Supplementary methods

1.1 Patient material

Test cohort Our testing material is referred to here as the Uppsala cohort. It consists of a population-based cohort of 315 primary breast cancers that were consecutively presented and subsequently resected in the Swedish county of Uppsala from January 1st 1987 to December 31st 1989. Of note, this figure represents 65% of all breast tumours operated on in Uppsala County during this time period. Immediately following surgery, the tumours were divided into two with one part frozen in isopentane and stored at -80°C until further analysis whilst the other part was immersed and fixed in 10% formalin for immunohistochemical assessment. Of the 315 patients, frozen tumours were available for 293. Of these, 263 had sufficient RNA quantity and quality for expression array profiling and after Affymetrix quality control checks 253 tumours remained. Patient characteristics of these 253 tumours divided according to ER status is displayed in Supplementary Table S1 and exclusion criteria are shown as a flow chart in Supplementary Figure S1A. In total, 109 patients received systemic adjuvant therapy, mostly intravenous, 3-weekly CMF-based therapy (in premenopausal patients) or adjuvant tamoxifen (postmenopausal patients). 143 patients did not receive adjuvant therapy and treatment information was not available for one case. All tumors were confirmed to have invasive cancer and were graded according to Elston–Ellis criteria with the exception of one patient for whom primary tumour slides were absent, however the presence of axillary lymph nodes confirmed invasiveness. Of note, breast cancer specific survival information (BCSS) has been updated up to 31st December 2008 giving a median follow-up time of 13 years. Both microarray studies were approved by the ethics committees at Karolinska Institutet and Karolinska university Hospital, respectively, Stockholm, Sweden.

Validation cohort Our validation material is referred to here as the Stockholm cohort. The Stockholm-Gotland breast cancer registry (established in 1976) was used to identify all breast cancer patients who were operated on in the Karolinska hospital from January 1st 1994 to December 31st 1996 (*n*=524). Of these 524 patients, frozen tumours were available for 293. Patients were excluded on the basis of living abroad (7), refused participation (6), degraded tumours (42), insufficient RNA (35), did not pass QC (12), profiled on another array type (14), received neo-adjuvant therapy (12),

in situ carcinoma (5) and stage IV cancer (1), leaving 159 patients who had sufficient RNA quantity and quality for expression array profiling. Patient characteristics of these 159 tumours divided according to ER status is displayed in Supplementary Table S10 and exclusion criteria are shown as a flow chart in Supplementary Figure S1B. In total, 126 received systemic adjuvant therapy, which was mostly intravenous CMF-based therapy or adjuvant tamoxifen. 33 patients did not receive any systemic adjuvant therapy. All tumors were graded according to Elston–Ellis and survival information has been updated up to 31st December 2010 giving a median follow-up time of 14.5 years.

1.2 Immunohistochemical assessment of Ki67 expression Ki67 immunohistochemical staining was performed on 4 µm thick serial sections from paraffin embedded primary breast tumors. Sections were deparaffinized and treated with Envision™ Flex target retrieval solution, low pH, (DAKO, Glostrup, Denmark) for 20 minutes at 85°C in the DAKO PT Link module. Staining was performed using a monoclonal mouse anti-human Ki67 antibody, clone MIB-1 (DAKO, Glostrup, Denmark) diluted 1:100, in the DAKO autostainer plus staining system. Full details of the autostainer program are shown in supplementary materials Table 1 (Table SM1). below. Staining was evaluated by an experienced pathologist (JC) without prior knowledge of patient outcome or tumor characteristics. Only tumor cells with nuclear expression of Ki67 were considered positive. Stromal and inflammatory cells were not counted. Counts were performed using a Zeiss microscope using a reticule to ensure counting proceeded in an unbiased fashion. Counting was performed at the invasive edge of the tumor, and included 200 tumor cells, as deemed important for Ki67 interpretation and scoring (Dowsett et al., 2011).

Table SM1.

