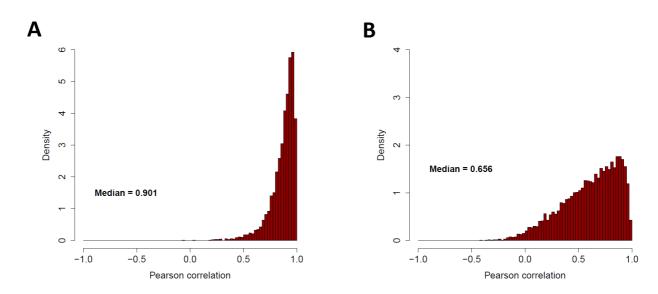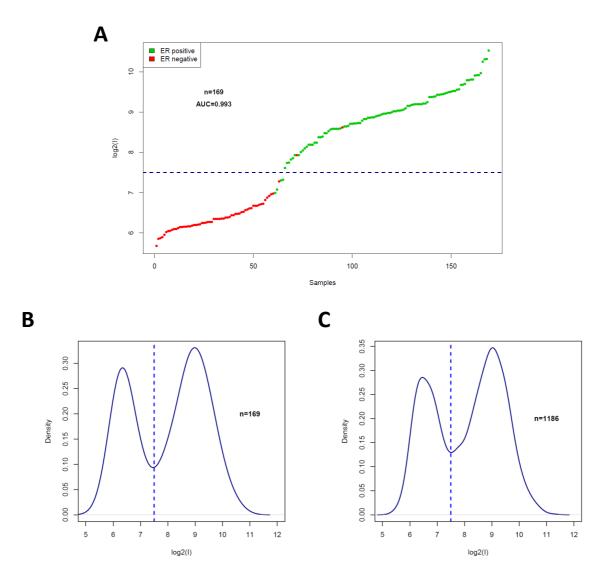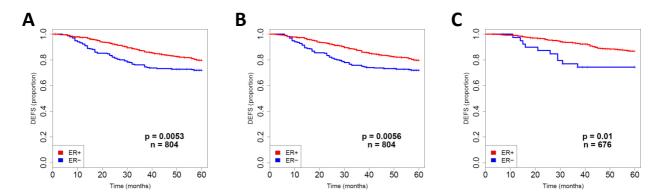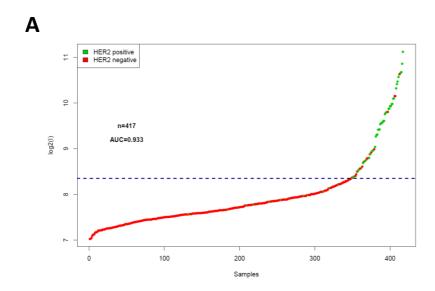# Supplementary Figures

Supplementary Figure 1 - Evaluation of chip bias in the GSE17700 dataset. (A) U133 Plus 2.0-U133A probesets correlation using our alternative CDF and fRMA normalization. (B) U133 Plus 2.0-U133A probesets correlation using the standard Affymetrix CDF and MAS5 normalization
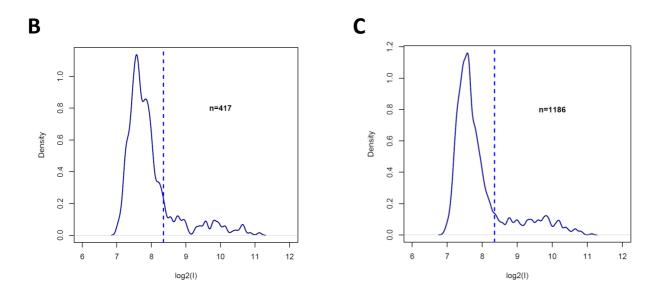
Supplementary Figure 2 - Definition of a metagene-based ER status classifier. (A) ER metagene compared with IHC derived ER status in the training subset from the GENERIC collection. (B) ER metagene distribution in the training subset and (C) in all the GENERIC collection.
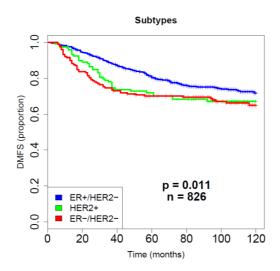

