NSABP B-41, a Randomized Neoadjuvant Trial: Genes and Signatures Associated with Pathologic Complete Response

Sandra M. Swain,[1,2] Gong Tang,[1,3] Heather Ann Brauer,[4] David S. Goerlitz,[2] Peter C. Lucas,[1,3] André Robidoux,[1,5] Brent T. Harris,[2] Hanna Bandos,[1,3] Yuqi Ren,[4] Charles E. Geyer Jr.,[1,6] Priya Rastogi,[1,3,7] Eleftherios P. Mamounas,[1,8] Norman Wolmark[1,3]

[1]National Surgical Adjuvant Breast and Bowel Project (NSABP), Pittsburgh, PA; [2]Georgetown Lombardi Comprehensive Cancer Center, Georgetown University Medical Center, Washington, DC; [3]University of Pittsburgh, Pittsburgh, PA; [4]NanoString Technologies, Inc., Seattle, WA; [5]Centre hospitalier de l'Universite´ de Montréal, Montréal, QC; [6]Houston Methodist Cancer Center, Houston, TX; [7]Magee-Womens Hospital, Pittsburgh, PA; [8]Orlando Health UF Health Cancer Center, Orlando, FL.

Supplementary Materials.

Table S1. REMARK checklist [a]

Item to be reported	Page no.
Introduction	
1. State the marker examined, the study objectives, and any pre-specified hypotheses.	6
Materials and Methods	
Patients	
2. Describe the characteristics (e.g., disease stage or co-morbidities) of the study patients, including their source and inclusion and exclusion criteria.	6, 7
3. Describe treatments received and how chosen (e.g., randomized or rule-based).	6, 7
Specimen characteristics	
4. Describe type of biological material used (including control samples) and methods of preservation and storage.	7
Assay methods	
5. Specify the assay method used and provide (or reference) a detailed protocol, including specific reagents or kits used, quality control procedures, reproducibility assessments, quantitation methods, and scoring and reporting protocols. Specify whether and how assays were performed blinded to the study endpoint.	7, 8
Study design	
6. State the method of case selection, including whether prospective or retrospective and whether stratification or matching (e.g., by stage of disease or age) was used. Specify the time period from which cases were taken, the end of the follow-up period, and the median follow-up time.	6, 11
7. Precisely define all clinical endpoints examined.	7
8. List all candidate variables initially examined or considered for inclusion in models.	9, 10
9. Give rationale for sample size; if the study was designed to detect a specified effect size, give the target power and effect size.	10, 11
Statistical analysis methods	
10. Specify all statistical methods, including details of any variable selection procedures and other model-building issues, how model assumptions were verified, and how missing data were handled.	9, 10
11. Clarify how marker values were handled in the analyses; if relevant, describe methods used for cutpoint determination.	8, 9, 10
Results	
Data	
12. Describe the flow of patients through the study, including the number of patients included in each stage of the analysis (a diagram may be helpful) and reasons for dropout. Specifically, both overall and for each subgroup extensively examined report the numbers of patients and the number of events.	10, 11

Item to be reported	Page no.
13. Report distributions of basic demographic characteristics (at least age and sex), standard (disease-specific) prognostic variables, and tumor marker, including numbers of missing values.	11, 4 (SM)
Analysis and presentation	
14. Show the relation of the marker to standard prognostic variables.	14
15. Present univariable analyses showing the relation between the marker and outcome, with the estimated effect (e.g., hazard ratio and survival probability). Preferably provide similar analyses for all other variables being analyzed. For the effect of a tumor marker on a time-to-event outcome, a Kaplan-Meier plot is recommended.	11,12, 26, 5 (SM), 8 (SM)
16. For key multivariable analyses, report estimated effects (e.g., hazard ratio) with confidence intervals for the marker and, at least for the final model, all other variables in the model.	12, 13, 28
17. Among reported results, provide estimated effects with confidence intervals from an analysis in which the marker and standard prognostic variables are included, regardless of their statistical significance.	28
18. If done, report results of further investigations, such as checking assumptions, sensitivity analyses, and internal validation.	12, 7 (SM)
Discussion	
19. Interpret the results in the context of the pre-specified hypotheses and other relevant studies; include a discussion of limitations of the study.	15, 16, 18, 19
20. Discuss implications for future research and clinical value.	17, 18, 19

[a] Source: McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Reporting recommendations for tumor marker prognostic studies (REMARK). *J Natl Cancer Inst* **2005**; 97(16): 1180-4.

