

Appendix

Description of each patient cohort

- 1) SOTIRIOU et al (<http://www.pnas.org/cgi/content/full/1732912100/DC1>) - This cohort contained 99 node-negative and node-positive breast cancer patients. All of the tumor samples were invasive ductal carcinomas. 46 individuals were node negative and 53 were node positive. 16 patients with tumor grade 1, 38 patients with tumor grade 2 and 45 patients with tumor grade 3. 65 patients were ER positive and 34 ER negative patients. Two patients received PMF Chemotherapy; 30 patients received CMF; and two received Adrymycin, CMF chemotherapy.
- 2) BILD et al (GSE3143) - This cohort contained 157 patients, 110 with ER level 1 and 47 with ER level 0.
- 3) WANG et al (GSE2034) - This cohort contained 286 lymph-node-negative patients of which, 146 were of stage T1, 132 of stage T2, and 8 of stage T3/4. This cohort contained 209 ER positive and 77 ER negative patients. There were 165 PR positive, 111 PR negative and 10 with unknown PR status. 148 patients were with poor grade, 42 with moderate grade, 7 with good grade and 89 with unknown grade. There were 139 pre-menopausal and 147 post-menopausal patients.
- 4) VAN DE VIJVER et al (<http://www.rii.com/publications/2002/nejm.html>) - There were 295 consecutive patients with primary breast carcinomas, 151 with lymph node negative disease, and 144 with lymph node positive disease.
- 5) MILLER et al (GSE3494) – This cohort contained 236 patients, 62 patients with Grade I, 121 with Grade II, 51 with Grade III and 2 patients with unknown grade information. 201 patients were ER positive and 31 patients were ER negative. There were 179 PR positive patients and 57 PR negative patients. The cohort contained 78 lymph node positive patients and 149 lymph node negative patients.
- 6) LOI et al (GSE6532) – This cohort contained 137 untreated patients and 277 tamoxifen treated patients. Gene expression profiles of 327 patients were screened on GPL96 Affymetrix Gene Chip Human Genome U133 Array Set HG-U133A platform and 87 patient expression profiles were generated on GPL570 Affymetrix GeneChip Human Genome U133 plus 2.0 Array. The cohort contained 250 lymph node negative patients and 143 lymph node positive patients. There were 82 patients with Grade I, 182 patients with Grade II, and 76 patients with Grade III. There were 349 ER positive patients and 45 ER negative patients.
- 7) IVSHINA et al (GSE4922) - This cohort contained patient and tumor samples of the Uppsala and Singapore sets. The Uppsala set was composed of 249 patients. The Singapore set contained 40 patients. There were 211 ER positive patients and 34 ER negative patients. This cohort contained 81 lymph node positive patients and 159 lymph node negative patients.
- 8) SORLIE et al (GSE4335) - This cohort contained 122 tissue samples of which, 77 carcinomas and 7 nonmalignant tissues were previously published. There were 83 ER positive patients and 32 ER negative patients. There were 34 Lymph node negative patients and 46 lymph node positive patients. The cohort contains 11 patients with Grade I, 49 patients with Grade II, and 53 patients with Grade III.

Supplementary Table 1: List of 28-gene breast cancer prognostic signature

Gene	Spot ID	Clone_IMAGE	UniGene Cluster ID
PSMC3	3912	198917	Hs.463079* (Hs.383019)
TOMM70A	4919	198312	Hs.227253
MCF2	2370	268412	Hs.387262
RAD52 homolog	418	1377154	Hs.552577
MCM2	1881	239799	Hs.477481
C18B11	5984	131988	Hs.173311
SEC13L	6497	757210	Hs.301048
SLC25A5	5182	291660	Hs.522767
PLSCR1	6959	268736	Hs.130759
TXNRD1	7296	789376	Hs.434367
RAD50	2925	261828	Hs.242635
-	6498	46196	
INPPL1	1987	703964	Hs.523875
-	583	501651	Hs.439445*
TXNRD1	6736	789376	Hs.434367
PBX2	536	80549	Hs.509545
SSBP1	3434	125183	Hs.490394
HSPCB (heat shock 90kD protein 1, beta)	2403	34396	Hs.448229
PDGFRA	6674	376499	Hs.74615
ACOT4	6555	488202	Hs.49433
DDOST	2416	50666	Hs.523145
Immunoglobulin alpha (1 or 2) heavy chain constant region IGHA1	2276	182930	Hs.497723
S100P	5593	135221	Hs.2962
FAT	7009	591266	Hs.481371
FGF2	3514	324383	Hs.284244
INSM1	3061	22895	Hs.89584
IRF5	5962	260035	Hs.521181
SMARCD2	2923	741067	Hs.250581
MAP2K2	1652	769579	Hs.465627