Ki67 automated staining protocol for the DAKO Autostainer Plus, consecutive steps

Reagent name	Volume (µl)	Incubation time (mins)
Buffer[*]	-	0
Peroxidase Block, DAKO Real S 2023	150	5
Buffer	-	0
Ki67 clone MIB-1, 1:100	150	40
Buffer	-	0
EnVision FLEX+ Mouse (LINKER)[**]	150	20
Buffer	-	5
EnVision horseradish peroxidase, K 5007	150	30
Buffer	-	0
Buffer	-	5
Diaminobenzidine, K 5007	300	10
Buffer	-	0
Buffer	-	5
Hematoxylin	150	5
DI Water	-	0
Buffer	-	5
DI Water	-	0
Buffer	-	5

[*] Buffered solution containing hydrogen peroxide and preservative

[**] Amplifies the signal of primary mouse antibodies

1.3 Ki67 cutoffs selection

Due to the considerable range of cutoffs used in the literature (Dowsett et al., 2011) (1% – 28.6%), we used a broad approach to identify all reasonable levels above or equal to which staining was considered positive. Firstly, a MEDLINE search for Ki67 AND breast cancer AND 2010 or 2011 retrieved 164 publications of which 22 were directly relevant, of these 22, 10% (n= 7) and 20% (n= 8) were the most commonly used cutoffs.

Secondly, as the relevance of a certain cutoff may to some extent be laboratory-specific, median values of 11% and 12% for Ki67 were identified in the Uppsala and Stockholm cohorts, respectively.

Thirdly, we applied a model of two normal distributions, attempting to capture the visually apparent distribution of low-expressing Ki67 tumors versus the dispersed high-expressing tumours using the mixtools package (Benaglia et al., 2009) in R. This is represented diagrammatically for the Uppsala dataset in supplementary materials Figure 1 (Figure SM1) below. The mixtools package identified 2 normal distributions in the Ki67 data of the Uppsala cohort, designated by the red and green lines in Fig. SM1, bars represent the distribution of Ki67. The red line captures the distribution of

tumours with low expression of Ki67 and encompasses 74% (171/233) of all tumours in the Uppsala cohort. This curve has mu= 7.14 and sigma= 5.18, meaning that a cutoff of Ki67=16% contains 90% of this lower modeled distribution: (7.14 + (5.18*1.64)= 16.43), rounded down to 16%. Using identical methods a distribution-based cutoff of 17% was identified for the Stockholm cohort.