Supplementary Figure 3 - ER status and prognosis. Association with distant metastasis free survival (DMFS) for ER status as originally reported (A) and as defined by our ER status predictor (B) in the PROGNOSTIC dataset. (C) Association with DMFS for ER status as defined by our metagene in the TAM dataset.
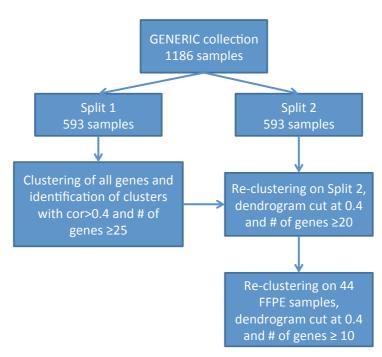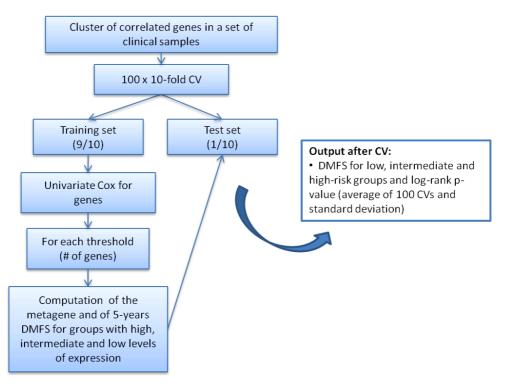
Supplementary Figure 4 - Definition of a metagene-based HER2 status classifier. (A) HER2 metagene compared with IHC/FISH derived HER2 status in the training subset from the GENERIC collection. (B) HER2 metagene distribution in the training subset and (C) in all the GENERIC collection (right).
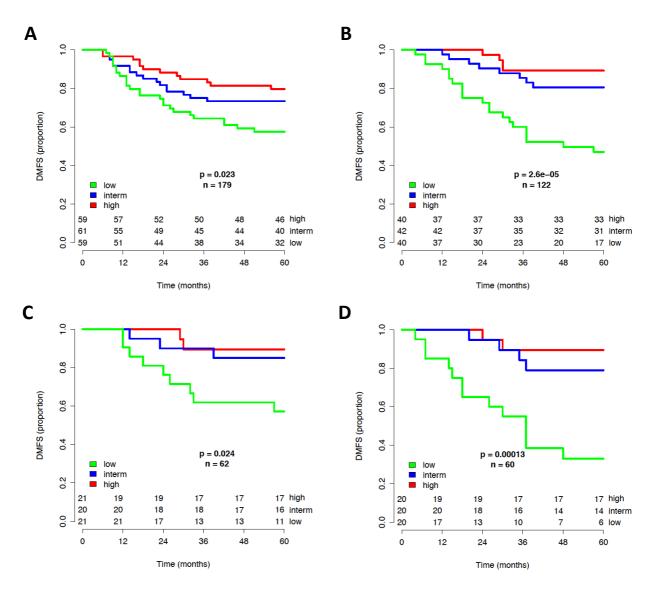
Supplementary Figure 5 - Follow-up of breast cancer subtypes as defined by our ER and HER2 metagenes in the PROGNOSTIC collection

