

Data driven derivation of cutoffs from a pool of 3,030 Affymetrix arrays to stratify distinct clinical types of breast cancer

Thomas Karn · Dirk Metzler · Eugen Ruckhäberle · Lars Hanker ·
Regine Gätje · Christine Solbach · Andre Ahr · Marcus Schmidt ·
Uwe Holtrich · Manfred Kaufmann · Achim Rody

Received: 4 February 2009 / Accepted: 29 April 2009 / Published online: 20 May 2009
© Springer Science+Business Media, LLC. 2009

Abstract Pooling of microarray datasets seems to be a reasonable approach to increase sample size when a heterogeneous disease like breast cancer is concerned. Different methods for the adaption of datasets have been used in the literature. We have analyzed influences of these strategies using a pool of 3,030 Affymetrix U133A microarrays from breast cancer samples. We present data on the resulting concordance with biochemical assays of well known parameters and highlight critical pitfalls. We further propose a method for the inference of cutoff values directly from the data without prior knowledge of the true result. The cutoffs derived by this method displayed high specificity and sensitivity. Markers with a bimodal distribution like ER, PgR, and HER2 discriminate different biological subtypes of disease with distinct clinical courses. In contrast, markers displaying a continuous distribution like proliferation markers as Ki67 rather describe the composition of the mixture of cells in the tumor.

Keywords Breast cancer · Microarray · Cutoff · Distribution · Pooling · Meta-analysis · Bimodal markers

Introduction

Breast cancer is a heterogeneous disease of many different subtypes. This is one of the reasons that large cohorts of hundreds to thousands of patients are often needed to analyze treatment effects and the prognosis of specific subgroups [1–3]. In contrast, microarray datasets encompass only tens to hundreds of samples because of the expenditure and complexity of these analyses compared to standard parameters like age, tumor size, or hormone receptor status. Thus pooling of microarray datasets or meta analyses are required to enlarge samples size [4]. In the majority of cases an adaption of the raw values is necessary before pooling different datasets. To this end common methods like scaling by *Z-transformation* [5] or magnitude normalization [6] have been applied. In some studies normalization across genes has also been performed [7]. Here we analyze influences of these methods on the resulting concordance with data from biochemical assays. A previous report already demonstrated that estrogen receptor (ER) and human epidermal growth factor receptor 2 (HER2) status can be deduced from Affymetrix microarray data with high confidence [8]. However, in this former study specific cutoff values were derived and optimized by comparison with immunohistochemistry as the gold standard. In contrast, in the present study we aimed to derive the cutoff values directly from the data without prior knowledge of the biochemical status of the samples. Finally we compared the results for bimodal and continuous markers with respect to their biological impact on disease. Our results demonstrate that reliable cutoffs can be derived from the distribution of

Electronic supplementary material The online version of this article (doi:10.1007/s10549-009-0416-z) contains supplementary material, which is available to authorized users.

T. Karn (✉) · E. Ruckhäberle · L. Hanker · R. Gätje ·
C. Solbach · A. Ahr · U. Holtrich · M. Kaufmann · A. Rody
Department of Obstetrics and Gynecology, J. W. Goethe
University, Theodor-Stern-Kai 7, 60590 Frankfurt, Germany
e-mail: t.karn@em.uni-frankfurt.de

D. Metzler
Department of Biology II, Ludwig-Maximilians-University,
Munich, Germany

M. Schmidt
Department of Obstetrics and Gynecology,
Johannes Gutenberg University, Mainz, Germany

expression values in a pooled dataset of individually normalized microarrays. These cutoffs led to exceptionally high concordance with biochemical data for bimodal markers like the ER. Clinical follow up data demonstrate that they correctly identify distinct subtypes of the disease.

Materials and methods

We combined a database of $n = 3,030$ Affymetrix HG-U133A microarrays from treatment-naïve primary breast cancer samples (Table 1). We included 238 of our own samples (datasets Frankfurt, Frankfurt-2, and Frankfurt-3) which have been described previously (Ahr et al. 2002 [9], Rody et al. 2007 [10], Rody et al. 2008 [11], Ruckhäberle et al. 2008 [12], and Rody et al. 2007 [13], respectively) as well as 2792 samples from 22 different publicly available datasets (Table 1): Rotterdam [14–16], Mainz [17], TransBIG [18], Oxford-Untreated [19], London [20], London-2 [21], Oxford-Tamoxifen, Veridex-Tam [22], Stockholm [23], Uppsala [24, 25], San Francisco [26], New York [27], MDA133 [28], EORTC [29], Edinburgh [30], ExpO [31], Signapore [32], Genentech [33], Boston [34], Berlin [35], Paris [36], and Tampa [37]. For comparability, only the ProbeSets from the Affymetrix HG-U133A microarray were used from seven datasets where HG-U133+ microarrays were applied. The clinical characteristics of the patients in the different datasets are given in Table 1.

Affymetrix expression data were analyzed by using the MAS5.0 [38] algorithm of the *affy* package [39] of the Bioconductor software project [40] (<http://www.bioconductor.org/>). Subsequently data were \log_2 transformed and median centered across arrays. Further scaling was performed in two different ways: In the first method the expression values of all the genes on the array were multiplied by a scale factor S so that the magnitude (sum of the squares of the values) equals 1 (we refer to these data as “magnitude-normalized”). This method is similar to scaling by Z score transformation but the latter uses mean-centering instead of the more robust median-centering. In addition the applied magnitude-normalization is sensitive to the total feature size but this does not have an effect as long as the same number of ProbeSets is used for all samples. In a second approach, mean-centering and magnitude-normalization were first applied across arrays and subsequently also across genes in each individual dataset. To these data we refer here as “gene-normalized”.

ER status as determined by immunohistochemistry (IHC) or biochemical assay was available for 2,198 samples from 18 of the 25 datasets (see Supplementary Table S1). We further refer to this parameter as “biochemical ER status” in this manuscript. Data on PgR status were available for 1,474 patients from 13 of the 25 datasets and

HER2 status was available for only 618 patients from 8 of the 25 datasets (Supplementary Table S1). Supplementary Table S1 also gives further information on the specific methods and cutoffs used in the different studies for the definition of the ER, PgR and HER2 status. Nine different ProbeSets of the ER (ESR1) gene are present on the Affymetrix U133A array. ProbeSet 205225_at was selected for most analyses because of its highest concordance with the biochemical ER status (see Results). The progesterone receptor (PgR) is represented by only one ProbeSet (208305_at). From the two ProbeSets for HER2 which are present on the U133A array ProbeSet 216836_s_at was used (see Results). Regarding Ki67, four different ProbeSets exist on the U133A array (212020_s_at to 212023_s_at). However, there is no established cutoff for Ki67 IHC [41, 42] and a gold standard is missing. In addition all four ProbeSets display similar strong correlations to each other. Thus, we used the mean of the magnitude-normalized data of all four ProbeSets in subsequent analyses. Cutoffs for ER, PgR and HER2 expression from microarray were derived from fitting two normal distributions to the observed distribution of Affymetrix expression values by maximum likelihood optimization using the *optim* function in R as described by Venables and Ripley [43].