SM, abbreviation for Supplementary Materials.

Table S2. Baseline characteristics of patients included and excluded in the current analysis

Characteristic, *n* (%)	**Included *n* = 194**	**Excluded *n* = 335**	***P*** [a]
Treatment			
Chemotherapy + tamoxifen	69 (36)	112 (33)	0.84
Chemotherapy + lapatinib	64 (33)	110 (33)	
Chemotherapy + tamoxifen + lapatinib	61 (31)	113 (34)	
Clinical nodal status			
N0	89 (46)	168 (50)	0.39
N1+	105 (54)	167 (50)	
Age			
< 50 years	98 (51)	179 (53)	0.58
≥ 50 years	96 (49)	156 (47)	
Clinical tumor size			
2.0-4.0 cm	98 (51)	173 (52)	0.87
4.1+ cm	96 (49)	162 (48)	
Hormone receptor status			
Negative	82 (42)	116 (35)	0.10
Positive	112 (58)	219 (65)	

[a] Pearson Chi square test.

Table S3. Differential expression in BC360 curated meta-gene signatures (shown in bold) and single genes (shown in italics) between patients with pCR and patients without pCR among all 194 study participants

	Unadjusted *P*	**Differential expression (pCR vs non-pCR) [95% CI]**
ESR1	0.00005	-1.37 (-2.01, -0.72)
IDO1	0.0001	0.74 (0.38, 1.11)
PGR	0.0003	-1.05 (-1.61, -0.5)
ER	0.0004	-0.54 (-0.83, -0.25)
Cytotoxic	0.0004	0.51 (0.23, 0.79)
ERBB2	0.001	0.68 (0.28, 1.08)
TIS	0.003	0.44 (0.15, 0.73)
p53	0.004	0.38 (0.13, 0.63)
Mast cells	0.004	-0.5 (-0.83, -0.16)
B7_H3	0.004	-0.21 (-0.35, -0.07)
TIGIT	0.0045	0.46 (0.15, 0.77)
MHC2	0.012	0.46 (0.1, 0.82)
Cytotoxicity	0.015	0.5 (0.1, 0.9)
PD1	0.016	0.31 (0.06, 0.56)
Treg	0.019	0.27 (0.05, 0.49)
CD8 T-cells	0.022	0.37 (0.06, 0.68)
BRCAness	0.06	0.18 (0, 0.36)
PDL1	0.07	0.25 (-0.02, 0.52)
Apoptosis	0.07	0.43 (-0.03, 0.88)
IFN gamma	0.08	0.62 (-0.07, 1.3)
PDL2	0.11	0.19 (-0.04, 0.42)
Hypoxia	0.13	0.39 (-0.12, 0.89)
Macrophages	0.21	0.12 (-0.07, 0.31)
Inflammatory =	0.33	-0.15 (-0.45, 0.15)
AR =	0.36	0.22 (-0.25, 0.68)
Stroma	0.47	-0.1 (-0.38, 0.17)
APM =	0.49	0.11 (-0.21, 0.44)
Differentiation	0.63	-0.05 (-0.24, 0.15)
Endothelial cells	0.66	-0.03 (-0.17, 0.11)
FOXA1	0.71	0.1 (-0.44, 0.64)
BRCAness	0.73	-0.03 (-0.21, 0.15)
TGF beta	0.80	-0.02 (-0.17, 0.13)
Claudin low	0.81	0.03 (-0.21, 0.27)
Proliferation	0.99	0 (-0.35, 0.35)

Table S4. Interaction between pCR status and treatment: trastuzumab-containing arms vs. lapatinib-only (n=194) for BC360 curated meta-gene signatures (shown in bold) and single genes (shown in italics).

