

Supplementary Table S1: Datasets used in the study (n=40)

Datasets used in the study:

Dataset *	Data source	Number of samples	Reference (PubMed link)	Ref.
Stockholm	GSE1456	159	http://www.ncbi.nlm.nih.gov/pubmed/16280042	1
EORTC	GSE1561	49	http://www.ncbi.nlm.nih.gov/pubmed/15897907	2
Rotterdam-EMC344	GSE2034, GSE5327	344	http://www.ncbi.nlm.nih.gov/pubmed/15721472 , 17420468	3,4
expO	GSE2109	301	http://www.intgen.org/expo/	
New York	GSE2603	99	http://www.ncbi.nlm.nih.gov/pubmed/16049480	5
Oxford-Untreated	GSE2990 (n=61), GSE6532 (n=8)	69	http://www.ncbi.nlm.nih.gov/pubmed/16478745	6
Uppsala	GSE3494 (n=251), GSE6232 (n=5), GSE4922 (n=1), GSE2990 (n=1)	258	http://www.ncbi.nlm.nih.gov/pubmed/16141321 ,%2 017079448	7,8
Boston	GSE3744	40	http://www.ncbi.nlm.nih.gov/pubmed/16473279	9
Signapore	GSE5364	183	http://www.ncbi.nlm.nih.gov/pubmed/18636107	10
Edinburgh	GSE5462	116	http://www.ncbi.nlm.nih.gov/pubmed/17885619	11
London	GSE6532	87	http://www.ncbi.nlm.nih.gov/pubmed/17401012	12
Oxford-Tamoxifen	GSE6532	109	http://www.ncbi.nlm.nih.gov/pubmed/16478745	6
Berlin	GSE6596	24	http://www.ncbi.nlm.nih.gov/pubmed/17410534	13
TransBIG	GSE7390	198	http://www.ncbi.nlm.nih.gov/pubmed/17545524	14
London-2	GSE9195	77	http://www.ncbi.nlm.nih.gov/pubmed/18498629	15
Tampa	GSE10780	39	http://www.ncbi.nlm.nih.gov/pubmed/19266279	16
Mainz	GSE11121	200	http://www.ncbi.nlm.nih.gov/pubmed/18593943	17
Veridex-Tam	GSE12093	136	http://www.ncbi.nlm.nih.gov/pubmed/18821012	18
Rotterdam-EMC204	GSE12276	204	http://www.ncbi.nlm.nih.gov/pubmed/19421193	19
Genentech	GSE12763	30	http://www.ncbi.nlm.nih.gov/pubmed/19567590	20
Paris	GSE13787	23	http://www.ncbi.nlm.nih.gov/pubmed/19055754	21
BIG1-98	GSE16391	55	http://www.ncbi.nlm.nih.gov/pubmed/19573224	14
TOP	GSE16446	120	http://www.ncbi.nlm.nih.gov/pubmed/21422418	22
MDA100	GSE16716	100	http://www.ncbi.nlm.nih.gov/pubmed/20064235	23
SET1	GSE17705	103	http://www.ncbi.nlm.nih.gov/pubmed/20697068	24
SET2	GSE17705	195	http://www.ncbi.nlm.nih.gov/pubmed/20697068	24
IPC_HER2	GSE17907	51	http://www.ncbi.nlm.nih.gov/pubmed/20932292	25
Seattle	GSE18728	21	http://www.ncbi.nlm.nih.gov/pubmed/20012355	26
Boston_Neo-Cisplatin	GSE18864	84	http://www.ncbi.nlm.nih.gov/pubmed/20100965	27
Boston_2	GSE19615	115	http://www.ncbi.nlm.nih.gov/pubmed/20098429	28
StLouis	GSE19697	24	http://www.ncbi.nlm.nih.gov/pubmed/19967557	29
Edinburgh	GSE20181	60	http://www.ncbi.nlm.nih.gov/pubmed/20697427	11
MAQC_add_GSE20194	GSE20194	45	http://www.ncbi.nlm.nih.gov/pubmed/20676074	30
MDA_139	GSE20271 (n=139 additional to MDA133, GSE16716, and GSE20194)	139	http://www.ncbi.nlm.nih.gov/pubmed/20829329	31
IPC	GSE21653	266	http://www.ncbi.nlm.nih.gov/pubmed/20490655	32
St-Cloud	GSE22035	43	http://www.ncbi.nlm.nih.gov/pubmed/21209903	33
Nashville	GSE22513	28	http://www.ncbi.nlm.nih.gov/pubmed/20068102	34
San Francisco	http://www.ebi.ac.uk/arrayexpress/ experiments/E-TABM-158	118	http://www.ncbi.nlm.nih.gov/pubmed/17157792	35
Neo-Trastuzumab	https://array.nci.nih.gov/caarray/pr oject/harri-00137	22	http://www.ncbi.nlm.nih.gov/pubmed/17317830	36
MDA133	http://bioinformatics.mdanderson.o rg/pubdata.html	133	http://www.ncbi.nlm.nih.gov/pubmed/16896004	37
TOTAL:		4467		

*** Remarks:** The complete *TransBIG* dataset contains independent replicate samples from 19 patients of the *Uppsala* cohort and 22 patients of the *Oxford-Untreated* cohort. Datasets "MAQC_add_GSE20194" and "MDA_139" contain only the subsets of 45 and 139 nonredundant samples from the GEO series GSE20194 and GSE20271, respectively, which are not already covered by the *MDA133* and *MDA100* datasets.

