Supplementary Materials

Supplementary Methods

TRYPHAENA Study design

Briefly, 225 patients with centrally confirmed HER2-positive early breast cancer were randomized to receive three different preoperative regimens: either 3 cycles of FEC (Fluorouracil, Epirubicine, Cyclophosphamide) followed by 3 cycles of docetaxel in combination with trastuzumab and pertuzumab that was initiated together with the 1st cycle of FEC or the same chemotherapy regimen in combination with trastuzumab and pertuzumab starting with the first cycle of docetaxel or 6 cycles of an anthracycline-free regimen of docetaxel and carboplatin in combination with trastuzumab and pertuzumab(1). This trial provided evidence that the combination of trastuzumab and pertuzumab with standard chemotherapy is safe in terms of cardiac toxicity (primary objective) and resulted in high pCR rates(1).

The independent ethics committees that approved the TRYPHAENA protocols, and amendments when appropriate, were: CE università Cattolica del S. Cuore -Policl. Gem, Italy; Comitato Etico A.O. San Gerardo di Monza, Italy; EK Heidelberg, Germany; Comisia Nationala de Etica, Romania; Ethic Committee Institute for Oncology and Radiology of Serbia; Agency for medicinal products and medical devices, Bosnia and Herzegovina; Ottawa Hospital Research Ethics Boards; Canada; UBC BCCA Research Ethics Board (BCCA REB), Canada; Comitê de Ética em Pesquisa da PUCRS, Brazil; Comite de Etica em Pesquisa do Centro de Referencia da Saude da Mulher, Brazil; Northern X Ethics Committee, New Zealand; Pharma-Ethics Independent Research Ethics committee, South Africa; Kyungpook

National Uni Hospital, Republic of Korea; TVGH Institutional Review Board, Taiwan; Comité Etico Hospital Vall de Hebron, Spain; Comité de Etico Hospital Universitario Puerta de Hierro, Spain; Western Institutional Review Board, United States; Hospital Angeles Metropolitano, Mexico; Southampton & South West Hampshire LREC (B), UK; Kantonale Ethikkommission Aarau, Switzerland; Korea University Guro Hospital, Republic of Korea; Kantonale Ethikkommission Zürich (KEK), Switzerland; CEIC - Comissão de Ética para Investigação Clínica, Portugal; Regionala Etikprövningsnämnden i Stockholm, Sweden; McGill University Health Centre – Research Ethics Board, Canada; University of Pretoria Research Ethics committee, Faculty of Health Sciences, South Africa; Comité Etico de Euskadi, Spain; Comité Ético de Investigación Clinica Hospital Universitario Reina Sofía, Spain; Central Ethics Committee Agency for medicines and medical devices, Croatia; IRB,China Medical University Hospital, Taiwan; Centro Estatal De Cancerologia Dr Miguel Dorantes Mesa, Mexico; Ministry of Health and Social Welfare, Croatia.

TILs evaluation

The geometric mean of TIL percentages from the two pathologists was computed after adding 0.5 to values equal to 0 consistently with previous methodological developments on TILs evaluation(2). Agreement between the 2 pathologists was evaluated using the concordance correlation coefficient (CCC) as a summary measure of reproducibility(3). For binary variables with no repetitions, CCC is equivalent to Cohen’s kappa.

Sample processing and RNA extraction for gene expression profiling

Macrodissection was performed in order to enrich for tumor cell content. If it was possible to obtain a tumor area of at least 5 mm^2 containing at least 70% viable tumor cells only those area(s) were macrodissected. If collection of the required amount of tumor cells (less than 70% viable tumor cell content) was not feasible, macrodissection was performed on a minimum of 5mm^2 tumor area containing at least 50% viable tumor cells. If this was not feasible or only less than 5mm^2 tumor area are available, the total amount of tumor area is dissected.

A proprietary method developed in house was applied to extract RNA using kit reagents that are equivalent to the now commercially available High Pure FFPET RNA Isolation Kit manufactured by Roche Diagnostics.