	Unadjusted *P* for the interaction term [a]
ERBB2	0.025
p53	0.081
FOXA1	0.144
Hypoxia	0.154
AR	0.19
Cytotoxicity	0.197
Claudin low	0.215
Stroma	0.224
TGF beta	0.24
IFN gamma	0.282
BRCAness	0.302
Differentiation	0.307
Mast cells	0.324
B7H3	0.346
PD1	0.432
PDL2	0.509
PGR	0.524
Endothelial cells	0.553
Inflammatory chemokines	0.583
MHC2	0.629
Proliferation	0.66
Macrophages	0.661
Treg	0.688
ER signaling	0.786
Apoptosis	0.802
CD8 T-cells	0.813
TIS	0.827
Cytotoxic cells	0.855
PDL1	0.857
IDO1	0.885
TIGIT	0.939
APM	0.939
ESR1	0.952
BRCAness	0.987

[a] (pCR and treatment: trastuzumab-containing vs. lapatinib-only)

Table S5. Selected 38 candidate genes that were prognostic of pCR among 758 non-housekeeping genes on the BC360 panel in patients on trastuzumab-containing regimens with FDR controlled at 0.1 (n = 130)

Genes	**OR (95% CI)**	***P***	**Adjusted *P***
HEMK1	0.22 (0.11, 0.47)	7.9E-5	0.06
GRB7	1.70 (1.30, 2.21)	8.8E-5	0.07
ERBB2	1.73 (1.30, 2.31)	1.6E-4	0.12
ITGB6	1.53 (1.23, 1.91)	1.6E-4	0.12
SOCS2	0.47 (0.32, 0.70)	1.8E-4	0.14
LRP2	0.67 (0.55, 0.83)	2.4E-4	0.18
ADCY9	0.42 (0.26, 0.67)	3.6E-4	0.27
ELOVL2	0.66 (0.53, 0.84)	5.7E-4	0.43
NPEPPS	0.20 (0.08, 0.50)	6.1E-4	0.46
DUSP6	2.22 (1.40, 3.52)	6.8E-4	0.51
MYC	0.33 (0.41, 0.79)	6.9E-4	0.51
IFT140	0.33 (0.18, 0.63)	6.9E-4	0.52
ZNF205	0.21 (0.08, 0.52)	7.7E-4	0.58
TMPRSS4	1.46 (1.16, 1.84)	1.2E-3	0.88
NKG7	1.66 (1.22, 2.25)	1.2E-3	0.93
GNLY	1.61 (1.20, 2.15)	1.4E-3	1
IDO1	1.58 (1.19, 2.09)	1.4E-3	1
MAPT	0.70 (0.57, 0.87)	1.5E-3	1
CXCL9	1.41 (1.14, 1.74)	1.7E-3	1
ESR1	0.79 (0.68, 0.91)	1.8E-3	1
CXCR6	1.78 (1.23, 2.56)	2.0E-3	1
PRKDC	0.37 (0.20, 0.70)	2.0E-3	1
IGF1R	0.62 (0.45, 0.84)	2.2E-3	1
COL27A1	0.56 (0.39, 0.81)	2.2E-3	1
FGFR2	0.65 (0.50, 0.86)	2.3E-3	1
BCL2	0.58 (0.41, 0.83)	2.3E-3	1
GZMB	1.94 (1.27, 2.96)	2.3E-3	1
PRKCB	1.75 (1.22, 2.52)	2.6E-3	1
IKZF3	1.60 (1.18, 2.17)	2.7E-3	1
PGR	0.76 (0.64, 0.91)	2.7E-3	1
SYTL4	0.50 (0.32, 0.79)	2.8E-3	1
PRF1	1.94 (1.25, 3.01)	2.9E-3	1
TCEAL1	0.35 (0.18, 0.71)	3.2E-3	1
NRCAM	0.61 (0.44, 0.85)	3.2E-3	1
GZMA	1.66 (1.18, 2.33)	3.3E-3	1
MYCN	1.38 (1.11, 1.73)	4.4E-3	1
CCR5	1.81 (1.20, 2.72)	4.5E-3	1
TFDP1	0.30 (0.13, 0.69)	4.6E-3	1

Table S6. Selected top 5 BC360 meta-gene signatures and top 10 overall genes for prediction of pCR from univariate analysis among patients on lapatinib-only regimens (n = 64)