(*Retired)

Supplementary Table 2: Association between the gene expression-defined risk groups and clinicopathologic parameters in predicting disease-free survival.

Clinical Parameters	Good Signature Group (n=653)	Poor Signature Group (n=684)	p-value
Age <=50 (n=430) >50 (n=621) Unknown (n=286)	143/653 252/653 258/653	287/684 369/684 28/684	0.019
Lymph Node Status Positive(n=444) Negative(n=870) Unknown(n=23)	163/653 484/653 6/653	281/684 386/684 17/684	6.6 e-10
ER Status Positive(n=1075) Negative(n=248) Unknown(n=14)	549/653 98/653 6/653	526/684 150/684 8/684	0.0013
Grade I (n=168) II(n=327) III(n=245) Unknown(n=597)	88/653 124/653 41/653 400/653	80/684 203/684 204/684 197/684	9.8 e-14

Supplementary Table 3: Association between the gene expression-defined risk groups and clinicopathologic parameters in predicting overall survival.

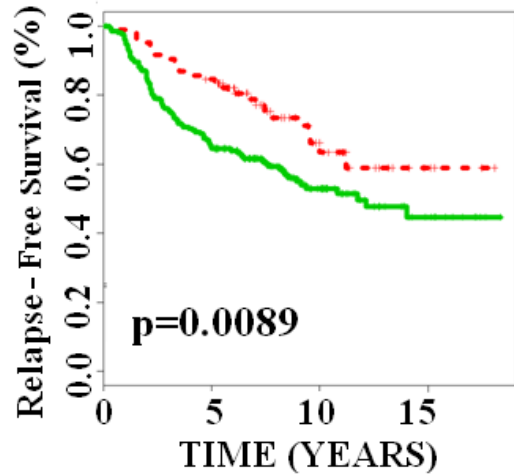
Clinical Parameters	Good Signature Group(n=336)	Poor Signature Group(n=470)	p-value
Age <=50 (n=300) >50 (n=112) Unknown (n=394)	96/336 40/336 200/336	204/470 72/470 194/470	0.5515
Lymph Node Status Positive(n=300) Negative(n=334) Unknown(n=172)	108/336 135/336 93/336	192/470 199/470 79/470	0.2887
ER Status Negative(n=179) Positive(n=618) Unknown(n=9)	57/336 274/336 5/336	122/470 344/470 4/470	0.003723
Grade I (n=148) II (n=270) III (n=223) Unknown(n=165)	82/436 117/436 43/436 94/436	66/505 153/505 180/505 71/505	8.549e-13

Supplementary Table 4: Drug responses in breast cancer cell lines.