Single genes / signature scores calculations and PAM50 subtype definition

The signature scores were calculated as a weighted sum of the log-expressions of their genes, with gene-specific weights equal to +1 or -1 depending on the direction of their association with the gene expression immune phenotype. The selected immune genes and three immune gene signature scores were scaled so that the 2.5% and 97.5% quantiles equaled −1 and +1, respectively.

Gene expression levels were made comparable to those of HER2-positive samples from the C9741 NanoString cohort(4) by using the cross-studies normalization of the R package genefu(5) (R package version 2.4.2). A merged data set was obtained by adding the renormalized TRYPHAENA samples to all C9741 samples. PAM50 subtypes were determined using genefu on the merged data set.

Statistical analysis for gene expression data generated using the NanoString® nCounter Assay

Correlation analysis:

Spearman correlation analysis was performed between the baseline TILs, the selected immune genes, the immune gene signatures and the *ERBB2* and *ESR1* mRNA genes.

Spearman correlation analysis was also performed between the baseline TILs level and the whole 800 gene panel. Genes with spearman coefficient rho higher than 0.30 or below -0.30 and p-values corrected for multiple testing (FDR) of less than .05 were considered statistically significant for the gene ontology (GO) analysis. GO enrichment was performed using the R package GO.db (version 3.4.1) and limma (version 3.32.4).

Association with pCR

The *ERBB2*, *ESR1* and the immune single genes/gene signatures, as well as the HER2 enriched subtype defined with PAM50 were assessed for their association with pCR status using logistic regression. The regression models were evaluated with and without adjustment for baseline clinicopathologic variables including age (<50y vs ≥50y), histology grade (III vs I/II), ER status (positive vs negative), clinical stage (III vs II), chemotherapy (anthracyclines vs non-anthracyclines) and Baseline TILs (10 units increase).

Association with EFS

The prognostic value of *ERBB2*, *ESR1* and the immune single genes/gene signatures, as well as the HER2 enriched subtype defined with PAM50, was assessed using univariate Cox proportional hazard models. The prognostic value of clinicopathological characteristics (age (<50y vs ≥50y), histology grade (III vs I/II), ER status (positive vs negative), clinical stage (III vs II), chemotherapy (anthracyclines vs non-anthracyclines)), PAM50 subtype (HER2-enriched vs non-HER2-enriched), pCR

and baseline TILs were evaluated in a multivariate cox proportional hazard model. The proportional hazards assumption was checked by examining scaled Schoenfeld residuals (6).

P-values were corrected for multi-testing with the Benjamini-Hochberg procedure (FDR).

References

(1) Schneeweiss A, Chia S, Hickish T, et al. - Pertuzumab plus trastuzumab in combination with standard neoadjuvant anthracycline-containing and anthracycline-free chemotherapy regimens in patients with HER2-positive early breast cancer: a randomized phase II cardiac safety study (TRYPHAENA). *Ann Oncol* 2013;24(9):2278-84

(2) Buisseret L, Desmedt C, Garaud S, et al. - Reliability of tumor-infiltrating lymphocyte and tertiary lymphoid structure assessment in human breast cancer. *Mod Pathol* 2017 ;30(9):1204-1212.

(3) Lin L, Hedayat A, Wu W. Statistical Tools for Measuring Agreement. *Springer-Verlag New York* 2012; doi:10.1007/978-1-4614-0562-7.

(4) Liu M, Pitcher BN, Mardis E, et al. PAM50 gene signatures and breast cancer prognosis with adjuvant anthracycline- and taxane-based chemotherapy: correlative analysis of C9741 (Alliance). *npg Breast Cancer* 2016;(doi:10.1038/npjbcancer.2015.23).

(5) Gendoo DM, Ratanasirigulchai N, Schroder MS, et al. genefu: Computation of gene expression-based signatures in breast cancer. *R-package version* 2 4 2 2015.

(6) Grambsch P and Therneau T Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika,* 1994;81(3):515-26

Supplementary Tables

Supplementary Table 1. Univariate linear regressions of baseline TILs on clinicopathological characteristics

Characteristics	No. patients	Estimated difference in TILs means	95% CI	P*
Age, y : ≥50 vs <50	213	-2.7%	-8.1% to 2.8%	.34
Histology grade : III vs I/II	172	6.9%	0.6% to 13.3%	.03
Estrogen Receptor : positive vs negative	213	-0.2%	-5.7% to 5.3%	.95
Clinical Stage : III vs II	211	-1.8%	-7.4% to 3.7%	.51
Chemotherapy : anthracycline-based vs not	213	-3.6%	-9.3% to 2.2%	.23

*P-values based on two-sided t-tests from univariate linear regression models. TILs = tumor infiltrating lymphocytes; CI = confidence interval.