	OR (95% CI)	***P***	**Adjusted *P***
BC360 Single & Meta-Gene Signatures			
P53	2.58 (1.31, 5.09)	0.006	0.21
Mast cells	0.62 (0.40, 0.97)	0.038	1
Hypoxia	1.37 (1.01, 1.85)	0.050	1
TIS	1.69 (1.00, 2.85)	0.051	1
Cytotoxic cells	1.69 (1.00, 2.86)	0.058	1
Genes			
IFT140	0.09 (0.02, 0.33)	0.0004	0.27
ZNF205	0.02 (0.00, 0.18)	0.0004	0.29
TCEAL1	0.15 (0.04, 0.51)	0.0025	1
NEIL2	0.15 (0.04, 0.53)	0.0031	1
PTGER3	0.43 (0.25, 0.76)	0.0035	1
MUS81	0.15 (0.04, 0.56)	0.0046	1
PALB2	0.06 (0.01, 0.44)	0.0060	1
BBC3	0.19 (0.06, 0.64)	0.0072	1
CREBBP	0.09 (0.02, 0.53)	0.0075	1
DNAJC12	0.59 (0.40, 0.88)	0.0090	1

Figure S1. Heatmap of expression of 40 gene signatures over all 194 study participants. The heatmap uses unsupervised hierarchical clustering to group signature scores which are scaled by signature to have a mean 0 and standard deviation 1. The signatures are identified at the bottom of the heatmap. Each row is a unique sample. Abbreviations: PCR, pathologic complete response; TRT1, trastuzumab; TRT2, lapatinib; TRT3, trastuzumab and lapatinib. (* indicates single gene expression)