Supplementary Figure 6 - Workflow for metagene identification and confirmation
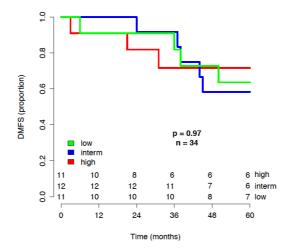
Supplementary Figure 7 - Outline of metagene refinement procedure to improve association with patients' outcome.
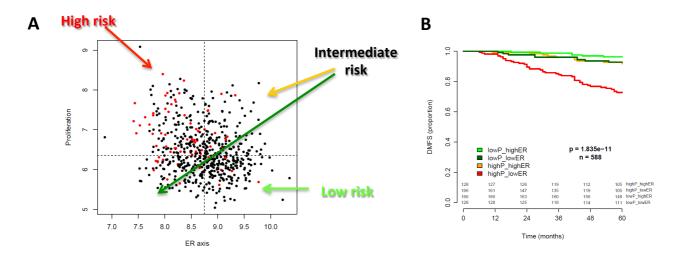
Supplementary Figure 8 - Non cross-validated prognostic value for the CTM in ER-HER2- (A), HER2+ (B), ER-HER2+ (C) and ER+HER2+ subtypes.
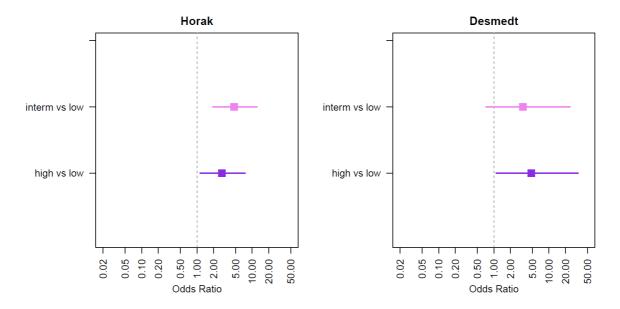
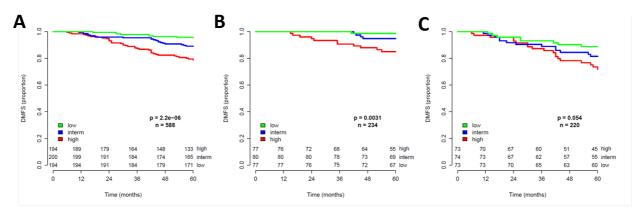**Supplementary Figure 9 - Refined proliferation metagene in all ER+HER2- samples of the TAM collection (A) and separately for N- (B) and N+ samples (C).**
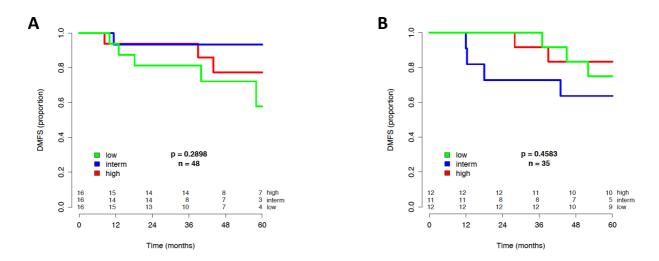
Supplementary Figure 10 - Association between ER-related metagene by tertile and risk of distant relapse in the ER+HER2+ subtype of patients treated with endocrine treatment (TAM collection).
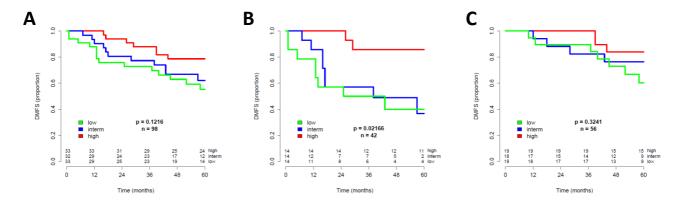
Supplementary Figure 11 - Combination of ER-related and proliferation metagenes. (A) Scatter plot of expression levels of the two metagenes in ER+HER2- samples of the TAM collection; red dots indicate cases having an event within 5 years. Median cut-point for each metagene was plotted defining four distinct groups. (B) Kaplan-Meier analysis for the four groups defined in (A).
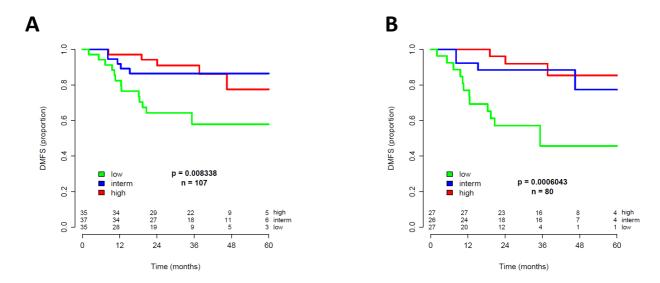

Supplementary Figure 12 - Logistic regression analysis of CTM in association with treatment response in ER- or HER2+ samples from Horak (left) or Desmedt (right) dataset.
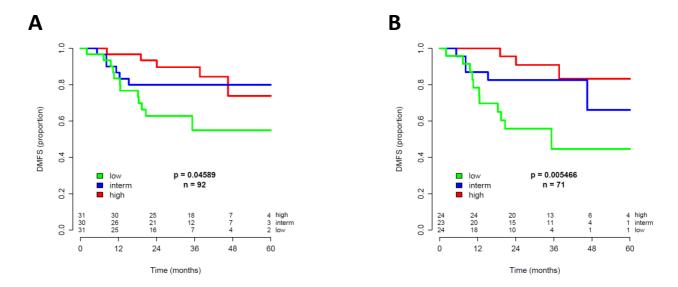
Supplementary Figure 13 - Association between CTM by tertile and risk of distant relapse in ER-HER2+ (A) and ER+HER2+ (B) in patients treated with chemotherapy +/- endocrine treatment (CHEMO collection)

Supplementary Figure 14 - CTM association with long-term outcome in ER- or HER2+ (left), ER-HER2- (center) or HER2+ (right) samples from Petel dataset.