Supplementary Table 2. Multivariate logistic regression of pCR on clinicopathological characteristics and baseline TILs (number of patients included in the model = 170, number of patients with pCR = 98)

Variable	OR	95% CI	P*
Age, y : ≥50 vs <50	1.17	0.61 to 2.26	.64
Histology grade : III vs I/II	1.17	0.59 to 2.29	.66
Estrogen receptor : positive vs negative	0.27	0.14 to 0.54	<.001
Clinical stage : III vs II	0.44	0.22 to 0.88	.02
Chemotherapy : anthracycline-based vs not	0.71	0.35 to 1.42	.33
Baseline TILs (10 units increase)	1.12	0.95 to 1.31	.17

*P-values based on two-sided Wald tests from the multivariate logistic regression model including age, histology grade, estrogen receptor status, clinical stage, chemotherapy and baseline TILs. OR = odds ratio; CI = confidence interval; TILs = tumor infiltrating lymphocytes.

Supplementary Table 3. Univariate linear regressions of TILs at surgery on clinicopathological characteristics for patients with no pCR

Characteristics	No. patients	Estimated difference in TILs means	95% CI	P*
Age, y : ≥50 vs <50	64	-1.6%	-8.8% to 5.7%	0.67
Histology grade : III vs I/II	54	3.3%	-5.2% to 11.8%	0.45
Estrogen receptor : positive vs negative	63	2.4%	-5.1% to 9.9%	0.53
Clinical stage : III vs II	64	-5.4%	-12.3% to 1.6%	0.13
Chemotherapy : anthracycline-based vs not	64	1.3%	-6.6% to 9.2%	0.75

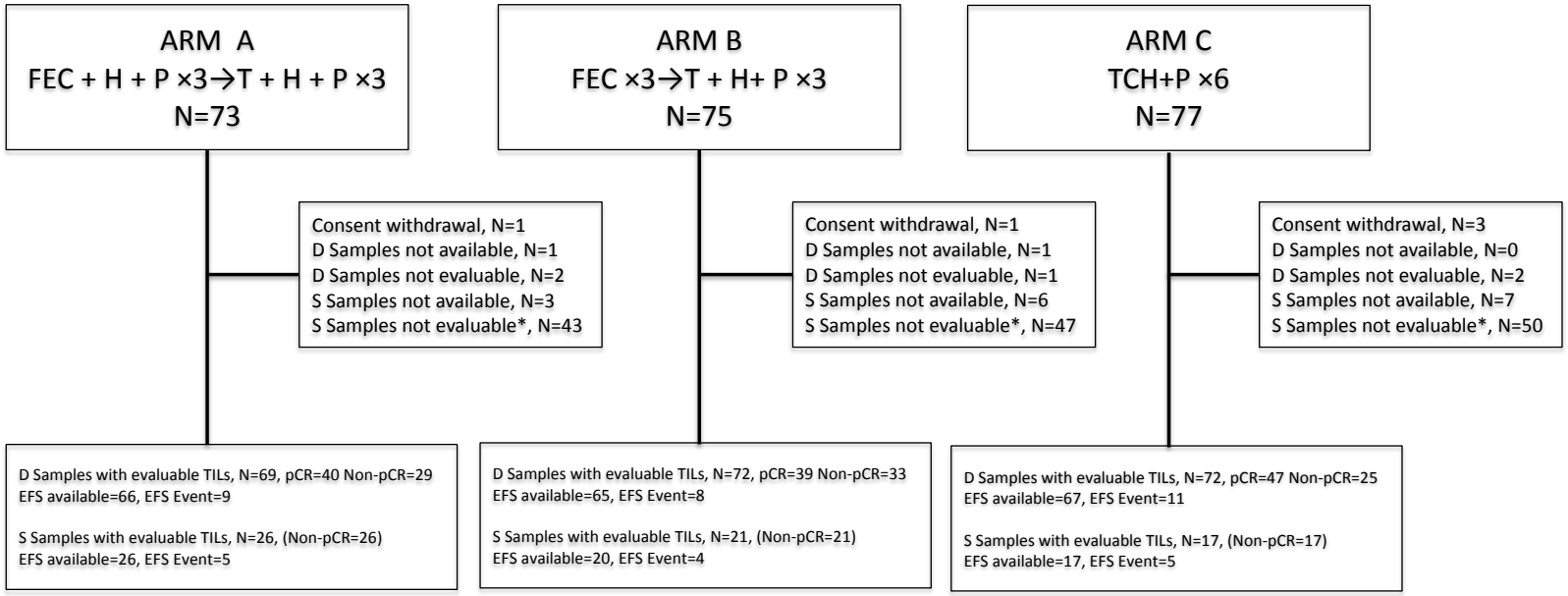
*P-values based on two-sided t-tests from univariate linear regression models. TILs = tumor infiltrating lymphocytes; CI = confidence interval.