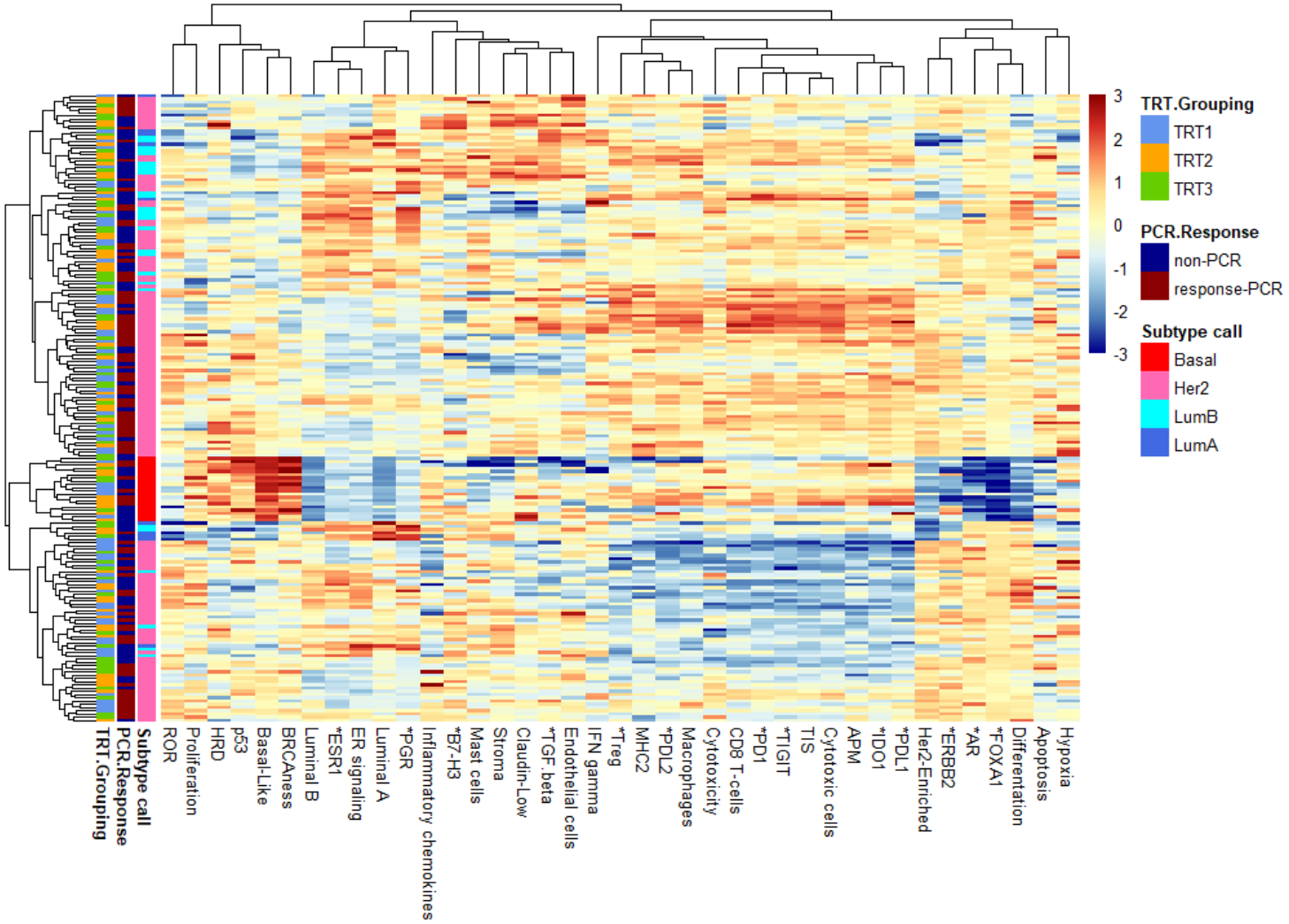

Figure S2. The ROC curve from a 10-fold cross validation in predicting pCR with 19 selected genes and one gene signature in patients on trastuzumab-containing regimens (n = 130)

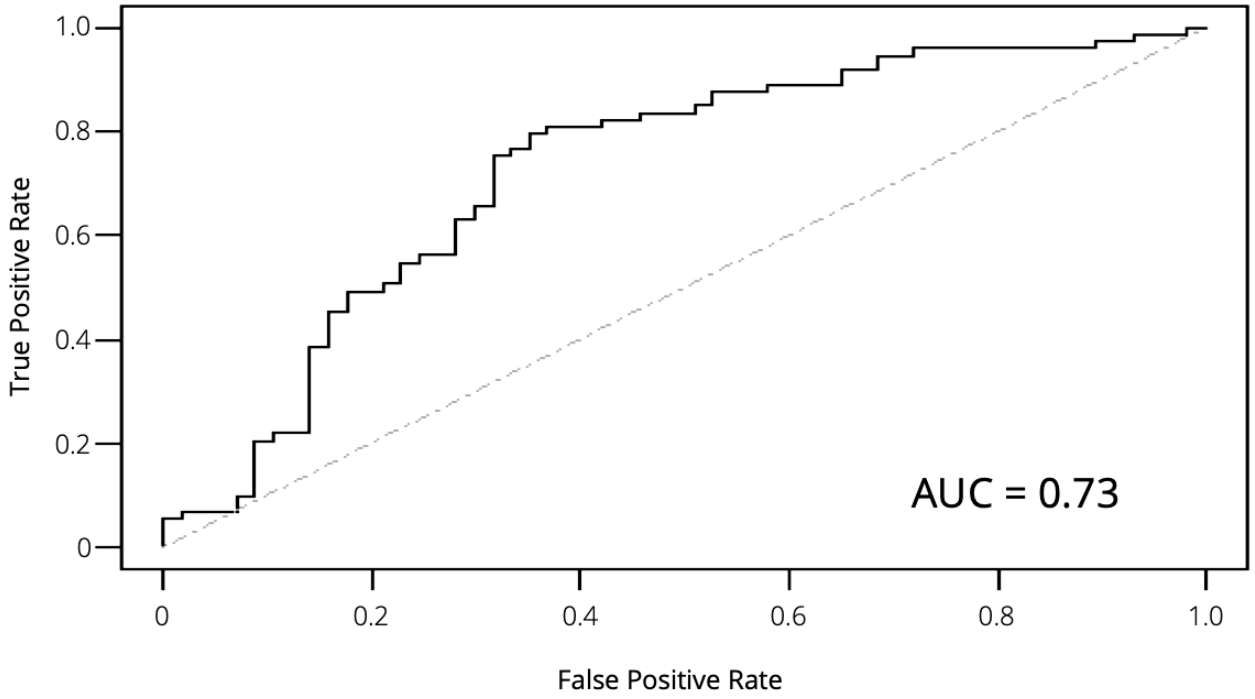

Figure S3. EFS: prognostic utility of individual genes among 758 genes on the BC360 panel in patients under trastuzumab-containing regimens