Supplementary Figure 15 - CTM association with long-term outcome in ER- or HER2+ (A) and ER-HER2- samples (B) from Desmedt dataset

Supplementary Figure 16 - Association with long-term outcome according with our CTM in patients with residual disease. ER- or HER2+ (A) or ER-HER2- cases (B) from Desmedt dataset.

A

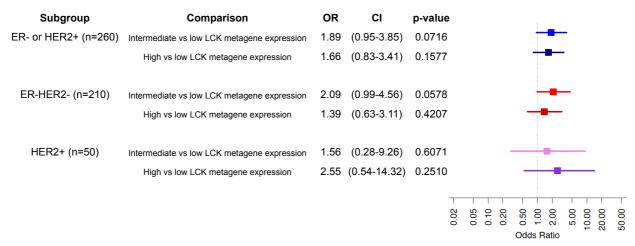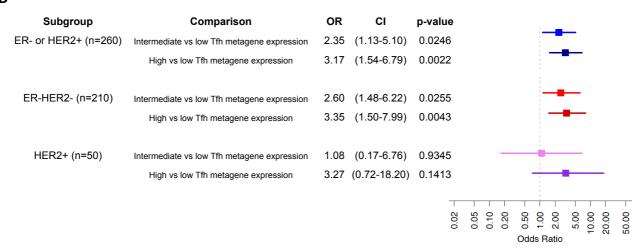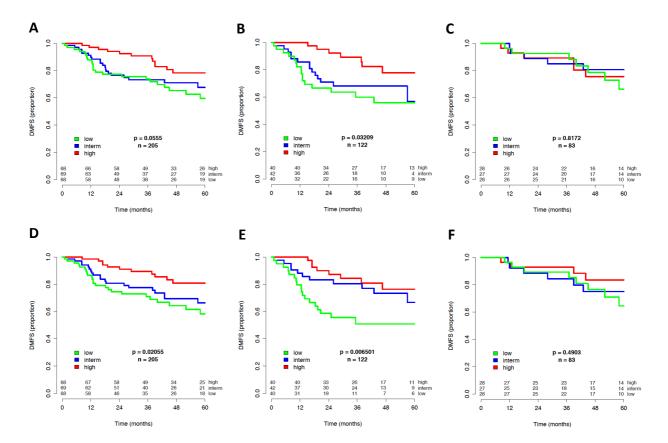| Subgroup	Comparison	OR	CI	p-value
ER- or HER2+ (n=260)	Intermediate vs low LCK metagene expression	1.89	(0.95-3.85)	0.0716
	High vs low LCK metagene expression	1.66	(0.83-3.41)	0.1577
ER-HER2- (n=210)	Intermediate vs low LCK metagene expression	2.09	(0.99-4.56)	0.0578
	High vs low LCK metagene expression	1.39	(0.63-3.11)	0.4207
HER2+ (n=50)	Intermediate vs low LCK metagene expression	1.56	(0.28-9.26)	0.6071
	High vs low LCK metagene expression	2.55	(0.54-14.32)	0.2510

B

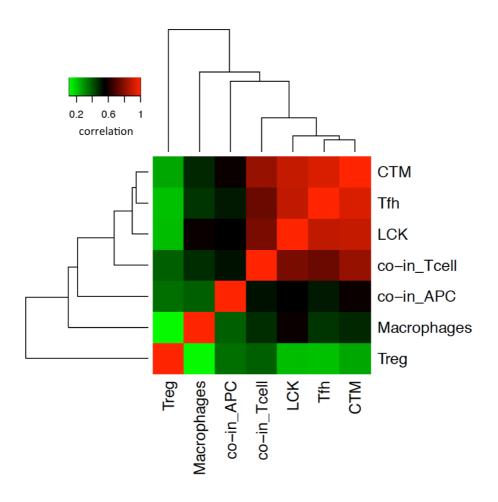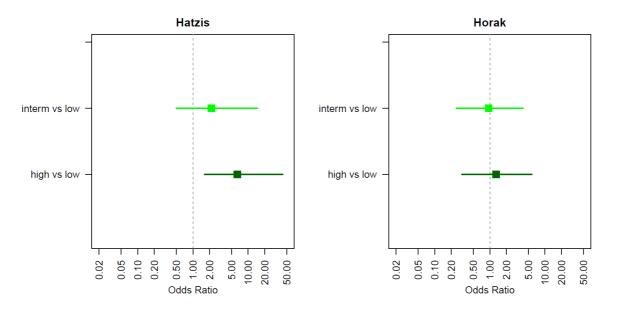| Subgroup	Comparison	OR	CI	p-value
ER- or HER2+ (n=260)	Intermediate vs low Tfh metagene expression	2.35	(1.13-5.10)	0.0246
	High vs low Tfh metagene expression	3.17	(1.54-6.79)	0.0022
ER-HER2- (n=210)	Intermediate vs low Tfh metagene expression	2.60	(1.48-6.22)	0.0255
	High vs low Tfh metagene expression	3.35	(1.50-7.99)	0.0043
HER2+ (n=50)	Intermediate vs low Tfh metagene expression	1.08	(0.17-6.76)	0.9345
	High vs low Tfh metagene expression	3.27	(0.72-18.20)	0.1413

