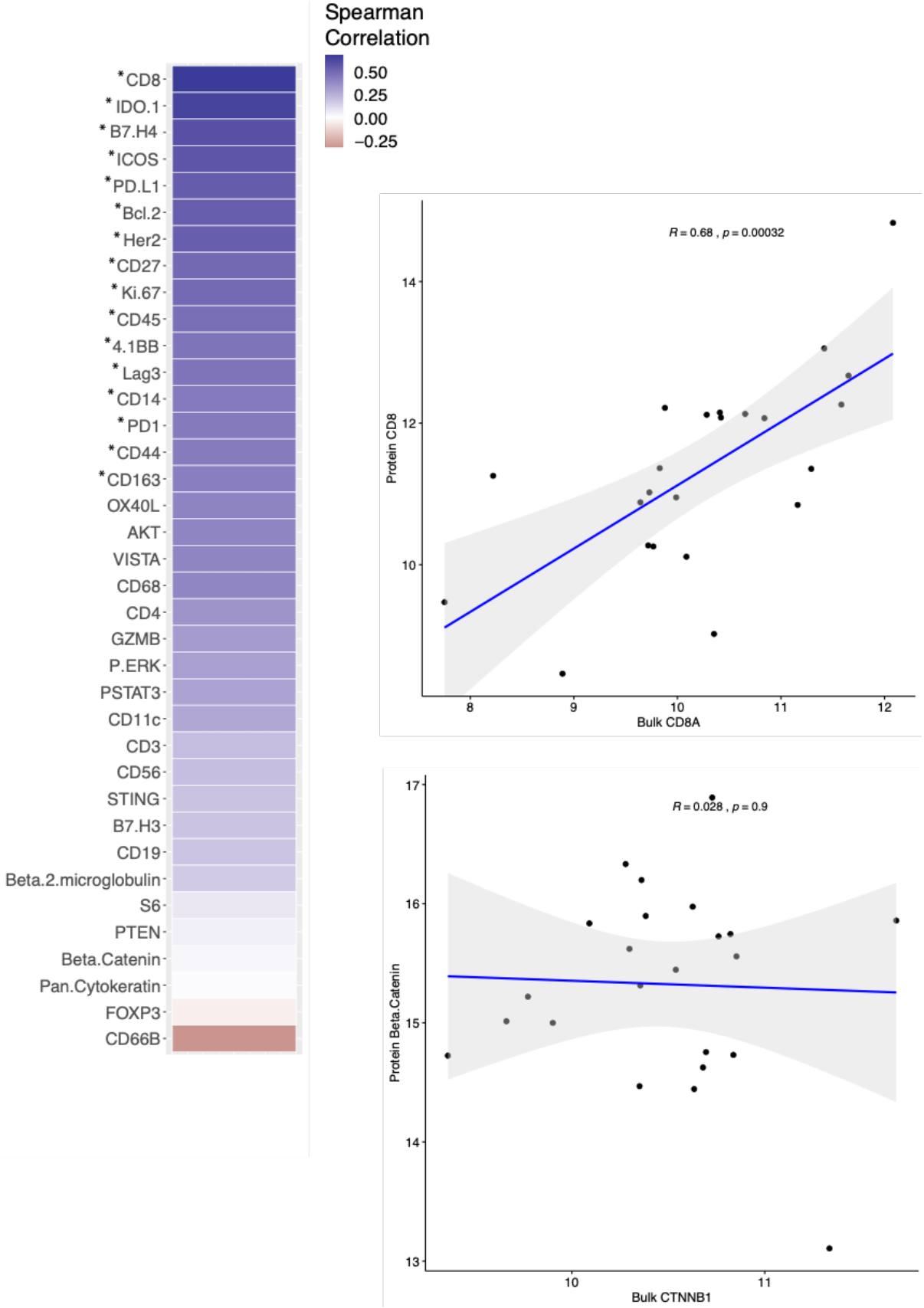
Supplementary Figure 18. a. Illustration of panCK-enriched binary masks and perimetric complexity-based quantification of the tumor-microenvironment border. b. Comparison of perimetric complexity values pre-treatment between pCR cases and non-pCR cases. c. Comparison of pre-treatment versus on-treatment perimetric complexity values. PanCK-enriched ROIs were used to quantify perimetric complexity. P-values computed with a linear model, blocked by patient. Analyses are based on the discovery cohort.