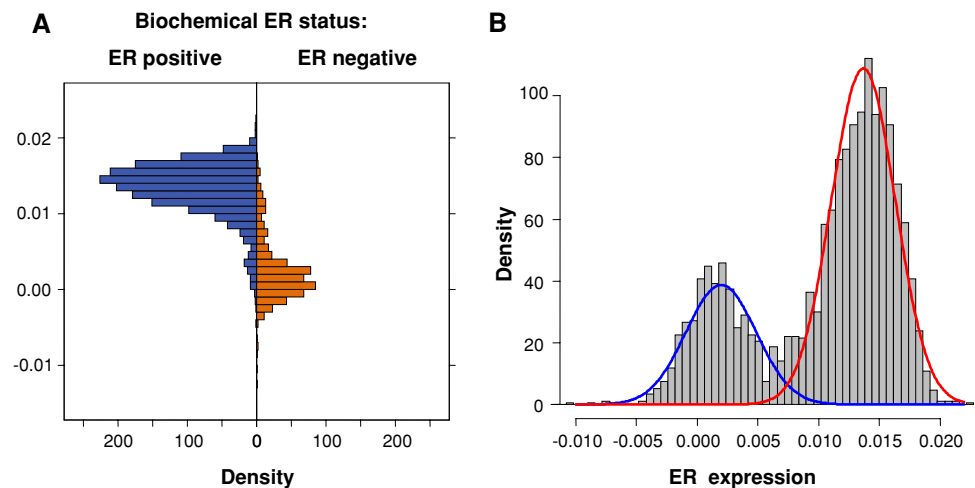
Follow up data were available for 2,058 of the samples (11 datasets without follow up, see Table 1). Survival intervals were measured from the time of surgery. For nine datasets relapse free survival (RFS) was used as an endpoint ($n = 1,180$) while for five dataset only distant metastasis free survival (DMFS) was available ($n = 879$). Thus any local recurrence events are missing from these five datasets. However, the effect of using these different endpoints was rather small in the overall dataset. Supplementary Figure S1 shows that no significant difference in relative survival was found when comparing the 879 samples where only the DMFS endpoint was available to the 1,180 samples using the RFS endpoint. Thus we used in the context of this study either the RFS endpoint as disease free survival (DFS) or the DMFS endpoint if RFS was not available. Data for women in whom the envisaged end point was not reached were censored as of the last follow-up date or at 120 months. We constructed Kaplan–Meier curves and used the log rank test to determine the univariate significance of the variables. A Cox proportional-hazards regression model was used to examine simultaneously the effects of multiple covariates on survival. The effect of each variable was assessed with the use of the Wald test and described by the hazard ratio, with a 95% confidence interval. Subjects with missing values were excluded from the analyses and all reported P values are two sided. P values of less than 0.05 were considered to indicate a significant result. All analyses were performed using the R software environment

Table 1 Summary of Affymetrix microarray datasets used in this study

| Dataset | Data source | No. of samples | % of Samples | | Systemic treatment | | | Median follow up months | No. of relapses | Event type | References |
|------------------|--|----------------|--------------|-------------------|--------------------|---------|------|-------------------------|-----------------|------------|------------|
| | | | Age ≤ 50 | Tumor size ≤ 2 cm | LNN | ER pos. | G3 | | | | |
| Rotterdam | GSE2034, GSE5327 | 344 | n.a. | n.a. | n.a. | 62 | n.a. | 286 untreated, 58 n.a. | 118 | DMFS | [14, 15] |
| Mainz | GSE11121 | 200 | 35 | 56 | 100 | 84 | 23 | Untreated | 41 | DMFS | [17] |
| TransBIG | GSE7390 | 198 | 69 | 37 | 100 | 68 | 42 | Untreated | 91 | RFS | [18] |
| Oxford-Untreated | GSE2990 (<i>n</i> = 61), GSE6532 (<i>n</i> = 8) | 69 | 44 | 64 | 100 | 71 | 41 | Untreated | 29 | RFS | [19] |
| London | GSE6532 | 87 | 6 | 35 | 33 | 98 | 23 | Endocrine | 28 | RFS | [20] |
| London-2 | GSE9195 | 77 | 5 | 44 | 53 | 96 | 41 | Endocrine | 13 | RFS | [21] |
| Oxford-Tamoxifen | GSE6532 | 109 | 14 | 34 | 64 | 98 | 19 | Endocrine | 30 | RFS | [20] |
| Veridex-Tam | GSE12093 | 136 | n.a. | n.a. | 100 | 100 | n.a. | Endocrine | 20 | DMFS | [22] |
| Frankfurt-3 | This study | 52 | 6 | 9 | 61 | 98 | 10 | Endocrine | 19 | RFS | [12] |
| Stockholm | GSE1456 | 159 | n.a. | n.a. | n.a. | 79 | 42 | Yes/no | 40 | RFS | [23] |
| Uppsala | GSE3494 (<i>n</i> = 251), GSE6232 (<i>n</i> = 5), GSE4922 (<i>n</i> = 1), GSE2990 (<i>n</i> = 1) | 258 | 22 | 51 | 65 | 80 | 22 | Yes/no | 91 | RFS | [24, 25] |
| San Francisco | E-TABM-158 | 118 | 46 | 33 | 43 | 69 | 54 | Yes/no | 36 | DMFS | [26] |
| New York | GSE2603 | 99 | 37 | 9 | 34 | 58 | n.a. | n.a. | 27 | DMFS | [27] |
| Frankfurt | This study | 119 | 55 | 50 | 56 | 66 | 47 | Chemotherapy | 29 | RFS | [10] |
| Frankfurt-2 | This study | 67 | 51 | 0 | 49 | 58 | 30 | Chemotherapy | n.a. | – | [13] |
| MDA133 | www.mdanderson.org | 133 | 41 | 9 | 30 | 63 | 58 | Chemotherapy | n.a. | – | [28] |
| EORTC | GSE1561 | 49 | n.a. | n.a. | n.a. | 57 | n.a. | Chemotherapy | n.a. | – | [29] |
| Edinburgh | GSE5462 | 116 | n.a. | n.a. | n.a. | 100 | n.a. | Endocrine | n.a. | – | [30] |
| expO | GSE2109 | 301 | 31 | 32 | 47 | 67 | 49 | n.a. | n.a. | – | [31] |
| Signapore | GSE5364 | 183 | n.a. | n.a. | n.a. | 55 | n.a. | n.a. | n.a. | – | [32] |
| Genentech | GSE12763 | 30 | n.a. | n.a. | n.a. | 70 | n.a. | n.a. | n.a. | – | [33] |
| Boston | GSE3744 | 40 | n.a. | n.a. | n.a. | 30 | 100 | n.a. | n.a. | – | [34] |
| Berlin | GSE6596 | 24 | 21 | 63 | n.a. | 67 | 46 | n.a. | n.a. | – | [35] |
| Paris | GSE13787 | 23 | n.a. | n.a. | n.a. | 0 | 100 | n.a. | n.a. | – | [36] |
| Tampa | GSE10780 | 39 | n.a. | n.a. | n.a. | 70 | n.a. | n.a. | n.a. | – | [37] |
| Total | | 3,030 | 35 | 36 | 70 | 74 | 38 | | 80 | | |

Notes: The TransBIG cohort contains independent replicate samples from 19 patients of Uppsala cohort and 22 patients of Oxford-Untreated cohorts. Affymetrix HG-U133A microarrays were applied in all studies except for datasets expO, London, London-2, Genentech, Boston, Paris, and Tampa where the identical ProbeSets from HG-U133Plus arrays were used

Fig. 1 Distribution of ER expression values in the combined dataset. **a** Distribution of ER expression values (ProbeSet 205225_at) stratified by biochemical ER status in those 2,198 samples with data from immunohistochemistry/biochemical assay. **b** Distribution of ER expression values (ProbeSet 205225_at) among all 3,030 samples



(<http://www.r-project.org/>) and SPSS version 17.0 (SPSS Inc., Chicago, IL).

Results

Concordance of different Affymetrix ProbeSets with biochemical data of ER, PgR, and HER2 status

In a first approach, arrays from different datasets were adapted using magnitude normalization (see methods) and the concordance with biochemical data for ER, PgR, and HER2 status was assessed. For 2,198 of 3,030 total samples (72.5%), data on the estrogen receptor (ER) status from immunohistochemistry (IHC) or biochemical assay were available. Of these, 1,635 (74.4%) were characterized as ER positive and 563 (25.6%) as ER negative. We used receiver operating characteristics (ROC) analysis to demonstrate the correlation of magnitude normalized data of different ProbeSets from the Affymetrix HG-U133A microarray with the biochemical ER status (Supplementary Figure S2). The area under curve (AUC) of the ROC analysis provides a quantitative value of the concordance with the biochemical data. The Affymetrix ProbeSet 205225_at displayed the highest concordance with an AUC of 0.949 (95% CI 0.938–0.960). This confirms the results of Gong et al. [8], who obtained the strongest correlation of this ProbeSet with ER status by IHC in their training set of 193 samples.