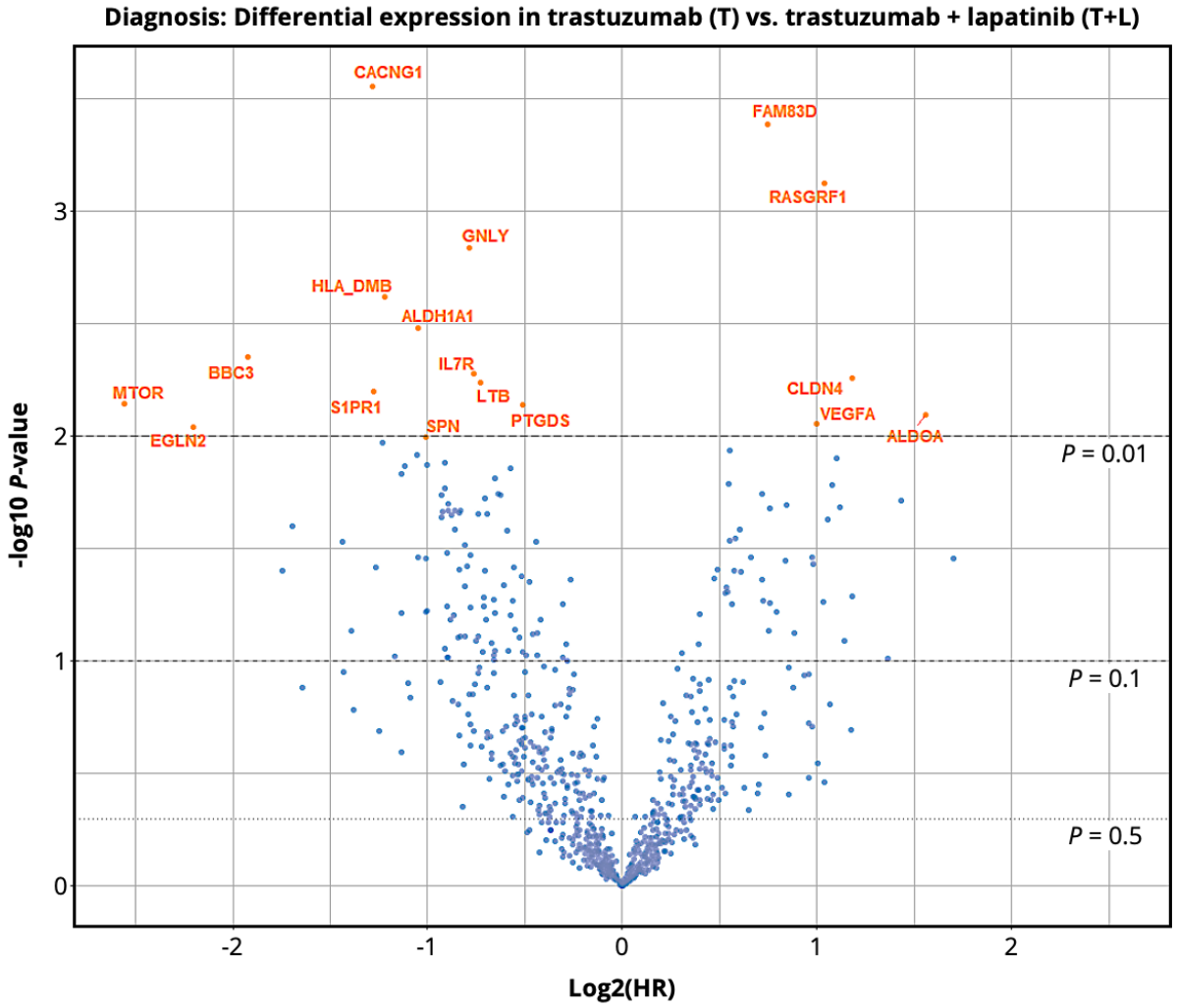

Figure S4. Box plots of the difference between pre-treatment and post-treatment expression levels of 10 selected genes among nine patients on trastuzumab-containing regimens. The P values are from the paired t-tests for comparison.