Supplementary Figure 19. Spearman correlation between the DSP protein expression values and perimetric complexity per region of interest (ROI) in the pre-treatment and on-treatment tissue specimens from the discovery cohort. Significantly correlated probes: p-value < .05. Correlation plot for Ki-67, the marker with the highest correlation with perimetric complexity, where each dot represents an individual ROI.

Model comparisons by marker type						
Data type	**Timepoints**	**Marker Set 1**	**Mean AUC 1**	**Marker Set 2**	**Mean AUC 2**	**Adjusted P-value**
DSP protein	On-treatment + Pre-treatment	Tumor marker means, Immune marker means	0.565	Tumor marker SEM, Immune marker means	0.679	<0.0001
DSP protein	On-treatment + Pre-treatment	Tumor marker means, Immune marker means	0.565	Tumor marker SEM, Immune marker SEM	0.589	0.239
DSP protein	On-treatment + Pre-treatment	Tumor marker SEM, Immune marker means	0.679	Tumor marker SEM, Immune marker SEM	0.589	<0.0001

Supplementary Figure 20. AUROC (Area Under the Receiver Operating Characteristics) performance (using nested cross-validation with Holm-Bonferroni correction for multiple hypotheses) comparing DSP protein on- plus pre-treatment L2-regularized classifiers trained using marker means versus marker standard error of the mean (SEM) for tumor markers and immune markers. Model comparisons were performed in the discovery cohort.



Supplementary Figure 21. Spearman correlation between DSP protein probes (averaged across all ROIs per case) and bulk RNA transcripts corresponding to these markers pre-treatment (see **Supplementary Figure 7**). Significantly correlated probes (with p-value < .05) are indicated by an asterisk. Two exemplary correlation plots are shown, where each dot represents a single case. Analyses based on the discovery cohort.