For 1,474 of 3,030 samples (48.6%), data on the progesterone receptor (PgR) status from immunohistochemistry (IHC) or biochemical assays were available. Among these, about 858 (58.2%) of them were characterized as PgR positive and 616 (41.8%) as PR negative. ROC analysis of the single ProbeSet (208305_at) for PgR on the HG-U133A array resulted in an AUC of 0.786 (95% CI 0.763–0.809; Supplementary Figure S3A).

The HER2 status of the tumor was available for 618 of the 3,030 samples (20.4%). 139 (22.5%) of them were characterized as HER2 positive (3+ IHC or FISH positive) and 479 (77.5%) as HER2 negative. Affymetrix ProbeSet 216836_s_at revealed a slightly better result in ROC analysis with an AUC of 0.856 (95% CI 0.814–0.897; Supplementary Figure S3B) than ProbeSet 210930_s_at (AUC 0.799; 95% CI 0.752–0.846). The superiority of this ProbeSet for HER2 status was also demonstrated previously by Gong and coworkers [8].

Derivation of a cutoff value for the ER status from the distribution of ER microarray data

We selected the ER Affymetrix ProbeSet 205225_at which worked best in ROC analysis for further study. Figure 1a presents the distribution of the expression values for this ProbeSet separately in ER positive and ER negative samples as defined by IHC/biochemical assay. In Fig. 1b we analyzed the combined distribution of the expression values among all 3,030 samples from the combined datasets. A mixture of two normal distributions was fitted to these data as demonstrated by the blue and red lines in Fig. 1b. Subsequently, the interception of the two fitted distributions was selected as a cutoff value (0.0075) for the definition of ER positive samples based on microarray. This cutoff resulted in a specificity of 86.7% and a sensitivity of 93.3% when compared with the biochemical ER status available for 2,198 of the samples (Table 2). The positive predictive value (PPV) was 95.3%, the negative predictive value (NPV) 81.7% and the overall accuracy 91.6%. Among the individual datasets the specificities ranged from 66.7 to 100%, sensitivities from 80.0 to 100%, the PPVs from 76.7 to 100%, and the NPVs from 55.6 to 100% (see Table 2).

We also performed the fitting on the distribution of ER expression values separately in each individual dataset.

Table 2 Concordance of ER status based on microarray and biochemical data

| Dataset | ER detection method ^a | General cutoff (0.0075) | | | | | Dataset specific cutoff (%) | | | | | |
|------------------|----------------------------------|-------------------------|---------------|---------|---------|--------------|-----------------------------|---------------|---------|---------|--------------|--------|
| | | Sensitiv. (%) | Specific. (%) | PPV (%) | NPV (%) | Accuracy (%) | Sensitiv. (%) | Specific. (%) | PPV (%) | NPV (%) | Accuracy (%) | Cutoff |
| Rotterdam | LBA, EIA, IHC | 89.0 | 83.7 | 89.4 | 83.1 | 86.9 | 90.0 | 83.0 | 89.1 | 84.2 | 87.2 | 0.0072 |
| TransBIG | IHC | 90.3 | 82.8 | 91.7 | 80.3 | 87.9 | 92.5 | 78.1 | 89.9 | 83.3 | 87.9 | 0.0051 |
| Oxford-untreated | Not given | 100.0 | 66.7 | 76.7 | 100.0 | 84.1 | 100.0 | 40.0 | 64.7 | 100.0 | 71.4 | 0.0034 |
| London | Not given | 97.7 | n.a. | 100.0 | n.a. | 97.7 | 97.7 | n.a. | 100.0 | n.a. | 97.7 | 0.0082 |
| London-2 | Not given | 94.8 | n.a. | 100.0 | n.a. | 94.8 | 89.6 | n.a. | 100.0 | n.a. | 89.6 | 0.0101 |
| Oxford-Tamoxifen | Not given | 97.2 | n.a. | 100.0 | n.a. | 97.2 | 100.0 | n.a. | 100.0 | n.a. | 100.0 | 0.0018 |
| Veridex-Tam | IHC, LBA | 99.3 | n.a. | 100.0 | n.a. | 99.3 | 99.3 | n.a. | 100.0 | n.a. | 99.3 | 0.0031 |
| Frankfurt-3 | IHC | 91.8 | n.a. | 97.8 | n.a. | 90.0 | 98.0 | n.a. | 98.0 | n.a. | 96.0 | 0.0040 |
| Uppsala | EIA | 88.8 | 88.2 | 97.9 | 55.6 | 88.8 | 84.7 | 88.2 | 97.8 | 47.6 | 85.1 | 0.0092 |
| San Francisco | Not given | 94.7 | 81.4 | 89.9 | 89.7 | 89.8 | 96.0 | 79.1 | 88.9 | 91.9 | 89.8 | 0.0064 |
| New York | Not given | 94.7 | 95.2 | 96.4 | 93.0 | 94.9 | 94.7 | 92.9 | 94.7 | 92.9 | 93.9 | 0.0066 |
| Frankfurt | IHC | 94.4 | 90.5 | 94.4 | 90.5 | 93.0 | 97.2 | 88.1 | 93.3 | 94.9 | 93.9 | 0.0056 |
| Frankfurt-2 | IHC | 89.7 | 88.5 | 92.1 | 85.2 | 89.2 | 89.7 | 88.5 | 92.1 | 85.2 | 89.2 | 0.0064 |
| MDA133 | IHC | 93.9 | 92.2 | 95.1 | 90.4 | 93.2 | 96.3 | 92.2 | 95.2 | 94.0 | 94.7 | 0.0072 |
| EORTC | IHC | 96.4 | 100.0 | 100.0 | 95.2 | 97.9 | 96.4 | 95.0 | 96.4 | 95.0 | 95.8 | 0.0067 |
| Edinburgh | IHC | 99.1 | n.a. | 100.0 | n.a. | 99.1 | 100.0 | n.a. | 100.0 | n.a. | 100.0 | 0.0061 |
| expO | Not given | 90.2 | 88.2 | 93.9 | 81.8 | 89.5 | 88.2 | 90.2 | 94.7 | 79.3 | 88.9 | 0.0080 |
| Boston | Not given | 80.0 | 100.0 | 100.0 | 88.9 | 92.3 | 73.3 | 100.0 | 100.0 | 85.7 | 89.7 | 0.0093 |
| All combined | | 93.3 | 86.7 | 95.3 | 81.7 | 91.6 | 93.4 | 84.0 | 94.4 | 81.4 | 91.0 | 0.0075 |

^a LBA ligand binding assay (≥ 10 fmol/mg), EIA enzyme immunoassay (>0.05 fmol/ μ g DNA), IHC immunohistochemistry ($\geq 10\%$ positive tumor cells)

n.a. not applicable (if all samples were ER positive by the biochemical method)

This procedure yielded only slightly different cutoff values (range 0.0018–0.0101, Table 2 and Supplementary Figure S4). When these dataset specific cutoffs were used, a somewhat lower overall specificity of 84.0% and identical sensitivity of 93.4% was obtained (Table 2). Thus the differences between the individual datasets are small and the simultaneous use of all samples seems to improve the fitting to the distribution.

Derivation of cutoff values for PgR and HER2 microarray data

The same method of fitting two normal distributions to the expression data of the combined sample cohort was applied to identify cutoff values for the expression of the progesterone receptor gene (PgR, ProbeSet 208305_at) and HER2 (ProbeSet 216836_s_at). The corresponding graphs are given in Fig. 2a, b, respectively. The resulting cutoff (−0.0078) from Fig. 2a for PgR expression corresponded to an overall accuracy of 71.8%, a specificity of 67.4% and a sensitivity of 74.9%. The positive predictive value (PPV) was 76.2% and the negative predictive value (NPV) 65.9%.