References to Supplementary Table S1

1. Pawitan, Y. *et al.* Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* **7**, R953-64 (2005).
2. Farmer, P. *et al.* Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* **24**, 4660–4671 (2005).
3. Wang, Y. *et al.* Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**, 671–679 (2005).
4. Minn, A. J. *et al.* Genes that mediate breast cancer metastasis to lung. *Nature* **436**, 518–524 (2005).
5. Minn, A. J. *et al.* Lung metastasis genes couple breast tumor size and metastatic spread. *Proc Natl Acad Sci U S A* **104**, 6740–6745 (2007).
6. Sotiriou, C. *et al.* Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst* **98**, 262–272 (2006).
7. Miller, L. D. *et al.* An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A* **102**, 13550–13555 (2005).
8. Ivshina, A. V. *et al.* Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* **66**, 10292–10301 (2006).
9. Richardson, A. L. *et al.* X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell* **9**, 121–132 (2006).
10. Yu, K. *et al.* A precisely regulated gene expression cassette potentially modulates metastasis and survival in multiple solid cancers. *PLoS Genet* **4**, e1000129 (2008).
11. Miller, W. R., Larionov, A., Anderson, T. J., Evans, D. B. & Dixon, J. M. Sequential changes in gene expression profiles in breast cancers during treatment with the aromatase inhibitor, letrozole. *Pharmacogenomics J* **12**, 10–21 (2012).
12. Loi, S. *et al.* Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* **25**, 1239–1246 (2007).
13. Klein, A. *et al.* Comparison of gene expression data from human and mouse breast cancers: identification of a conserved breast tumor gene set. *Int J Cancer* **121**, 683–688 (2007).
14. Desmedt, C. *et al.* The Gene expression Grade Index: a potential predictor of relapse for endocrine-treated breast cancer patients in the BIG 1-98 trial. *BMC Med Genomics* **2**, 40 (2009).
15. Loi, S. *et al.* Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics* **9**, 239 (2008).
16. Chen, D.-T. *et al.* Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue. *Breast Cancer Res Treat* **119**, 335–346 (2010).
17. Schmidt, M. *et al.* The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res* **68**, 5405–5413 (2008).
18. Zhang, Y. *et al.* The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. *Breast Cancer Res Treat* **116**, 303–309 (2009).
19. Bos, P. D. *et al.* Genes that mediate breast cancer metastasis to the brain. *Nature* **459**, 1005–1009 (2009).
20. Hoeflich, K. P. *et al.* In vivo antitumor activity of MEK and phosphatidylinositol 3-kinase inhibitors in basal-like breast cancer models. *Clin Cancer Res* **15**, 4649–4664 (2009).
21. Marty, B. *et al.* Frequent PTEN genomic alterations and activated phosphatidylinositol 3-kinase pathway in basal-like breast cancer cells. *Breast Cancer Res* **10**, R101 (2008).
22. Desmedt, C. *et al.* Multifactorial approach to predicting resistance to anthracyclines. *J Clin Oncol* **29**, 1578–1586 (2011).
23. Popovici, V. *et al.* Effect of training-sample size and classification difficulty on the accuracy of genomic predictors. *Breast Cancer Res* **12**, R5 (2010).
24. Symmans, W. F. *et al.* Genomic index of sensitivity to endocrine therapy for breast cancer. *J Clin Oncol* **28**, 4111–4119 (2010).
25. Sircoulomb, F. *et al.* Genome profiling of ERBB2-amplified breast cancers. *BMC Cancer* **10**, 539 (2010).
26. Korde, L. A. *et al.* Gene expression pathway analysis to predict response to neoadjuvant docetaxel and capecitabine for breast cancer. *Breast Cancer Res Treat* **119**, 685–699 (2010).
27. Silver, D. P. *et al.* Efficacy of neoadjuvant Cisplatin in triple-negative breast cancer. *J Clin Oncol* **28**, 1145–1153 (2010).
28. Li, Y. *et al.* Amplification of LAPTM4B and YWHAZ contributes to chemotherapy resistance and recurrence of breast cancer. *Nat Med* **16**, 214–218 (2010).
29. Lin, Y. *et al.* A gene expression signature that predicts the therapeutic response of the basal-like breast cancer to neoadjuvant chemotherapy. *Breast Cancer Res Treat* **123**, 691–699 (2010).
30. Shi, L. *et al.* The MicroArray Quality Control (MAQC)-II study of common practices for the development and validation of microarray-based predictive models. *Nat Biotechnol* **28**, 827–838 (2010).
31. Tabchy, A. *et al.* Evaluation of a 30-gene paclitaxel, fluorouracil, doxorubicin, and cyclophosphamide chemotherapy response predictor in a multicenter randomized trial in breast cancer. *Clin Cancer Res* **16**, 5351–5361 (2010).
32. Sabatier, R. *et al.* A gene expression signature identifies two prognostic subgroups of basal breast cancer. *Breast Cancer Res. Treat.* **126**, 407–420 (2011).
33. Cizkova, M. *et al.* Gene expression profiling reveals new aspects of PIK3CA mutation in ERalpha-positive breast cancer: major implication of the Wnt signaling pathway. *PLoS ONE* **5**, e15647 (2010).
34. Bauer, J. A. *et al.* Identification of markers of taxane sensitivity using proteomic and genomic analyses of breast tumors from patients receiving neoadjuvant paclitaxel and radiation. *Clin Cancer Res* **16**, 681–690 (2010).
35. Chin, K. *et al.* Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10**, 529–541 (2006).

36. Harris, L. N. *et al.* Predictors of resistance to preoperative trastuzumab and vinorelbine for HER2-positive early breast cancer. *Clin Cancer Res* **13**, 1198–1207 (2007).
37. Hess, K. R. *et al.* Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* **24**, 4236–4244 (2006).