Supplementary Table 4. Multivariate cox regression of EFS on clinicopathological characteristics, PAM50 subtype, pCR and baseline TILs (number of patients included in the model =131, number of patients with pCR =22)

Variable	HR	CI	P*
Age : ≥50y vs <50y	1.36	0.57 - 3.24	0.49
Histology grade : III vs II	1.04	0.39 - 2.79	0.94
Estrogen Receptor : positive vs negative	0.85	0.33 - 2.22	0.74
Clinical Stage : III vs II	1.6	0.65 - 3.99	0.31
Chemotherapy : anthracycline-based vs not	1.67	0.67 - 4.18	0.27
HER2-enriched (PAM50): yes vs no	1.56	0.60 – 4.10	0.37
pCR: yes vs no	0.29	0.1 - 0.82	0.02
Baseline TILs (10 units increase)	0.74	0.56 - 1.01	0.06

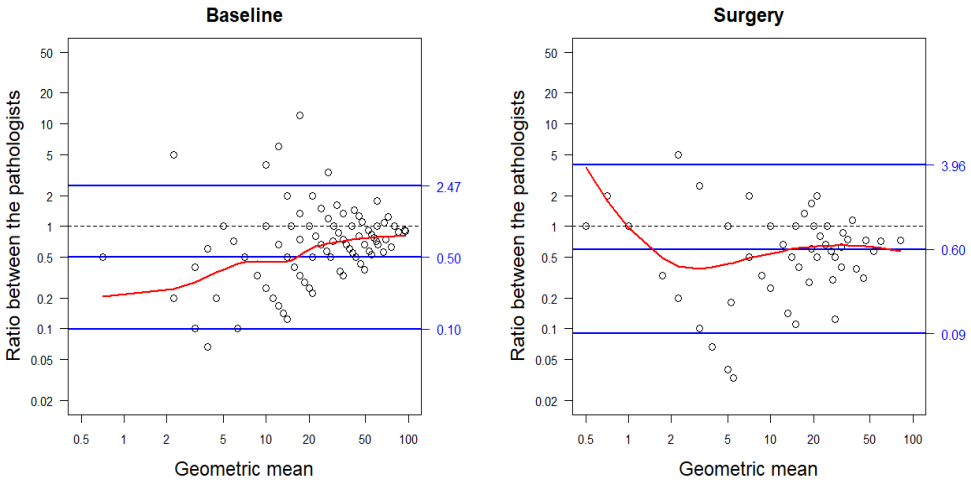
*P-values based on two-sided Wald tests from the multivariate cox regression model including age, histology grade, estrogen receptor status, clinical stage, chemotherapy, baseline TILs and HER2-enriched status. HR = Hazard ratio; CI = confidence interval; TILs = tumor infiltrating lymphocytes.

Supplementary Figure 1. CONSORT diagram for TILs.