Again, as shown in Table 3 fitting separately each dataset Supplementary Figure S5) resulted in similar cutoffs (range −0.0099 to −0.0047) and an identical overall accuracy. The HER2 cutoff (0.0135) from Fig. 2b led to an accuracy of 89.2%, a specificity of 97.9% but a rather low sensitivity of 59.0% when compared to HER2 status based on “3+” staining in immunohistochemistry or FISH ratio >2.0 (Table 4). The PPV was 89.1% and the NPV 89.2%. Similar cutoffs (range 0.0119–0.0146) were obtained when datasets were fitted separately (Supplementary Figure S6). In contrast, using either cutoff values the sensitivity for HER2 detection differed markedly between datasets (range 32–100%, Table 4).

Influence of gene normalization on different cohorts

Some analyses of microarray datasets have used “gene normalization” to bring the data to a uniform scale. By this method the expression values of each gene are adjusted across all samples of the respective cohort. We analyzed the effect of this transformation on the distribution of ER expression values in the individual datasets. After “gene

Fig. 2 Distribution of PgR and HER2 Affymetrix expression values in the combined dataset. **a** Distribution of PgR expression values (ProbeSet 208305_at) among all 3,030 samples. **b** Distribution of HER2 expression values (ProbeSet 216836_s_at) among all 3,030 samples

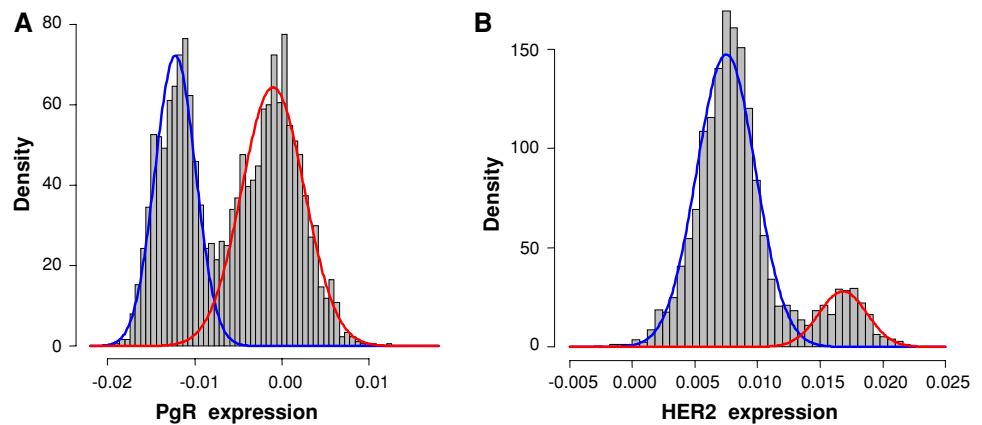


Table 3 Concordance of PgR status based on microarray and biochemical data

| Dataset | PgR detection method | General cutoff (−0.0078) | | | | | Dataset specific cutoff | | | | | |
|---------------|----------------------|--------------------------|---------------|---------|---------|--------------|-------------------------|---------------|---------|---------|--------------|---------|
| | | Sensitiv. (%) | Specific. (%) | PPV (%) | NPV (%) | Accuracy (%) | Sensitiv. (%) | Specific. (%) | PPV (%) | NPV (%) | Accuracy (%) | Cutoff |
| Rotterdam | LBA, EIA | 68.3 | 80.4 | 86.2 | 58.7 | 72.7 | 68.3 | 80.4 | 86.2 | 58.7 | 72.7 | −0.0075 |
| London | Not given | 82.8 | 42.9 | 81.5 | 45.0 | 72.9 | 82.8 | 42.9 | 81.5 | 45.0 | 72.9 | −0.0088 |
| London-2 | Not given | 94.9 | 66.7 | 90.3 | 80.0 | 88.3 | 89.8 | 77.8 | 93.0 | 70.0 | 87.0 | −0.0064 |
| Frankfurt-3 | IHC | 80.8 | 50.0 | 67.7 | 66.7 | 67.4 | 80.8 | 45.0 | 65.6 | 64.3 | 65.2 | −0.0095 |
| Uppsala | EIA | 68.4 | 73.8 | 89.0 | 42.9 | 69.7 | 68.9 | 73.8 | 89.1 | 43.3 | 70.1 | −0.0084 |
| San Francisco | Not given | 75.8 | 72.5 | 78.1 | 69.8 | 74.4 | 78.8 | 66.7 | 75.4 | 70.8 | 73.5 | −0.0099 |
| New York | Not given | 67.4 | 90.9 | 85.3 | 78.1 | 80.6 | 67.4 | 90.9 | 85.3 | 78.1 | 80.6 | −0.0069 |
| Frankfurt | IHC | 77.4 | 64.4 | 66.1 | 76.0 | 70.5 | 77.4 | 69.5 | 69.5 | 77.4 | 73.2 | −0.0064 |
| Frankfurt-2 | IHC | 72.4 | 77.8 | 72.4 | 77.8 | 75.4 | 72.4 | 77.8 | 72.4 | 77.8 | 75.4 | −0.0081 |
| MDA133 | IHC | 72.7 | 68.0 | 62.5 | 77.3 | 70.0 | 72.7 | 66.7 | 61.5 | 76.9 | 69.2 | −0.0079 |
| EORTC | IHC | 66.7 | 93.1 | 85.7 | 81.8 | 83.0 | 61.1 | 93.1 | 84.6 | 79.4 | 80.9 | −0.0054 |
| expO | Not given | 91.0 | 19.2 | 54.6 | 66.7 | 56.3 | 91.0 | 19.2 | 54.6 | 66.7 | 56.3 | −0.0076 |
| Boston | Not given | 53.8 | 76.9 | 53.8 | 76.9 | 69.2 | 30.8 | 92.3 | 66.7 | 72.7 | 71.8 | −0.0047 |
| Combined | | 74.9 | 67.4 | 76.2 | 65.9 | 71.8 | 74.5 | 68.0 | 76.4 | 65.7 | 71.8 | −0.0078 |

LBA ligand binding assay (≥ 10 fmol/mg), EIA enzyme immunoassay (>0.05 fmol/ μ g DNA), IHC immunohistochemistry ($\geq 10\%$ positive tumor cells)

normalization” has been performed it was still possible to derive cutoff values from the mixed distribution as described above. However, the specific cutoffs are different for each individual dataset after “gene normalization” since they depend on the proportions of ER positive and ER negative samples in the specific dataset. We analyzed the impact of this effect by deliberately subdividing the dataset Frankfurt in two subgroups containing either only the ER positive or the ER negative samples. Supplementary Figure S7 demonstrates the influence of gene normalization on the full cohort (Supplementary Figure S7A) and the two subcohorts (Supplementary Figure S7B, C). “Gene normalization” leads to a broadening of the distribution of expression values in the ER positive and the ER negative

subsets as compared to the full cohort (Supplementary Figure S7). Importantly, the derivation of a cutoff from such gene normalized data by fitting two distributions was only possible when at least some ER negative samples were enclosed in the ER positive cohort and vice versa.