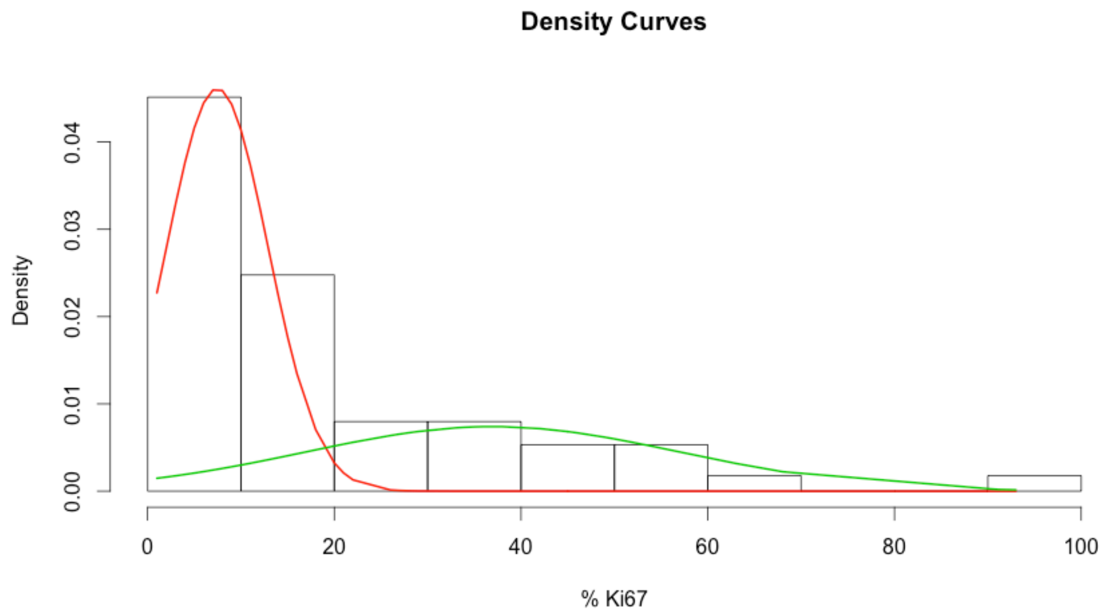

Figure SM1, mixtools plot for the Uppsala cohort.

Fourthly, as we aimed to determine whether Ki67 can achieve similar good vs. poor tumour classifications as gene expression signatures, we selected cutoffs most relevant to each signature using receiver operating characteristics (ROC). For this analysis we employed the Daim package within the R environment. The Daim ROC function requires a continuous variable (in this case Ki67) and a binary variable (here we used binary collapses for each gene expression signature in turn) in order to determine the sensitivity and specificity for each Ki67 value and plot an ROC curve. Next we employed the "point-on-curve" method to determine which Ki67 value gave the minimum distance from the (0,1) point to the ROC curve.

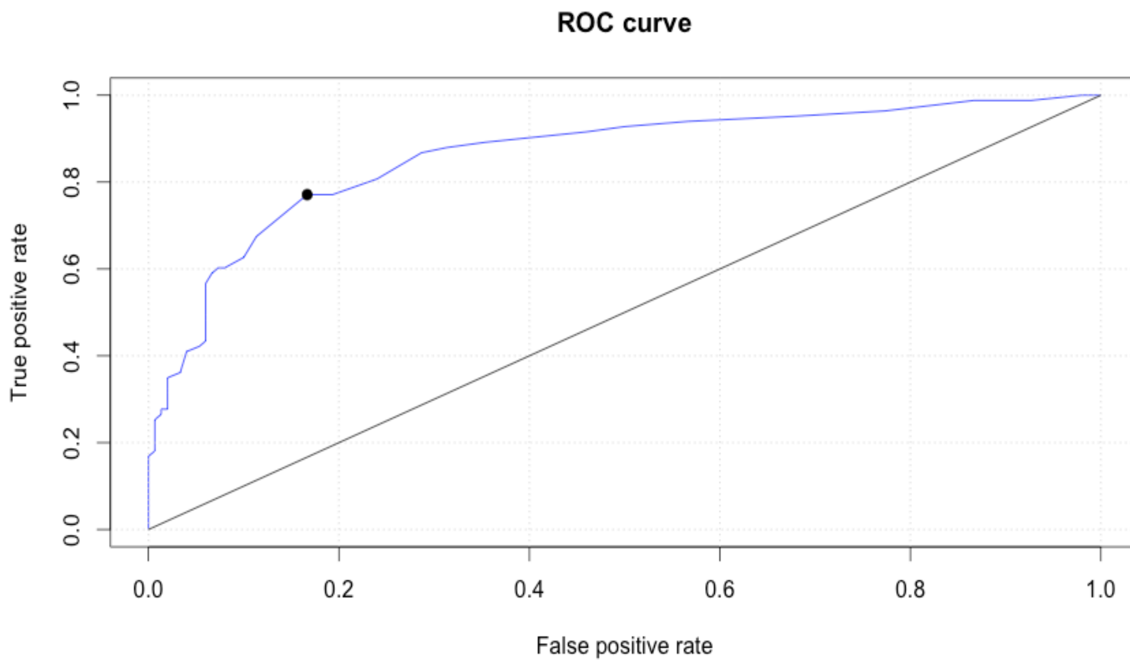

Figure SM2, Typical ROC curve

This is calculated from the output from the ROC function that includes the True positive rate (TPR= sensitivity) and the False positive rate (FPR= 1- specificity). From this we can work out the distance from each Ki67 value to the ROC curve using the equation:

$$d = \sqrt{(1 - Sensitivity)^2 + (1 - Specificity)^2}$$

Finally, the Ki67 value that gives the smallest distance to the curve is chosen as the most appropriate cutoff (shown as a black dot on the ROC curve in supplementary materials Figure SM2). Combining all of these methods we came up with the cutoffs shown in supplementary materials Table SM2 in the Uppsala and Stockholm cohorts (Table SM2).