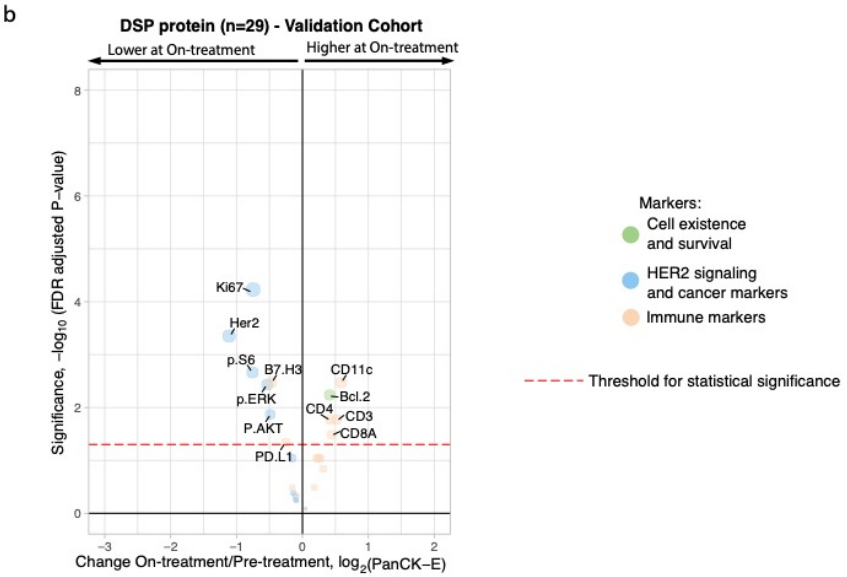
a

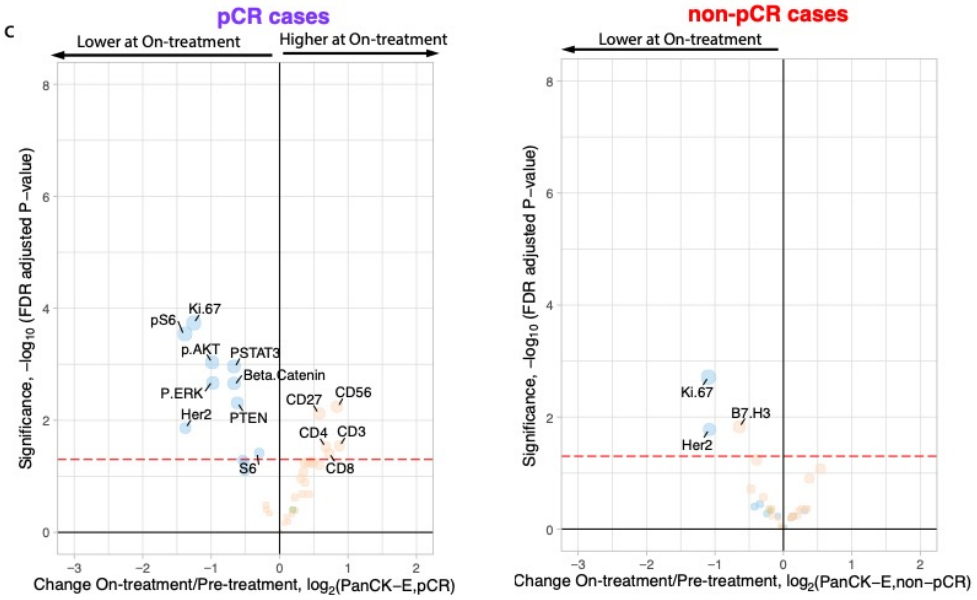
Validation Cohort Characteristics	n=29 (100%)
Treatment Arm	
Arm 1 (Trastuzumab)	11 (38%)
Arm 2 (Lapatinib)	7 (24%)
Arm 3 (Trastuzumab + Lapatinib)	11 (38%)
pCR Status	
pCR	10 (34%)
non-pCR	19 (66%)
ER Status	
ER+	13 (45%)
ER-	16 (55%)
PAM50 Status: Pre-treatment	
HER2-Enriched	15 (52%)
Normal-like	5 (17%)
Basal	2 (6.9%)
LuminalA	2 (6.9%)
LuminalB	2 (6.9%)
No Data	3 (10%)

	pCR	non-pCR
Arm 1	6 (21%)	5 (17%)
Arm 2	2 (6.9%)	7 (24%)
Arm 3	2 (6.9%)	9 (31%)

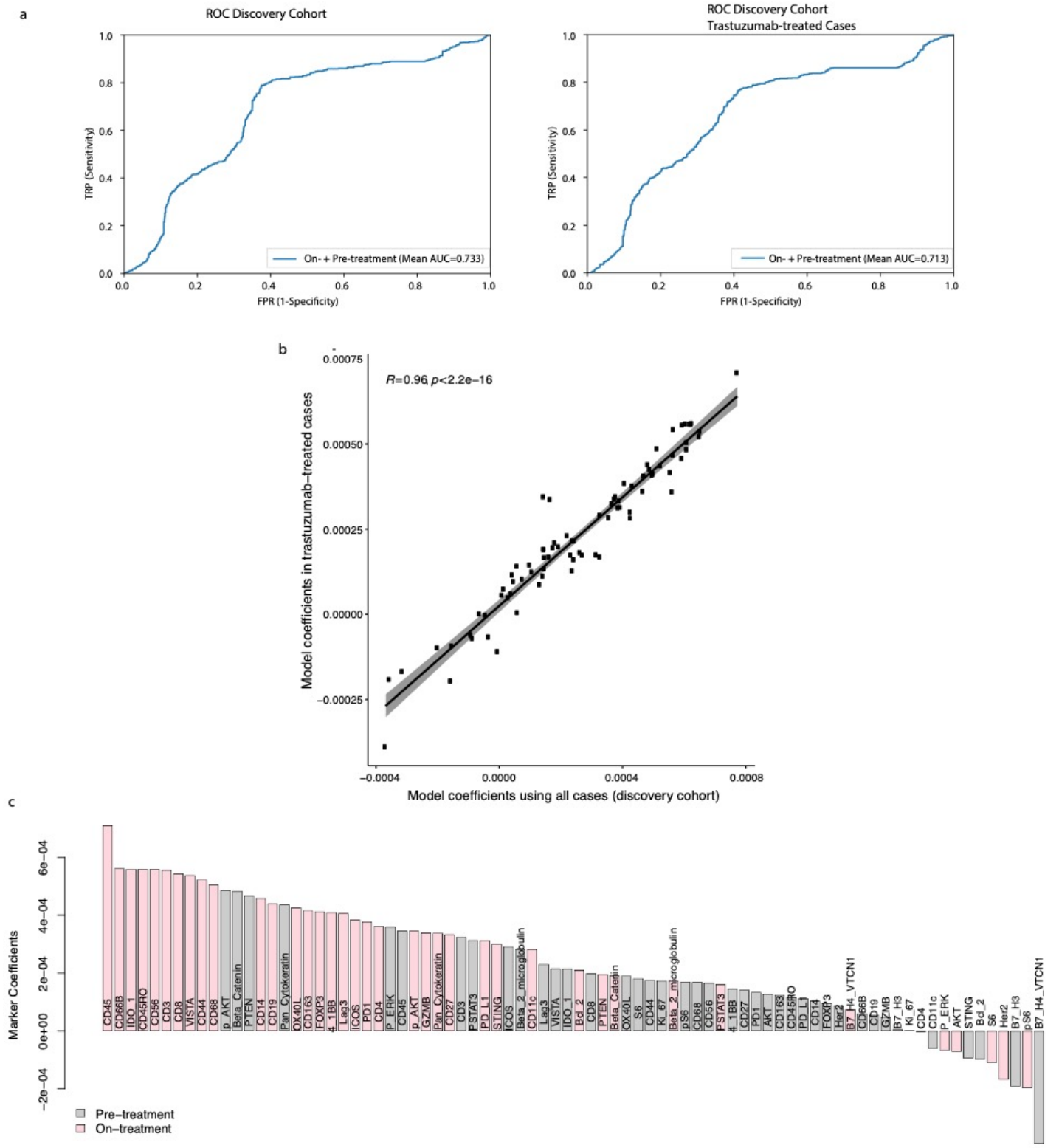
	pCR	non-pCR
ER+	3 (10%)	10(34%)
ER-	7 (24%)	9 (31%)