Plausibility of the derived cutoffs through analysis of patients’ prognosis

ER positive and ER negative breast cancers are generally assumed as separate types of disease with a different clinical course [44–46]. We reasoned that a correct classification according to ER status should result in patient subgroups with a distinct prognosis while further subdivision

Table 4 Concordance of HER2 status based on microarray and biochemical data

| Dataset | IHC/FISH HER2 available | HER2 positive ^a (%) | General cutoff (0.0135) | | | | | Dataset specific cutoff | | | | | |
|---------------|-------------------------------|--------------------------------------|-------------------------|------------------|------------|------------|-----------------|-------------------------|------------------|------------|------------|-----------------|--------|
| | | | Sensitiv. (%) | Specific. (%) | PPV (%) | NPV (%) | Accuracy (%) | Sensitiv. (%) | Specific. (%) | PPV (%) | NPV (%) | Accuracy (%) | Cutoff |
| Frankfurt-3 | 19 | 2 (11%) | 0.0 | 94.1 | 0.0 | 88.9 | 84.2 | 0.0 | 94.1 | 0.0 | 88.9 | 84.2 | 0.0163 |
| San Francisco | 79 | 8 (10%) | 100.0 | 98.6 | 88.9 | 100.0 | 98.7 | 100.0 | 98.6 | 88.9 | 100.0 | 98.7 | 0.0119 |
| New York | 88 | 9 (10%) | 77.8 | 94.9 | 63.6 | 97.4 | 93.2 | 77.8 | 94.9 | 63.6 | 97.4 | 93.2 | 0.0151 |
| Frankfurt | 65 | 22 (34%) | 59.1 | 95.3 | 86.7 | 82.0 | 83.1 | 50.0 | 97.7 | 91.7 | 79.2 | 81.5 | 0.0146 |
| Frankfurt-2 | 57 | 20 (35%) | 35.0 | 97.3 | 87.5 | 73.5 | 75.4 | 35.0 | 97.3 | 87.5 | 73.5 | 75.4 | 0.0145 |
| MDA133 | 132 | 33 (25%) | 81.8 | 99.0 | 96.4 | 94.2 | 94.7 | 75.8 | 99.0 | 96.2 | 92.5 | 93.2 | 0.0144 |
| expO | 141 | 37 (26%) | 35.1 | 100.0 | 100.0 | 81.3 | 83.0 | 32.4 | 100.0 | 100.0 | 80.6 | 82.3 | 0.0141 |
| Boston | 37 | 8 (22%) | 87.5 | 100.0 | 100.0 | 96.7 | 97.3 | 87.5 | 100.0 | 100.0 | 96.7 | 97.3 | 0.0155 |
| Combined | 618 | 139 (22%) | 59.0 | 97.9 | 89.1 | 89.2 | 89.2 | 55.4 | 98.1 | 89.5 | 88.3 | 88.5 | 0.0135 |

^a IHC 3+ OR FISH > 2.0 if method given, see Supplementary Table S1

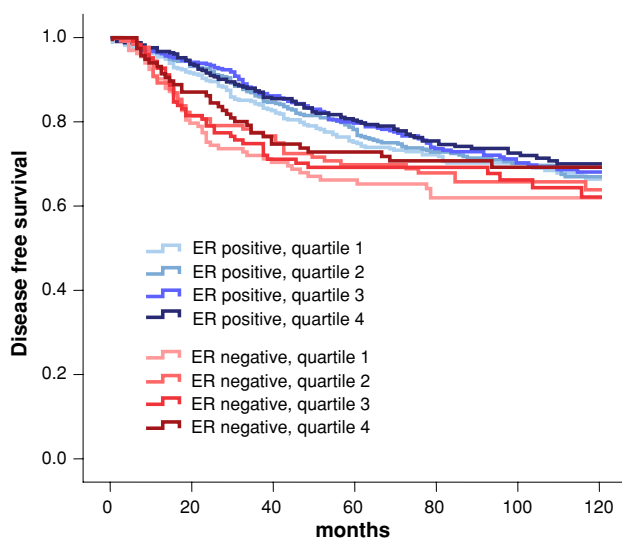


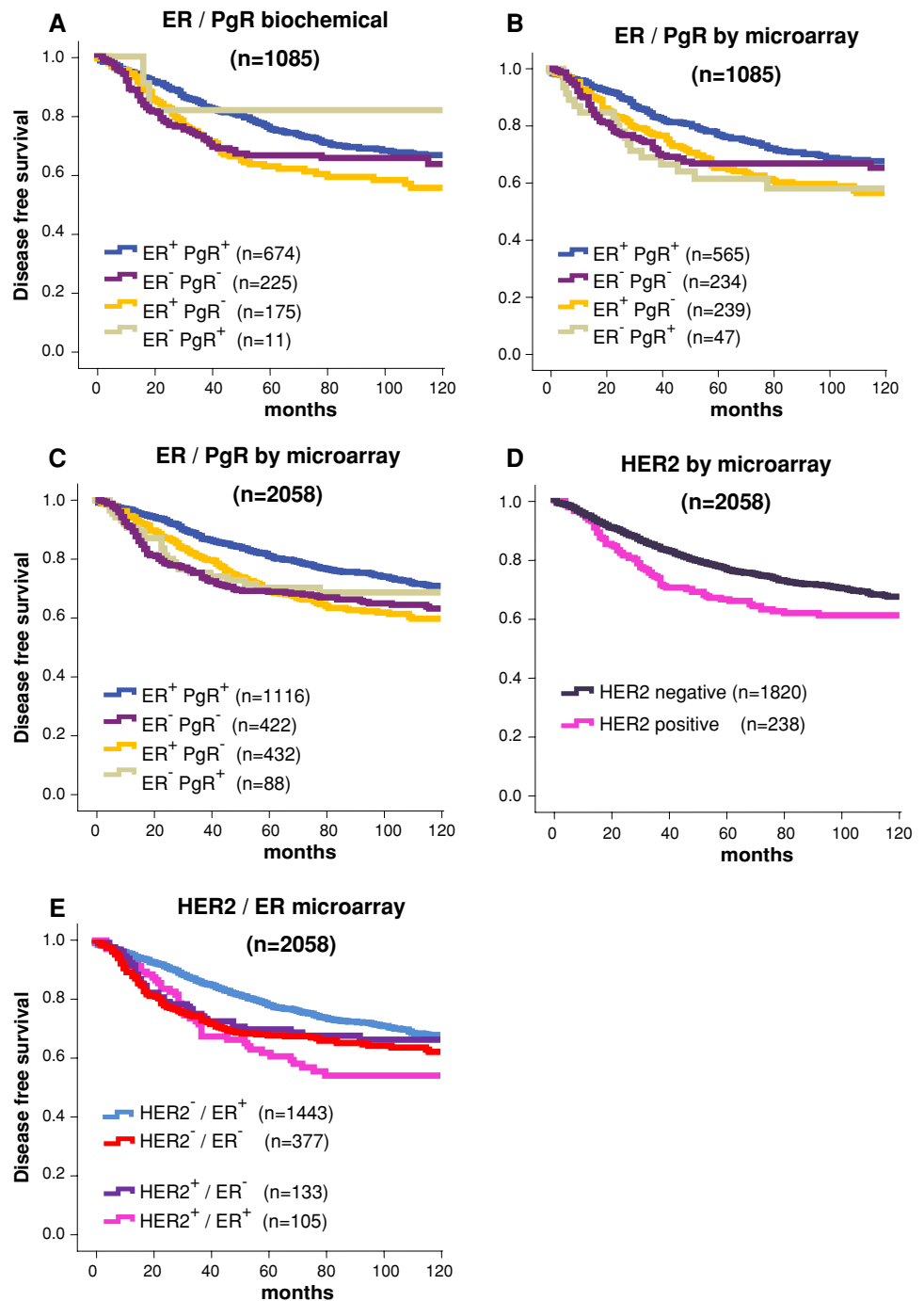
Fig. 3 Different prognosis of patients supports plausibility of the derived cutoff for ER expression. Samples were defined as ER positive and ER negative based on microarray. The two subgroups of patients were further stratified into quartiles based on the expression values of ER. Kaplan Meier analysis of disease free survival of 2,058 samples with follow up data is presented for the resulting eight subgroups. While ER positive and negative subtypes clearly differ in survival ($P = 0.001$) no correlation of the prognosis with the relative expression of ER within the subgroups of ER positive and ER negative tumors is detectable

according to the level of ER expression should have no significant effect on survival. Thus we stratified both the microarray-derived ER positive and ER negative subgroups in four quartiles each according to ER expression (ProbeSet 205225_at). Follow up data were available for 2,058 of the samples. Figure 3 presents the results from Kaplan Meier analyses of disease free survival of the patients in the eight resulting subgroups. Patients in the ER negative

subgroup as defined by the distribution-derived cutoff had a high risk for a relapse especially in the first three to five years. Survival in the ER positive subgroup is significantly better but steadily declining even after five to ten years ($P = 0.001$). These differences between ER negative and ER positive cancer types has been repeatedly described previously [44, 45]. In contrast we observed no significant differences in survival among each subtype when the patients were further substratified according to ER expression.