Table SM2.

All Ki67 cutoffs, Uppsala and Stockholm Cohorts		
Signature/property	**Ki67 Cutoff Uppsala**	**Ki67 cutoff Stockholm**
Literature search	10, 20	10, 20
Median	11	12
Distribution based	16	17
GGI	15	18
70-gene	11	8
p53	16	28
Recurrence score	12*	8†
Sorlie	16§	20‡
Hu	13¥	27‡
Parker	13¥	10‡

*Cutoff for high or intermediate recurrence score; low as reference
†Cutoff for high recurrence score; others as reference
§Cutoff for luminal B subtype; others as reference
‡Cutoff for luminal B/HER2 or Basal subtypes; others as reference
¥Cutoff for luminal B or HER2 subtype; others as reference

1.4 Supplementary gene set information Probe annotations for U133A and B chips were updated with versions 2.4.1 of the hgu133a.db and 2.5.0 of hgu133b.db packages from the Bioconductor website. In cases of numerous probes for one original signature probe or probe set, an average expression over all probes was taken. Classifications of all tumors by all signatures was confirmed by two analysts. Delineation of the gene expression signature status for each tumor was performed in the following manner and we have appended a .pdf document to this publication displaying the R code used for all signature classifications:

-The genomic grade index The gene expression grade index (GGI) signature was applied as detailed in the original manuscript (Sotiriou et al., 2006). Briefly, 128 probe sets and weights corresponding to the signature were accessed from the online supplements. Probe expression designated as having increased expression in grade 1 tumors was summed and subtracted from those relevant to grade 3 tumors. Scale and offset terms were calculated such that the mean gene expression score of grade 1 tumors was -1 and grade 3 tumors, +1. Tumors were classified as low grade if their

GGI score was negative, and high grade if it was zero or positive. In addition we performed a leave-one out cross validation as per the original manuscript, meaning a patient's own data was never used to calculate their own GGI. Results were reproduced without discrepancies with the R "genefu" package (http://bioconductor.org).

-The 70-gene signature Original probe IDs were obtained from the following report prepared by The Johns Hopkins University Evidence-based Practice Center, Baltimore, MD:
http://www.ahrq.gov/downloads/pub/evidence/pdf/brcancergene/brcangene.pdf.
For 70 original probe IDs, all relevant Affymetrix U133A and U133B probe sets were identified via Entrez Gene, GenBank and UniGene identifiers. In cases of numerous probes for one original signature probe or probe set, an average expression over U133A+B probes was taken. Each of the tumors was then classified based on the correlation of expression levels of the 70 original probe identifiers with a previously determined average "good" profile (in 44 breast cancers with more than 5 years favorable outcome, retrieved at http://www.rii.com/publications/2002/vantveer.html). In line with the original study (van't Veer et al., 2002)**,** tumors with a correlation coefficient of ≥ 0.3 were classified as "good" prognosis and < 0.3 as "poor". The distribution of correlations across patients was however narrower and slightly left-shifted compared to that reported by the authors in the original data resulting in a bias towards "poor" profile calls; to make an as true as possible reproduction in the current data we therefore normalized the correlations with scale and offset terms as follows:

$$\text{normalized correlation} = (\text{correlation} \times \text{scale}) - \text{offset}$$

where scale was equal to the standard deviation of the correlations in the original data divided by the standard deviation of the correlations in our data, and offset equal to the mean of correlations in our data subtracted from the mean of correlations in the original data. Example, supplementary materials Figure SM3: The distributions of author's original (black solid), raw Uppsala (red) and normalized Uppsala (green), are shown, with the vertical dashed line indicating the 0.3 cutoff.