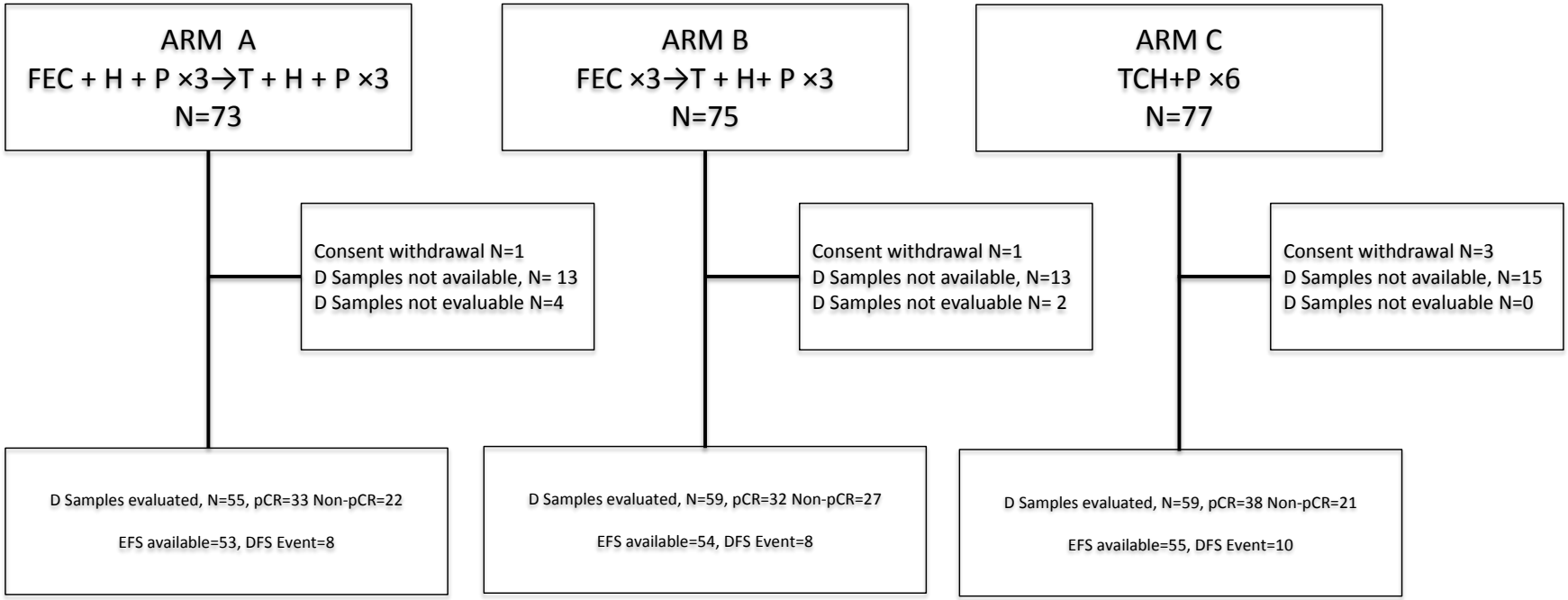


FEC: 5-Fluorouracil, Epirubicin, Cyclophosphamide, H: Trastuzumab, P:Pertuzumab, T: Docetaxel, C: Carboplatin, N: Number of patients, D Samples: Diagnosis Samples, S Samples: Surgery Samples, TILs: Tumor Infiltrating Lymphocytes, pCR pathological complete response, *Surgery Samples were not evaluable because no tumor bed could be identified or patients had pCR

Supplementary Figure 2. Agreement between the two pathologists for logTILs at baseline and at surgery.

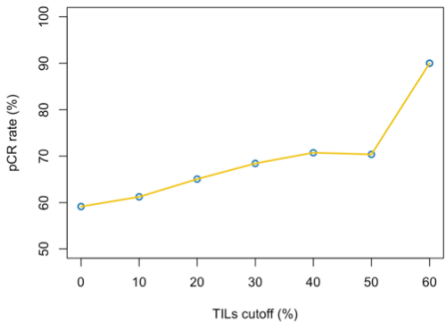


Supplementary Figure 3. CONSORT diagram for NanoString gene expression data.