Next we analyzed the impact of the combined stratification according to both ER and PgR determined by either biochemical methods (Fig. 4a) or microarray (Fig. 4b) on the prognosis of the 1,085 patients for which the biochemical data as well as follow up information were available. The obtained results were similar and detailed results for single markers are also presented in Supplementary Figure S8. Microarray data resulted in a higher portion (4%) of ER negative PgR positive tumors than biochemical methods (1%) which might represent false positive PgR results (see Supplementary Figure S10 and the section Discussion). Figure 4c–e present the results of Kaplan–Meier analyses in which all 2,058 patients with available follow up data were included using the distribution derived cutoffs described above. Results for ER and PgR (Fig. 4c) were comparable to those of the smaller subset in Fig. 4b. HER2 positive patients had a worse prognosis in the complete cohort (Fig. 4d). As shown in Fig. 4e, the largest impact of HER2 expression was observed in the ER positive subgroup. To analyze the relative impact of the three variables (ER, PgR, and HER2) on the prognosis of patients simultaneously we performed univariate and multivariate Cox regression analysis as presented in Supplementary Table 2. While all three markers were highly significant in univariate analysis, only

Fig. 4 Disease free survival of patients according to stratifications using the derived cutoff values. **a, b** Samples with available biochemical status of ER and PgR ($n = 1,085$) were stratified according to either the biochemical status (**a**) or the microarray derived status (**b**). Disease free survival of the respective subgroups according to Kaplan–Meier analysis is presented. Detailed individual comparisons are given in Supplementary Figure S8. **c–e** All samples with available follow up information ($n = 2,058$) were stratified according to microarray derived status for ER and PgR (**c**), HER2 status (**d**), as well as HER2 and ER status (**e**)



PgR remained significant with a hazard ratio of 1.48 (95% CI 1.23–1.78, $P < 0.001$, Supplementary Table 2). For a subset of 1,589 of the patients in the analyzed cohorts information on endocrine treatment was available. As shown in Fig. 5, the worse prognosis of PgR negative tumors was observed among both 661 endocrine treated and 928 untreated samples (results were similar for the subset of 722 patients with biochemical status data; Supplementary Figure S9).

Analysis of continuous markers

The markers analyzed so far demonstrated bimodal distributions. Both for ER and PgR as well as HER2, we observed two clearly different subgroups of samples in the cohorts. These results are in line with the widely accepted concept that these subgroups characterize biologically distinct subtypes of breast cancer [47, 48]. In contrast Ki67 as well as other proliferation markers represent a different

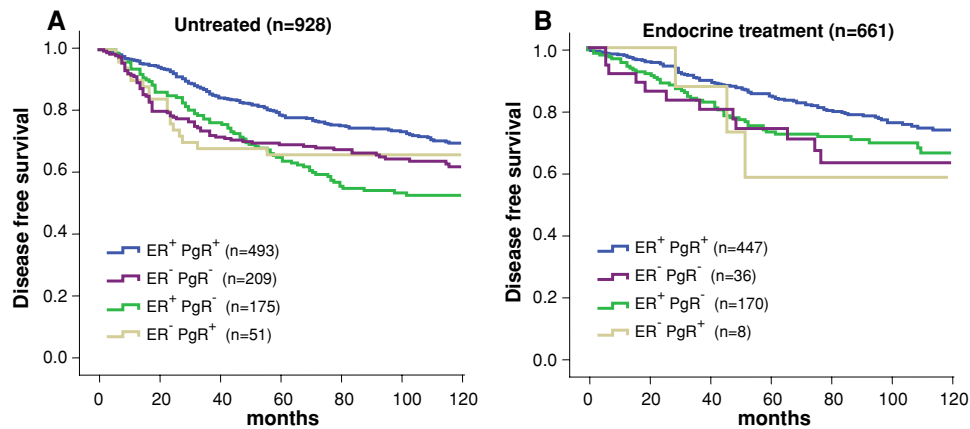


Fig. 5 Disease free survival of untreated and endocrine treated patients stratified according to ER and PgR status based on microarray. The 1,589 patients with available follow up information were selected which were either untreated (**a**) or treated only with

adjuvant endocrine therapy (**b**). Patients were stratified according to ER and PgR status based on microarray and Kaplan–Meier analysis performed to determine the disease free survival in the respective subgroups

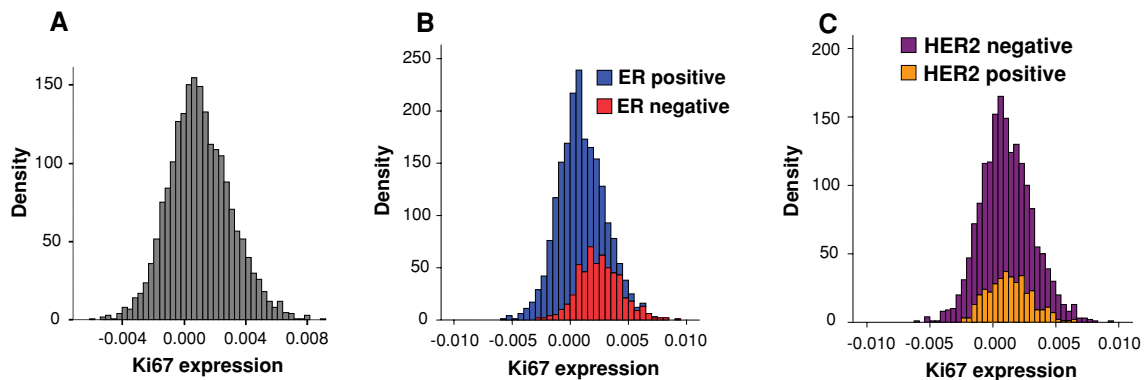


Fig. 6 Distribution of Ki67 Affymetrix expression values in the combined dataset. **a** Distribution of Ki67 expression values (ProbeSets 212020_s_at–212023_s_at) among all 3,030 samples. **b** Distribution of

Ki67 expression values stratified according to ER status among all 3,030 samples. **c** Distribution of Ki67 expression values stratified according to HER2 among all 3,030 samples

type of parameter. The observed distribution of Ki67 expression among the samples is not bimodal but rather continuous as shown in Fig. 6a. Thus the approach used above for bimodal distributions does not seem to be appropriate for Ki67. The continuous distribution of Ki67 expression might suggest that in contrast to the ER status those tumors with high and low Ki67 expression values, respectively, does not represent distinct types of disease. The level of Ki67 expression could rather be a surrogate marker for the proportion of Ki67 expressing cells in the tumor sample and display a quantitative correlation with prognosis. Consequently a multiple substratification according to Ki67 expression should result in multiple groups with a different clinical prognosis contrary to the results obtained for the ER above in Fig. 3. As shown in Fig. 6b, ER negative breast cancers are generally characterized by high expression of Ki67 while the influence of the HER2 status on the distribution of Ki67 expression

values was not so profound as demonstrated in Fig. 6c. Thus to avoid a confounding effect ER positive tumors need to be analyzed separately from ER negative tumors. We therefore performed a quartile split according to Ki67 expression in the ER positive cohort. As shown in Fig. 7, Kaplan Meier analysis of the four groups of ER positive tumors suggests that the higher the level of Ki67 the worse the prognosis of the patients.