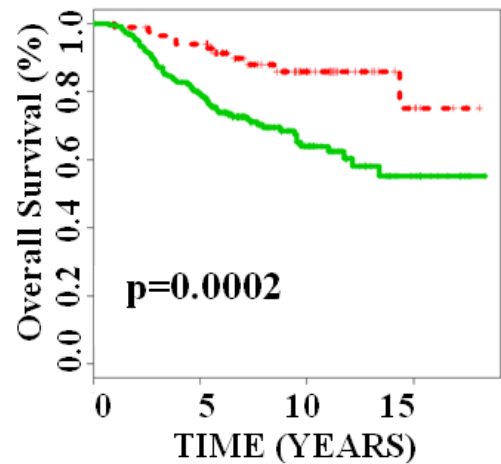
Cell Line	BR:BT_549	BR:HS578T	BR:MCF7	BR:MDA_MB_231	BR:T47D
5FU	Resistant	Resistant	Sensitive	Resistant	Resistant
Methotrexate	Resistant	Resistant	Sensitive	Resistant	Resistant
Tamoxifen	Resistant	Resistant	Sensitive	Intermediate	Sensitive
Adriamycin	Resistant	Intermediate	Sensitive	Resistant	Intermediate
Docetaxel	Sensitive	Sensitive	Sensitive	Intermediate	NA
Paclitaxel	Intermediate	Sensitive	Sensitive	Intermediate	Resistant
Doxorubicin	Resistant	Resistant	Sensitive	Resistant	Intermediate

DATASET: VAN DE VIJVER ET AL

van de Vijver's Cohort (n=295)



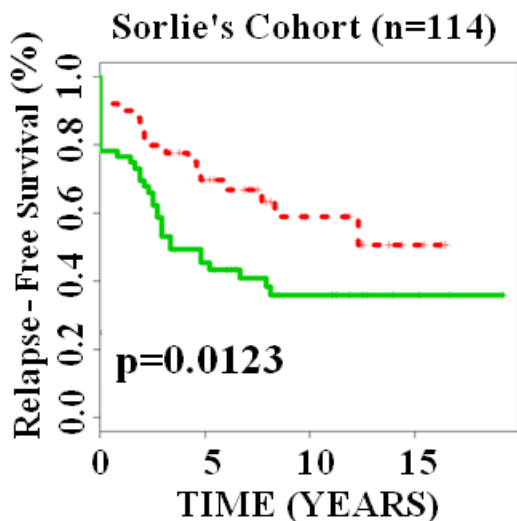
van de Vijver's Cohort (n=295)



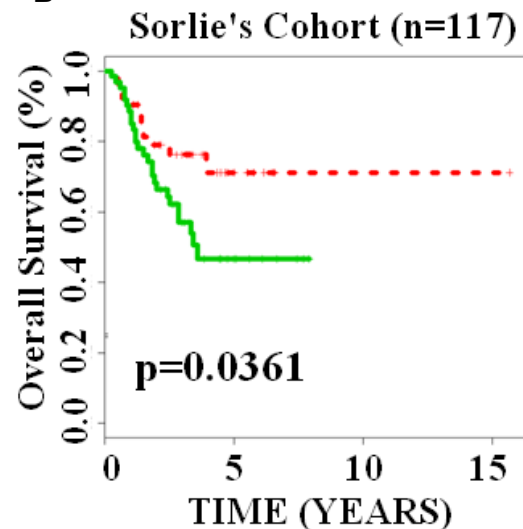
Suppl. Figure 1. The 28-gene signature classifies breast cancer patients from van de Vijver's cohort into prognostic subgroups with distinct relapse-free survival and overall survival in Kaplan-Meier analysis. Correlation of 0.15 with the "Good Prognosis" centroid in the training set was taken as a cut-off for stratifying patients into each sub-group. There were 84 patients in "Good Prognosis" group and 211 patients in "Poor Prognosis" group.