	ER+	ER-
Arm 1	6 (21%)	6 (17%)
Arm 2	0 (0%)	7 (24%)
Arm 3	7 (24%)	4 (14%)





Supplementary Figure 22. a. Summary of the clinical characteristics for the TRIO-US B07 clinical trial Digital Spatial Profiling (DSP) validation cohort used for model testing. Treatment arm, pathologic complete response (pCR), estrogen receptor (ER) status, and PAM50 status inferred based on pre-treatment bulk expression data are included. Two-way contingency tables compare the distribution of ER status, pCR status, and treatment arm. b. Volcano plot demonstrating treatment-associated changes based on comparison of pre-treatment versus on-treatment protein marker expression levels in pancytokeratin-enriched (PanCK-E) regions in the validation cohort. c. Volcano plots demonstrating treatment-associated changes in pCR versus non-pCR cases in the PanCK-E regions in the validation cohort. Significance, -log10(FDR adjusted p-value), is indicated along the y-axis.



Supplementary Figure 23. a. Receiver operating characteristic (ROC) curves for On- plus Pre-treatment DSP protein L2-regularized classifier in the discovery cohort using all cases with panCK-enriched data from both timepoints (n=23) and in the subset of cases treated with trastuzumab or trastuzumab+lapatinib (n=19). Model performance was assessed via cross-validation using the 40 DSP protein markers profiled in both cohorts b. Correlation plot comparing the marker coefficients for the On- plus Pre-treatment DSP protein trained using all cases in the discovery cohort and using only those cases treated with trastuzumab (arms 1 and 3). c. Coefficients for each marker in the L2-regularized On- plus Pre-treatment DSP protein model, trained using only those cases treated with trastuzumab (arms 1 and 3).

Protein marker panel

Markers of immune response	Her2 pathway members and other cancer markers
4-1BB^*	AKT
B7-H3	Beta-catenin
B7-H4	Her2
CD3	Ki67
CD4	PanCK
CD8	p-AKT
CD11c	p-ERK
CD14*	p-STAT3*
CD19*	p-S6
CD27*	PTEN
CD44	S6
CD45*	**Markers of cell existence and survival**
CD45RO	
CD56*	Beta-2 microglobulin
CD66B	Bcl-2
CD68	
CD163*	
FOXP3^*	
Granzyme B	
ICOS*	
IDO-1	
Lag3*	
PD1*	
PD-L1	
OX40L*	
STING	
VISTA	

Supplementary Figure 24. Markers with a signal to noise ratio < 3 (Methods) in the discovery cohort indicated by a caret (^) and those with an SNR < 3 in the validation cohort are indicated with an asterisk (*).