Discussion

In this study we used a data driven model for the definition of cutoffs without a priori information on the true result (e.g., the biochemical ER status). We derived the cutoff by fitting a mixture of two normal distributions to the data as the simplest approach. Obviously this simple model does not need to be correct or even a good model at all.

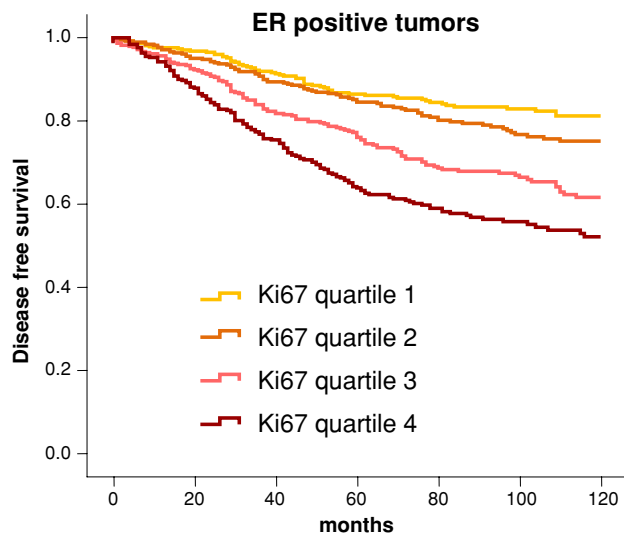


Fig. 7 Disease free survival of patients with ER positive tumors according to the level of Ki67 expression. Patients with ER positive tumors were stratified in quartiles according to the level of Ki67 expression by microarray. Kaplan–Meier analyses of disease free survival of the 1,549 patients in the respective four groups are given

However, we used this straight approach to avoid overfitting of the data. In contrast to previous studies [8] no external gold standard is necessary to derive a cutoff by this method. Nevertheless, the derived cutoffs for ER, PgR and HER2 showed high concordance with results from biochemical methods despite, e.g., widespread concerns of the inaccuracy of immunohistochemistry [49, 50]. The comparability of array normalized data from 25 different datasets was good leading to similar cutoffs. When individual datasets were used as “training” set similar accuracies were obtained in the resulting validation sets (Supplementary Figure 11). Our observed accuracy of ER (91.6%, Table 2) was higher than that between local and central IHC in a recent study from the ECOG 2,197 trial (90%) [51]. In the same study the PgR sensitivity and specificity between local and central IHC were 80.7 and 88.6%, respectively. Concordance of PgR IHC between core biopsy and surgical samples has been reported to be 83% [52]. Thus the observed concordance of 71% for PgR status in our study can not be considered satisfying (75% sensitivity and 67% specificity, Table 3). However, similar results for gene expression have been obtained by others [53] and might be related to substantially lower mRNA expression levels for PgR than ER (>20 fold lower mean MAS5 values, see Supplementary Figure S10). This problem might be overcome by using signatures combining several genes. Creighton et al. [16] have developed two gene signatures consisting of 182 and 1,005 genes for ER +/PR + and ER-/PR-tumors, respectively, using the Rotterdam and Uppsala datasets. When they applied these

signatures to classify the tumors from the same datasets they reached concordances of 89 and 84%, respectively.

The derived cutoff for HER2 expression led to a specificity of 97.9% but a rather low sensitivity of 59.0% among the 618 samples with biochemical HER2 status. However, the sensitivity differed markedly between datasets (range 32–100%, Table 4) and the possibility of some false positive biochemical data might be considered. Gong et al. [8] used a training approach to optimize the HER2 cutoff among 195 samples leading to 91% sensitivity and 95% specificity. This cutoff resulted in sensitivity and specificity of 79 and 94% as well as 100 and 88% in two different validation datasets. Microarray data of 133 of the 195 samples from this training set were available for our analysis (dataset MDA133). Importantly, among these 133 samples our distribution derived cutoff resulted in an 82% sensitivity, 99% specificity, and 95% overall accuracy. In other words our method resulted in a slightly higher cutoff value than the training approach of Gong et al. [8]. However, with respect to an analytical approach rather than a clinical test, this high cutoff value might be more preferable. It results in a specificity of 97.9% and a positive predictive value (PPV) of 89.1% among all 618 samples with biochemical HER2 status. Thus only 10.1% false positive HER2 tumors would be included in subsequent analyses when using this cutoff value. On the other hand because of the NPV of 89.2% those samples erroneously categorized as HER2 negative by this cutoff represent only 11.8% of the total number of samples in the larger group of HER2 negative tumors.

Some microarray studies performed normalization across genes [54]. However, the expression of many genes is highly correlated. For example it has been shown repeatedly that a large set of genes is strongly associated with the ER status in breast cancer [55, 56]. As we have demonstrated, “gene normalization” in cohorts with varying proportions of tumors differing in the ER status leads to a distortion of the distribution of expression values of such genes. If subsequent analysis steps involve a relative split of the cohort to stratify samples, this can lead to strange results [57]. Thus gene normalization is a very critical point when combining or comparing datasets.

Other studies have also characterized bimodal markers from microarray datasets [58–60]. Some suggested that bimodal “switch-like” genes differ from non-bimodal genes in transcriptional regulation [61]. Assuming a precise quantitation by microarrays, the different mRNA levels can either result from the level in individual cells or from the proportion of cells in the sample (or both). Breast cancer is a heterogeneous disease containing subtypes with different clinical behavior. It has been suggested that such distinct cancer subtypes may be derived from distinct progenitor cells which are arrested in their maturation [62–66]. When

we analyzed the prognosis of subgroups we have observed an essential difference between bimodal parameters like ER, PgR, HER2 on the one hand and the continuous marker Ki67 on the other. The bimodal markers seem to stratify distinct subtypes of tumors as revealed by their distinct clinical follow up. In contrast Ki67 did not define two distinct subtypes. Instead we observed a continuous relationship the higher the expression the worse the prognosis. From immunohistochemical studies it is known that the proportion of Ki67 expressing cells is relatively low with a median of 16–17% [41, 42]. It is not clear whether this represents a snapshot of transiently cycling cells or if Ki67 expression defines a distinct type of carcinoma cells which differ in their differentiation state. Regardless of this question it seems reasonable that the level of Ki67 mRNA measured by microarray predominantly results from the proportion of cells expressing the gene rather than the level of expression in the individual cells.

In summary our data demonstrate that pooling of microarray datasets seems to be recommended to enlarge sample size and to refine cutoffs derived from the data. Critical pitfalls which have to be considered include the introduction of bias from gene normalization which has been often applied to adjust different platforms.

Acknowledgments We thank Samira Adel and Katherina Kourtis for expert technical assistance and anonymous reviewers for their insightful suggestions. This work was supported by grants from the Deutsche Krebshilfe, the Margarete Bonifer-Stiftung, Bad Soden, the BANSS-Stiftung, Biedenkopf, and the Dr. Robert Pflieger-Stiftung, Bamberg. The efforts of the IGC and expO [31] are gratefully acknowledged.