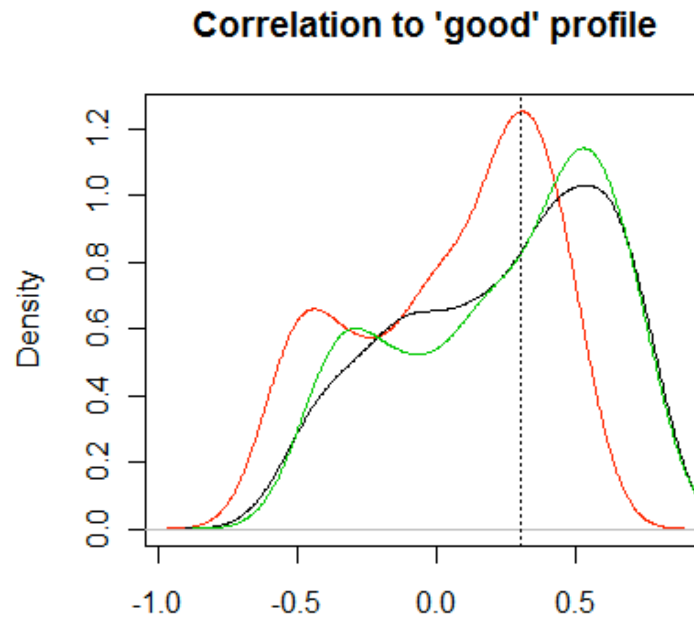

Figure SM3, Typical ROC curve

-The p53 signature As this signature was derived in the Uppsala dataset, classifications were as per the original publication (Miller et al., 2005). Classifications of the Stockholm dataset were performed as described in that manuscript. Briefly, the 32 Affymetrix probes comprising the p53 signature were selected from the Stockholm cohort U133 A and B chips. In cases of numerous probes for one original signature probe or probe set, an average expression over U133A+B probes was taken. Data was then scaled and used as the test object in a diagonal linear discrimination analysis (DLDA), performed in the R "supclust" package with the original Uppsala signature genes and p53 mutational calls as training objects.

-The recurrence score Gene lists and numerical constants for computation of recurrence score (RS) were retrieved from the original publication (Paik et al., 2004). For 16 prognostic and 5 reference genes, all relevant Affymetrix U133A and U133B probe sets were identified via Entrez Gene, GenBank and UniGene identifiers. In cases of numerous probes for one gene, an average expression over U133A+B probes was taken. Values were then transformed by scale and offset terms to achieve a range between 0 and 15 as in the original report:

$$\text{transformed expression} = (\text{raw expression} - \text{offset}) \times \text{scale}$$

where offset was equal to the minimum expression value for the relevant gene, and scale was equal to 15 divided by the maximum of *the raw expression – offset* values. The scores and resulting classifications were reproduced without discrepancies with the R genefu package (where, notably, the subtraction of a reference based on the 5 reference genes has to be done separately).

-Sorlie, Hu and PAM50 signatures: The relevant genes for each multi-level signature along with centroid values corresponding to subtypes were taken from the online supplementary material of Weigelt *et al.* (Weigelt et al., 2010). We employed the nearest centroid predictor based on Spearman correlation to classify tumors into Normal-like, Luminal A and B, Basal and ERBB2+ groups. Briefly, signature genes were extracted from Affymetrix A and B expression array sets, mean centered and probes with the same annotation were averaged. Each tumor was then classified into one subgroup based on the strongest correlation between its gene expression profile and published centroids. In the case of the Sorlie signature, tumours were classified as having "no subtype" if their closest correlation was less than 0.1. This caveat was detailed in the publication of Sorlie *et al.* (Sørlie et al., 2003), but not by the Hu or PAM50 groups, as such, we have only applied the cut-off when using the Sorlie classification scheme.