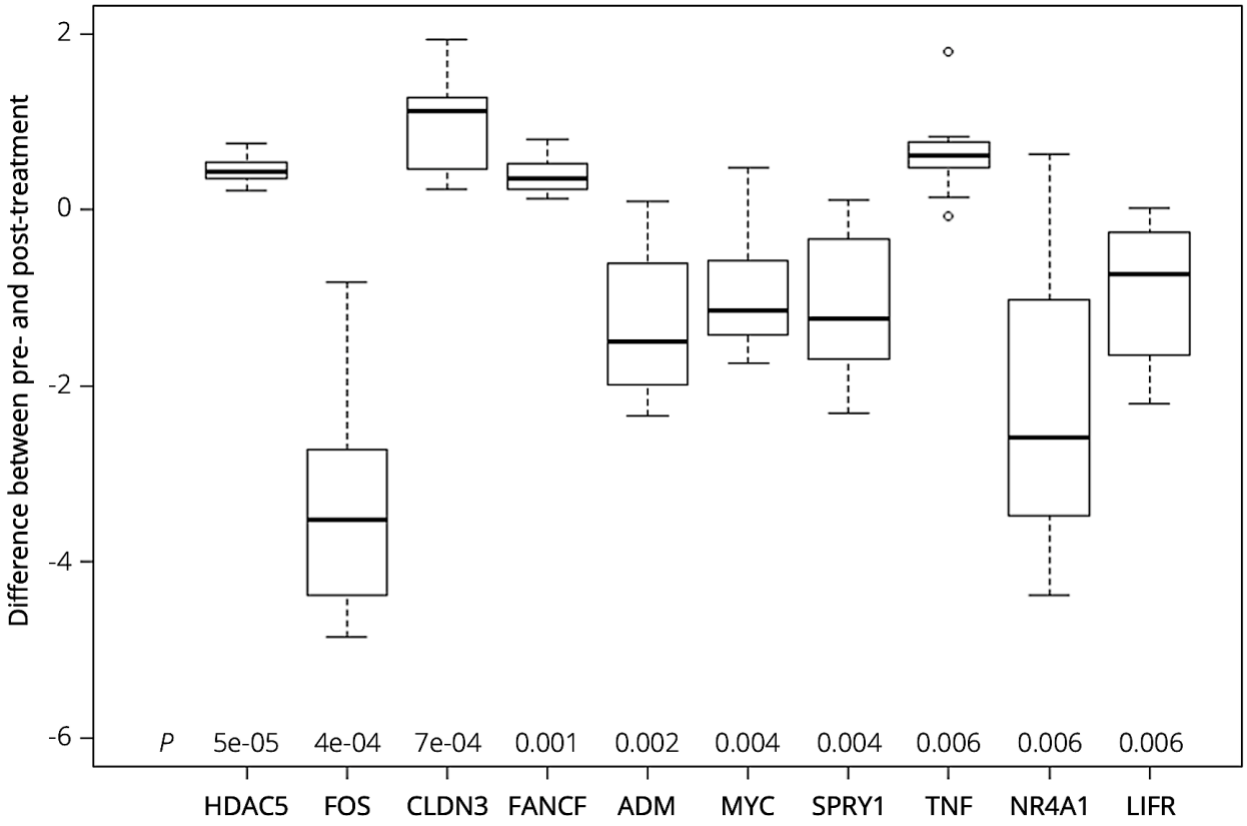

Figure S5. Wheel plots show BC360 signatures from 2 pairs of pre- and post-treatment samples from patients who did not achieve a pCR. In standard wheel plots, selected signatures are shown for the selected sample, with the TIS and PAM50 signatures at the core and other signatures shown as bars around the wheel. The signature scores for each sample is mapped to the empirical distribution of the calibrated breast invasive carcinoma cohort data in TCGA to get a quantile. The wheel plot shows values from 0 to 1 (bars) to indicate the quantiles of the signature scores in TCGA data. For the Pre and Post median wheel plots, we further calculate the median of the quantiles of each signature across samples in Pre and Post groups respectively and display the median of the quantiles for each group (bars around the wheel).

The PAM50 subtype correlations are displayed at the core of the wheel plot and TIS is displayed as a green radial arc around the PAM50 subtype calls. If the radial arc is clockwise, the subtype correlation is positive. A clockwise arc indicates a positive subtype correlation; a counterclockwise arc indicates a negative subtype correlation.