Supplementary Figure 17 - Prognostic role of immune scores (LCK and Tfh) in ER-HER2- and HER2+ breast cancer patients treated with neoadjuvant chemotherapy. Univariate logistic regression analysis of LCK (A) and Tfh (B) immune scores expression and pCR in patients treated with neoadjuvant chemotherapy. Immune scores expression was categorized in low, intermediate and high by tertiles.
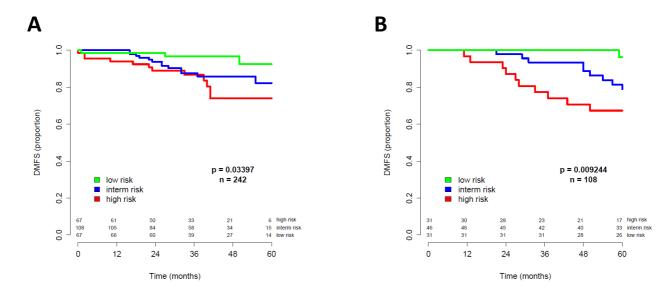
Supplementary Figure 18 - Prognostic role of immune scores in ER-HER2- and HER2+ breast cancer patients treated with systemic chemotherapy. (A, B and C) Kaplan-Meier analysis for the association of the LCK immune scores expression by tertiles with 5-year DMFS in chemotherapy treated patients for the subgroup of ER-HER2- and HER2+ (A), ER-HER2- (B) and HER2+ (C). (D, E and F) Kaplan-Meier analysis for the association of the Tfh immune scores expression by tertiles with 5-year DMFS in chemotherapy treated patients for the subgroup of ER-HER2- and HER2+ (D), ER-HER2- (E) and HER2+ (F). Survival differences were evaluated by log-rank test.
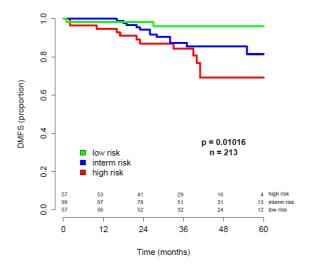
Supplementary Figure 19 - Reciprocal correlation between our immune metagene (IM), Tcell related signatures (LCK and Tfh), Treg, Macrophages and co-inhibitory immune molecules expressed on T cells (co-in_Tcell) or antigen-presenting cells (co-in_APC) (related signature were derived by Rooney et al).
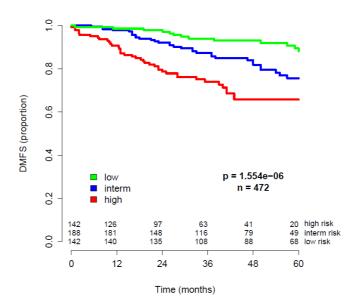
Supplementary Figure 20 - Logistic regression analysis of combined proliferation and ER-related metagenes in association with treatment response in ER+HER2- samples of the Hatzis dataset (left) or Horak dataset (right).

Supplementary Figure 21 - Association with long-term outcome for the combined proliferation and ER-related metagenes in ER+HER2- cases from Hatzis (left) and Petel (right) datasets

Supplementary Figure 22 - Association with long-term outcome for the combined proliferation and ER-related metagenes in the subgroup of ER+HER2- patients with residual disease (Hatzis dataset).

Supplementary Figure 23 - MBRP performance in HER2- (ER- or ER+) treated patients. Risk groups were defined by combining corresponding low, intermediate and high risk groups in ER-HER2- and ER+HER2-patients. Kaplan-Meier analysis for the association between these three risk groups with 5-year DMFS in patients treated with chemotherapy (ER-HER2) or chemo-endocrine treatment (ER+HER2-). Survival differences were evaluated by log-rank test.