1.5 Statistical analysis Univariate, bivariate and multivariate hazard ratios were calculated using the coxph function of the survival package in R, values were confirmed in SPSS version 19. Note, that for the multivariate analysis only cases where Ki67 staining was available were included in the model, leaving 233 and 113 patients available for analysis in the Uppsala and Stockholm cohorts, respectively. Kaplan-Meier curves were plotted using a modified version of the survplot function in R with BCSS as clinical endpoint. Log-rank values for the Kaplan-Meier curves were calculated in SPSS for both binary and multi-level signatures.

Similarity in classification between Ki67 and gene signatures was determined by adding all cases with identical classification and dividing by the total number of cases. An example is shown for the GGI signature in supplementary materials Table

SM3. We add the cases in agreement (125 +64) and divide by the total number (233), giving 0.811. Multiplying by 100 gives the percentage similarity/overlap of 81%.

Table SM3

Ki67 vs. GGI classification similarity, Uppsala Cohort		
	Ki67 Low (15)	Ki67 High (15)
GGI Low (GG1)	125	25
GGI High (GG3)	19	64

Analysis of Ki67 distribution and ROC curve plotting was performed using the R mixtools and Daim packages, respectively. Full details of this are provided in section 1.3 of this document.

1.6 REMARK guidelines This manuscript was carried out and is reported in accordance with REMARK guidelines. Full details are provided in supplementary materials Table SM4.

Table SM4

REMARK criteria checklist	
Introduction	
1.	Stated marker examined, study objectives and pre-specified hypothesis
Materials and methods	
Patients	
2.	Patient characteristics for both cohorts have been described extensively in the Supplementary Methods section 1.1 and in Supplementary Tables S1 and S10. Exclusion criteria are shown in Supplementary Figure S1.
3.	The treatments received have been described in Supplementary Methods section 1.1 and shown in Supplementary Figure S1 and Supplementary Tables S1 and S10.
Specimen characteristics	
4.	The biological material used and methods of preservation and storage have been detailed in Supplementary Methods section 1.1 and in accordance with BRISQ criteria (Supplementary Materials Table SM5).
Assay Methods	
5.	We have extensively outlined all staining and scoring procedures for Ki67 in the Supplementary Materials sections 1.2 and 1.3.
Study design	
6.	We have stated: the method of case selection (consecutive cases), the time period the cases were taken from, the end of follow up date and the median follow up time. Supplementary Methods section 1.1.
7.	Clinical end points are have been clearly stated in all figures and tables,

	we use BCSS as the only endpoint.
8.	All variables have been listed in Supplementary Tables S7 and S8. As our hypothesis centers on a comparison between Ki67 and gene expression signatures, only these variables were included in bivariate models.
9.	Our sample size was determined by the number of tumours available after exclusion criteria (Supplementary Figure S1)
Statistical analysis methods	
10.	We have extensively outlined all statistical aspects of this manuscript in the Supplementary Materials sections 1.3 to 1.5.
11.	Ki67 cutoff selections have been detailed in the Supplementary Materials section 1.3.
Results	
Data	
12.	Patient flowcharts are shown in Supplementary Figure S1 for both cohorts. Numbers of patients in relation to ER status and Ki67 low/high are shown in Supplementary Tables S1, S2, S10 and S11. The number of events is shown in the Kaplan-Meier curves in Figure 1 and Supplementary Figure S2.
13.	We have reported the distributions of basic demographic characteristics, standard prognostic variables and tumour markers, along with number of cases missing in Supplementary Tables S1, S2, S10 and S11.
Analysis and presentation	
14.	The relationship of Ki67 to standard prognostic variables is shown in Supplementary Tables S2, and S11 for the Uppsala and Stockholm cohorts respectively.
15.	Univariate analysis of clinicopathological parameters is shown in Supplementary Table S7. Kaplan-Meier curves for Ki67 and gene expression signatures are shown in Figure 1 and Supplementary Figure S2 for the Uppsala and Stockholm cohorts respectively.
16.	Our key bivariate comparison of Ki67 vs. gene expression signatures has been reported with confidence intervals in Tables 2 and 4 for the Uppsala and Stockholm cohorts respectively. Moreover the hazard ratios for all Ki67 cutoffs vs. all gene expression signatures are shown in Supplementary Tables S9 and S15 for the Uppsala and Stockholm cohorts respectively.
17.	We have performed a multivariate analysis where Ki67 and all clinicopathological parameters are included in Supplementary Table S8.
18.	Internal validation of all gene expression signatures was performed by second bioinformatician and also using the genefu package in the R programming environment.
Discussion	
19.	We interpreted our results in the context of our pre-defined hypothesis and in the context of other relevant studies and added a paragraph on the study limitations.
20.	We have discussed the future implications of this work.

