*Supplementary Table 1.
Clinical information of breast cancer patient cohorts used for gene expression analysis.

Variable	Cohort 1 N=1062 (100%)	Cohort 2 N=114 (100%)
Age		
≤50 yrs	615 (56.6%)	69 (60.5%)
>50yrs	414 (38.1%)	45 (39.5%)
unknown	58 (5.3%)	
Tumor size		
T1 ≤2cm	679 (62.5%)	20 (14.4%)
T2 >2cm	346 (31.8%)	98 (85.6%)
unknown	58 (5.3%)	
Nodal Status		
Node negative	932 (85.7%)	52 (45.6%)
Node positive	155 (14.3%)	62 (54.3%)
Histologic Grade		
Grade 1	148 (13.6%)	2 (1.8%)
Grade 2	373 (34.3%)	20 (17.5%)
Grade 3	417 (38.4%)	76.3 (76.3%)
Unknown	149 (13.7%)	5 (4.4%)
ER status		
Positive	761 (70%)	0
Negative	326 (30%)	114 (114%)
HER2 overexpressed		
Positive	162 (15%)	29 (25.4%)
Negative	921 (84.7%)	85 (74.6%)
unknown	4 (0.4%)	
Endpoint	Disease recurrence (28.6%)	pCR (14%)
Type of tumor biopsy	At primary surgery	At diagnosis (core biopsy)
Systemic anti cancer treatment given	none	Epirubicin monotherapy 4 cycles (100 mg/m^2) prior to surgery

*This table is in part reproduced in Ref 27.