DATASET: SORLIE ET AL

A

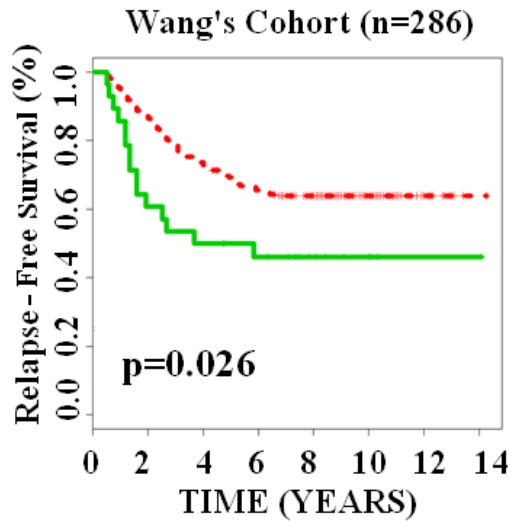


B



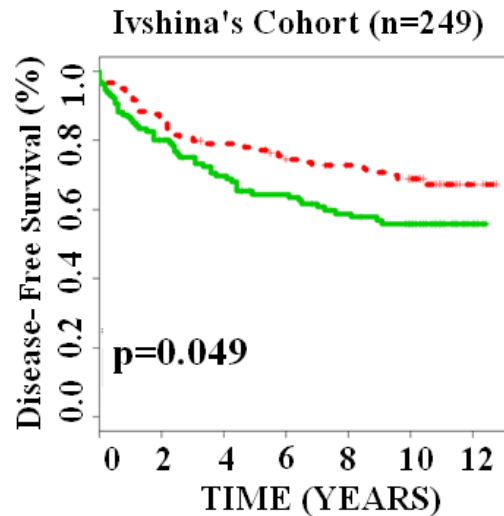
Suppl. Figure 2. The 28-gene signature classifies breast cancer patients from Sorlie's cohort into prognostic subgroups with distinct relapse-free survival and overall survival in Kaplan-Meier analysis. Correlation of 0.15 with the "Good Prognosis" centroid in the training set was taken as a cut-off for stratifying patients into each sub-group. (A) There were 50 patients in "Good Prognosis" and 64 patients in "Poor Prognosis" group. (B) There were 52 patients in "Good Prognosis" group and 65 patients in "Poor prognosis" group.

DATASET: WANG ET AL



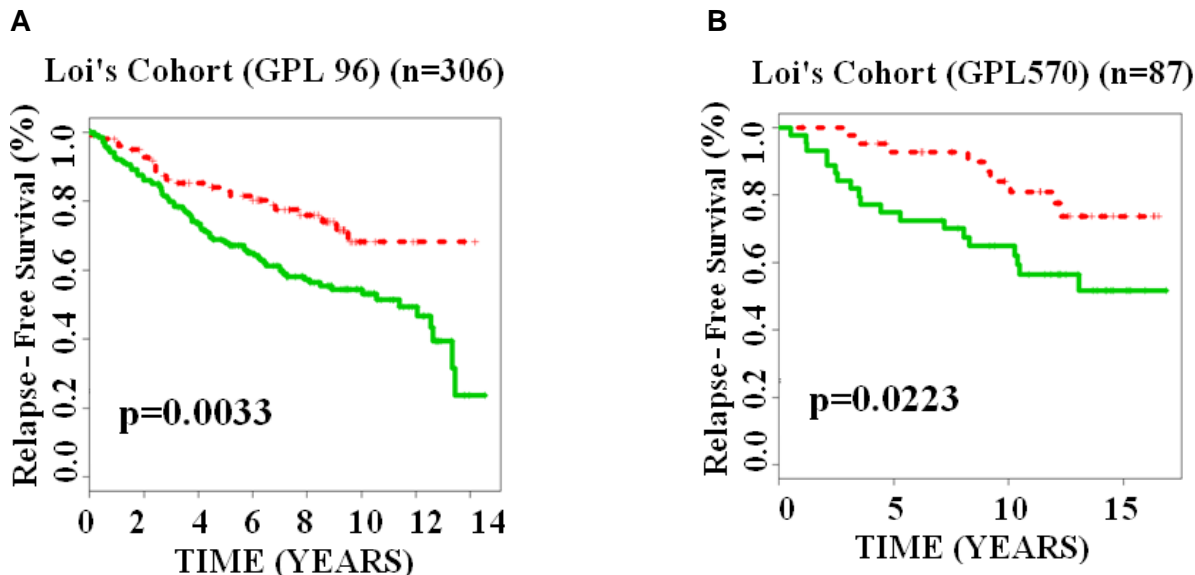
Suppl. Figure 3. The 28-gene signature classifies breast cancer patients from Wang's cohort into prognostic subgroups with distinct relapse-free survival in Kaplan-Meier analysis. Correlation of -0.4 with the "Good Prognosis" centroid in the training set was taken as a cut-off for stratifying patients into each sub-group. There were 258 patients in "Good Prognosis" group and 28 patients in "Poor Prognosis" group.