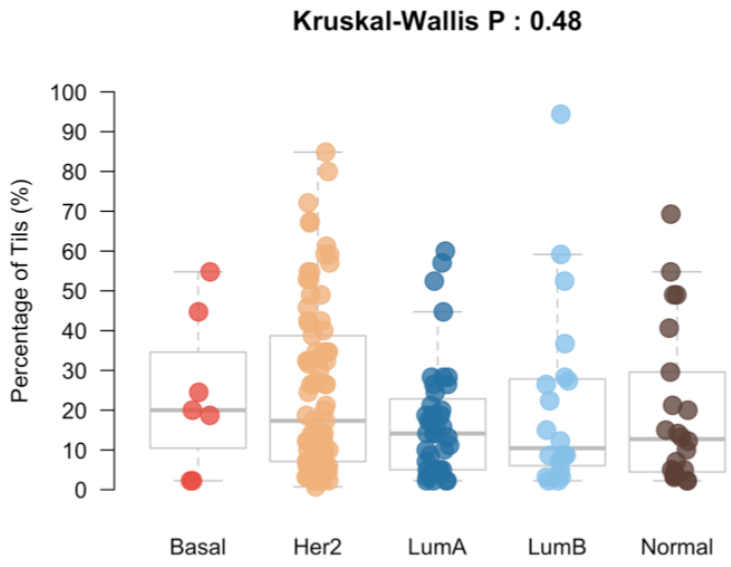
FEC: 5-Fluorouracil, Epirubicin, Cyclophosphamide, H: Trastuzumab, P:Pertuzumab, T: Docetaxel, C: Carboplatin, N: Number of patients, D Samples: Diagnosis Samples, pCR pathological complete response

Supplementary Figure 4. Distribution of pCR rate according to several baseline TILs cutoff (10% increment).



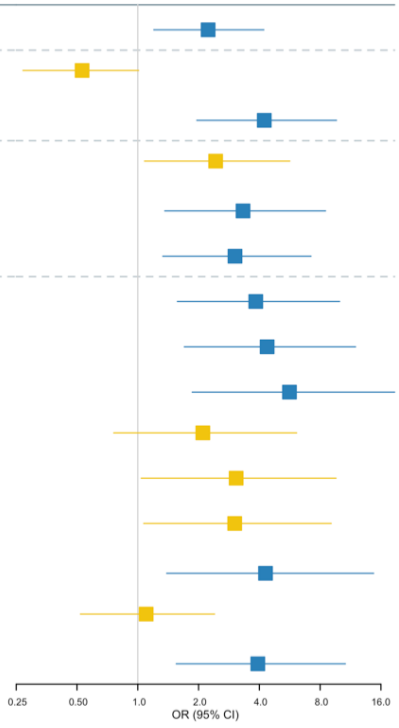
Baseline TILs cut-off	All n (%)	pCR (Yes) n (%)	pCR (No) n (%)
> 0	213 (100)	126 (59)	87 (39)
> 10	129 (100)	79 (61)	50 (39)
> 20	83 (100)	54 (65)	29 (35)
> 30	57 (100)	39 (68)	18 (32)
> 40	41 (100)	29 (71)	12 (29)
> 50	27 (100)	19 (70)	8 (30)
> 60	10 (100)	9 (90)	1 (10)

Supplementary Figure 5. Distribution of baseline TILs according to PAM50 subtypes within HER2-positive breast cancer. Association between TILs and PAM50 subtype was assessed using the two-sided Kruskal-Wallis test.