1.7 BRISQ guidelines Supplementary materials Table SM5 shows the "Biospecimen reporting for improved study quality" criteria for both the Uppsala and Stockholm cohort.

Table SM5

BRISQ criteria, Uppsala and Stockholm samples

Data Elements	**Information**
Biospecimen type	Primary tumour material
Anatomical Site	Breast
Clinical characteristics of patients	Pre and post- menopausal breast cancer patients
Vital State of patients	Alive at time of tumour excision
Clinical diagnosis of patients	Invasive breast cancer
Pathology diagnosis	See Supplementary Tables 1 (Uppsala) and 10 (Stockholm)
Collection mechanism	Whole tumour excision
Type of stabilization	Isopentane, on ice
Type of long-term preservation	Formalin fixation, freezing
Constitution of preservative	10% formalin
Storage temperature	-80°C (for frozen samples)
Storage duration	Uppsala: 13 yrs., Stockholm: 8 yrs.
Shipping temperature	Dry ice, -78.5 °C
Composition assessment and selection	Invasive breast cancer only

References

Benaglia, T., Chauveau, D., Hunter, D.R., Young, D.S., 2009. mixtools: An R package for analyzing finite mixture models. Journal of Statistical Software 32, 1–29.

Dowsett, M., Nielsen, T.O., A'Hern, R., Bartlett, J., Coombes, R.C., Cuzick, J., Ellis, M., Henry, N.L., Hugh, J.C., Lively, T., McShane, L., Paik, S., Penault-Llorca, F., Prudkin, L., Regan, M., Salter, J., Sotiriou, C., Smith, I.E., Viale, G., Zujewski, J.A., Hayes, D.F., 2011. Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group. J. Natl. Cancer Inst. 103, 1656–1664.

Miller, L.D., Smeds, J., George, J., Vega, V.B., Vergara, L., Ploner, A., Pawitan, Y., Hall, P., Klaar, S., Liu, E.T., Bergh, J., 2005. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. PNAS 102, 13550–13555.

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., Hiller, W., Fisher, E.R., Wickerham, D.L., Bryant, J., Wolmark, N., 2004. A Multigene Assay to Predict Recurrence of

Tamoxifen-Treated, Node-Negative Breast Cancer. New England Journal of Medicine 351, 2817–2826.

Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C.M., Lønning, P.E., Brown, P.O., Børresen-Dale, A.-L., Botstein, D., 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. PNAS 100, 8418–8423.

Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B., Desmedt, C., Larsimont, D., Cardoso, F., Peterse, H., Nuyten, D., Buyse, M., Van de Vijver, M.J., Bergh, J., Piccart, M., Delorenzi, M., 2006. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J. Natl. Cancer Inst. 98, 262–272.

Van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A.M., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T., Schreiber, G.J., Kerkhoven, R.M., Roberts, C., Linsley, P.S., Bernards, R., Friend, S.H., 2002. Gene expression profiling predicts clinical outcome of breast cancer. Nature 415, 530–536.

Weigelt, B., Mackay, A., A'hern, R., Natrajan, R., Tan, D.S.P., Dowsett, M., Ashworth, A., Reis-Filho, J.S., 2010. Breast cancer molecular profiling with single sample predictors: a retrospective analysis. Lancet Oncol. 11, 339–349.