DATASET: IVSHINA ET AL



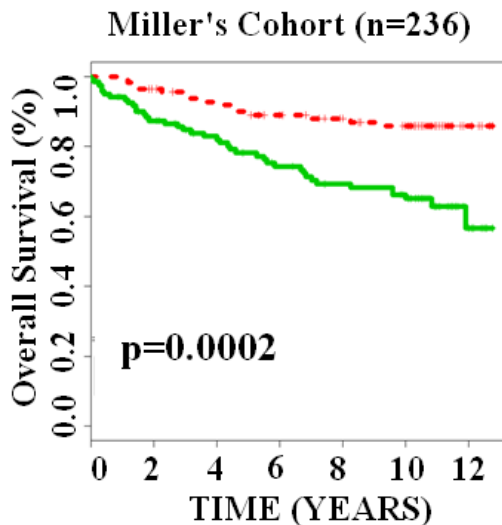
Suppl. Figure 4. The 28-gene signature classifies breast cancer patients from Ivshina's cohort into prognostic subgroups with distinct disease-free survival in Kaplan-Meier analysis. Correlation of -0.4 with the "Good Prognosis" centroid in the training set was taken as a cut-off for stratifying patients into each sub-group. There were 120 patients in "Good Prognosis" group and 129 patients in "Poor Prognosis" group.

DATASET: LOI ET AL



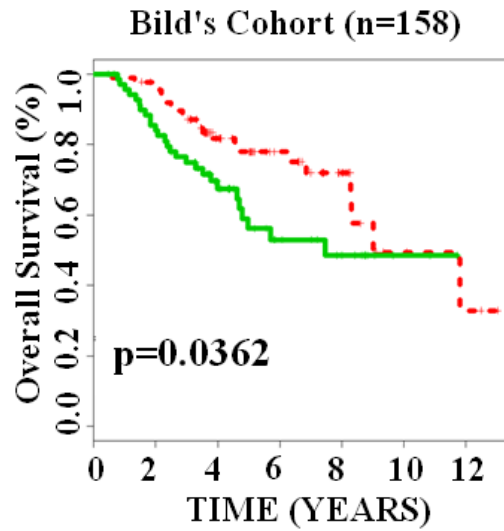
Suppl. Figure 5. The 28-gene signature classifies breast cancer patients from Loi's cohort into prognostic subgroups with distinct relapse-free survival in Kaplan-Meier analysis. A correlation of -0.4 with the "Good Prognosis" centroid of the training set was taken as a cut-off for Affymetrix platforms GPL 96, and a correlation of -0.5 was taken as a cut-off for GPL 570. (A) There were 98 patients in "Good Prognosis" group and 208 patients in "Poor Prognosis" group. (B) There were 43 patients in "Good Prognosis" group and 44 patients in "Poor Prognosis" group.

DATASET: MILLER ET AL



Suppl. Figure 6. The 28-gene signature classifies breast cancer patients from Miller's cohort into prognostic subgroups with distinct overall survival in Kaplan-Meier analysis. Correlation of -0.3 with "Good Prognosis" centroid of the training set was taken as a cut-off for stratifying patients into each sub-group. There were 113 patients in "Good Prognosis" group and 123 patients in "Poor Prognosis" group.

DATASET: BILD ET AL



Suppl. Figure 7. The 28-gene signature classifies breast cancer patients from Bild's cohort into prognostic subgroups with distinct overall survival in Kaplan-Meier analysis. Correlation of -0.3 with "Good Prognosis" centroid of the training set was taken as a cut-off for stratifying patients into each sub-group. There were 87 patients in "Good Prognosis" group and 71 patients in "Poor Prognosis" group.