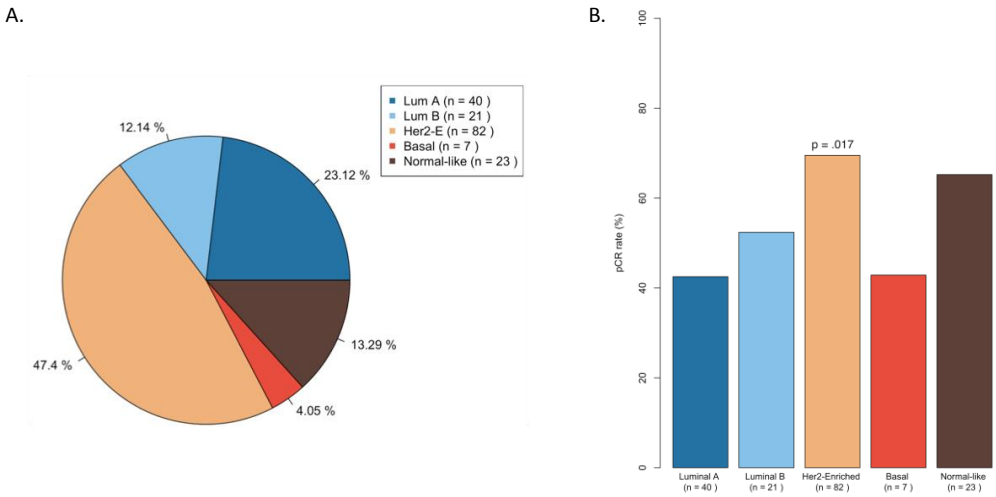


Supplementary Figure 6. Unadjusted odds ratios (ORs) for pCR for a unit increase in baseline single genes/gene signatures score and for change in baseline PAM50 subtype (HER2-enriched vs other) using logistic regression. Horizontal bars represent the 95% confidence intervals of ORs. Characteristics with significant effect (p<.05) after correction for multiple testing using the Benjamini-Hochberg procedure are shown in blue, n=number of patients.

Variable	OR	CI	p-value	FDR
Her2 Enriched (PAM50)	2.23	1.2 to 4.21	0.012	0.022
ESR1	0.53	0.27 to 1.01	0.055	0.064
ERBB2	4.23	1.96 to 9.64	4e-04	0.005
immune_STAT1	2.43	1.08 to 5.66	0.035	0.052
immune_TLS	3.32	1.36 to 8.51	0.010	0.021
immune_trastuzumab	3.03	1.33 to 7.21	0.010	0.021
GZMA	3.84	1.57 to 9.99	0.004	0.015
GZMB	4.37	1.7 to 11.98	0.003	0.015
PRF1	5.64	1.86 to 18.71	0.003	0.015
FOXP3	2.1	0.76 to 6.12	0.160	0.171
CTLA4	3.07	1.04 to 9.59	0.046	0.058
PDL1	3.02	1.07 to 9.08	0.042	0.057
PD1	4.29	1.39 to 14.72	0.015	0.025
IL4	1.1	0.52 to 2.4	0.804	0.804
IFNG	3.93	1.55 to 10.67	0.005	0.015



Supplementary Figure 7. A. Distribution of the baseline PAM50 intrinsic subtypes across the 173 patients. B. PCR rate distribution across the PAM50 intrinsic subtype. Differences between Her2-enriched subtype and the remaining intrinsic subtypes taken together was assessed using a chi-square test ($p<.05$).



Supplementary Figure 8. Unadjusted hazard ratios (HRs) for event-free survival (EFS) for a unit increase in baseline single genes/gene signatures score and for change in baseline PAM50 subtype (HER2-enriched vs other) using Cox regression. Horizontal bars represent the 95% confidence intervals of HRs. Correction for multiple testing was performed using the Benjamini-Hochberg procedure.

Variable	HR	CI	p-value	FDR
Her2 Enriched (PAM50)	1.24	0.58 to 2.68	0.6	0.7
ESR1	0.76	0.33 to 1.76	0.5	0.7
ERBB2	2.24	0.73 to 6.83	0.2	0.5
immune_STAT1	0.51	0.19 to 1.36	0.2	0.5
immune_TLS	0.48	0.16 to 1.42	0.2	0.5
immune_trastuzumab	0.75	0.28 to 1.97	0.6	0.7
GZMA	0.66	0.23 to 1.84	0.4	0.7
GZMB	0.38	0.13 to 1.11	0.08	0.5
PRF1	0.8	0.22 to 2.88	0.7	0.7
FOXP3	0.34	0.09 to 1.28	0.1	0.5
CTLA4	0.39	0.12 to 1.34	0.1	0.5
PDL1	0.53	0.14 to 2.06	0.4	0.7
PD1	1.39	0.36 to 5.32	0.6	0.7
IL4	1.55	0.66 to 3.63	0.3	0.7
IFNG	0.76	0.25 to 2.35	0.6	0.7