References

- Coates A, Goldhirsch A, Gelber R, International Breast Cancer Study Group (2002) Overhauling the breast cancer overview: are subsets subversive? *Lancet Oncol* 3(9):525–526. doi:10.1016/S1470-2045(02)00842-2
- Cole BF, Gelber RD, Gelber S, Coates AS, Goldhirsch A (2001) Polychemotherapy for early breast cancer: an overview of the randomised clinical trials with quality-adjusted survival analysis. *Lancet* 358(9278):277–286. doi:10.1016/S0140-6736(01)05483-6
- Sylvester R, Collette L, Duchateau L (2000) The role of meta-analyses in assessing cancer treatments. *Eur J Cancer* 36(11):1351–1358. doi:10.1016/S0959-8049(00)00125-8
- Ein-Dor L, Zuk O, Domany E (2006) Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc Natl Acad Sci USA* 103(15):5923–5928. doi:10.1073/pnas.0601231103
- Cheadle C, Vawter MP, Freed WJ, Becker KG (2003) Analysis of microarray data using Z score transformation. *J Mol Diagn* 5(2):73–81
- Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95(25):14863–14868. doi:10.1073/pnas.95.25.14863
- Getz G, Levine E, Domany E (2000) Coupled two-way clustering analysis of gene microarray data. *Proc Natl Acad Sci USA* 97(22):12079–12084. doi:10.1073/pnas.210134797
- Gong Y, Yan K, Lin F, Anderson K, Sotiriou C, Andre F, Holmes FA, Valero V, Booser D, Pippin JE Jr, Vukelja S, Gomez H, Mejia J, Barajas LJ, Hess KR, Sneige N, Hortobagyi GN, Pusztai L, Symmans WF (2007) Determination of oestrogen-receptor status and ERBB2 status of breast carcinoma: a gene-expression profiling study. *Lancet Oncol* 8(3):203–211. doi:10.1016/S1470-2045(07)70042-6
- Ahr A, Karn T, Solbach C, Seiter T, Strebhardt K, Holtrich U, Kaufmann M (2002) Identification of high risk breast-cancer patients by gene expression profiling. *Lancet* 359(9301):131–132. doi:10.1016/S0140-6736(02)07337-3
- Rody A, Holtrich U, Gaetje R, Gehrmann M, Engels K, von Minckwitz G, Loibl S, Diallo-Danebrock R, Ruckhäberle E, Metzler D, Ahr A, Solbach C, Karn T, Kaufmann M (2007) Poor outcome in estrogen receptor-positive breast cancers predicted by loss of Plexin B1. *Clin Cancer Res* 13(4):1115–1122. doi:10.1158/1078-0432.CCR-06-2433
- Rody A, Karn T, Ruckhäberle E, Hanker L, Metzler D, Müller V, Solbach C, Ahr A, Gätje R, Holtrich U, Kaufmann M (2009) Loss of Plexin B1 is highly prognostic in low proliferating ER positive breast cancers—results of a large scale microarray analysis. *Eur J Cancer* 45(3):405–413
- Ruckhäberle E, Rody A, Engels K, Gaetje R, von Minckwitz G, Schiffmann S, Grösch S, Geisslinger G, Holtrich U, Karn T, Kaufmann M (2008) Microarray analysis of altered sphingolipid metabolism reveals prognostic significance of sphingosine kinase 1 in breast cancer. *Breast Cancer Res Treat* 112(1):41–52. doi:10.1007/s10549-007-9836-9
- Rody A, Karn T, Solbach C, Gaetje R, Munnes M, Kissler S, Ruckhäberle E, Minckwitz GV, Loibl S, Holtrich U, Kaufmann M (2007) The erbB2+ cluster of the intrinsic gene set predicts tumor response of breast cancer patients receiving neoadjuvant chemotherapy with docetaxel, doxorubicin and cyclophosphamide within the GEPARTRIO trial. *Breast* 16(3):235–240. doi:10.1016/j.breast.2007.02.006
- Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EM, Atkins D, Foekens JA (2005) Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* 365(9460):671–679
- Minn AJ, Gupta GP, Padua D, Bos P, Nguyen DX, Nuyten D, Kreike B, Zhang Y, Wang Y, Ishwaran H, Foekens JA, van de Vijver M, Massagué J (2007) Lung metastasis genes couple breast tumor size and metastatic spread. *Proc Natl Acad Sci USA* 104(16):6740–6745. doi:10.1073/pnas.0701138104
- Creighton CJ, Kent Osborne C, van de Vijver MJ, Foekens JA, Klijn JG, Horlings HM, Nuyten D, Wang Y, Zhang Y, Chamness GC, Hilsenbeck SG, Lee AV, Schiff R (2009) Molecular profiles of progesterone receptor loss in human breast tumors. *Breast Cancer Res Treat* 114(2):287–299. doi:10.1007/s10549-008-0017-2
- Schmidt M, Böhm D, von Törne C, Steiner E, Puhl A, Pilch H, Lehr HA, Hengstler JG, Kölbl H, Gehrmann M (2008) The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res* 68(13):5405–5413. doi:10.1158/0008-5472.CAN-07-5206
- Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, Viale G, Delorenzi M, Zhang Y, d'Assignies, Bergh J, Lide-reau R, Ellis P, Harris AL, Klijn JG, Foekens JA, Cardoso F, Piccart MJ, Buyse M, Sotiriou C (2007) TRANSBIG Consortium. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multi-center independent validation series. *Clin Cancer Res* 13(11):3207–3214

19. 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 MJ, 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(4): 262–272
20. Loi S, Haibe-Kains B, Desmedt C, Lallemand F, Tutt AM, Gillet C, Ellis P, Harris A, Bergh J, Foekens JA, Klijn JG, Larsimont D, Buyse M, Bontempi G, Delorenzi M, Piccart MJ, Sotiriou C (2007) Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* 25(10):1239–1246. doi:10.1200/JCO.2006.07.1522
21. Loi S, Haibe-Kains B, Desmedt C, Wirapati P, Lallemand F, Tutt AM, Gillet C, Ellis P, Harris A, Reid JF, Daidone MG, Pierotti MA, Berns EM, Jansen MP, Foekens JA, Delorenzi M, Bontempi G, Piccart MJ, Sotiriou C (2008) Predicting prognosis using molecular profiling in estrogen receptor-positive breast cancer treated with tamoxifen. *BMC Genomics* 9:239. doi:10.1186/1471-2164-9-239
22. Zhang Y, Sieuwerts AM, McGreevy M, Casey G, Cufer T, Paradiso A, Harbeck N, Span PN, Hicks DG, Crowe J, Tubbs RR, Budd GT, Lyons J, Sweep FC, Schmitt M, Schittulli F, Golouh R, Talantov D, Wang Y, Foekens JA (2009) The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. *Breast Cancer Res Treat*. doi:10.1007/s10549-008-0183-2
23. Pawitan Y, Bjohle J, Amler L, Borg AL, Eghazi S, Hall P, Han X, Holmberg L, Huang F, Klaar S, Liu ET, Miller L, Nordgren H, Ploner A, Sandelin K, Shaw PM, Smeds J, Skoog L, Wedren S, Bergh J (2005) Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast Cancer Res* 7(6):R953–R964. doi:10.1186/bcr1325
24. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, Pawitan Y, Hall P, Klaar S, Liu ET, Bergh J (2005) An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci USA* 102(38):13550–13555. doi:10.1073/pnas.0506230102
25. Ivshina AV, George J, Senko O, Mow B, Putti T, Smeds J, Lindahl T, Pawitan Y, Hall P, Nordgren H, Wong John EL, Liu ET, Bergh J, Kuznetsov VA, Miller LD (2006) Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* 66(21):10292–10301. doi:10.1158/0008-5472.CAN-05-4414
26. Chin K, DeVries S, Fridlyand J, Spellman PT, Roydasgupta R, Kuo WL, Lapuk A, Neve RM, Qian Z, Ryder T, Chen F, Feiler H, Tokuyasu T, Kingsley C, Dairkee S, Meng Z, Chew K, Pinkel D, Jain A, Ljung BM, Esserman L, Albertson DG, Waldman FM, Gray JW (2006) Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* 10(6):529–541. doi:10.1016/j.ccr.2006.10.009
27. Minn AJ, Gupta GP, Siegel PM, Bos PD, Shu W, Giri DD, Viale A, Olshen AB, Gerald WL, Massagué J (2005) Genes that mediate breast cancer metastasis to lung. *Nature* 436(7050):518–524. doi:10.1038/nature03799
28. Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, Mejia JA, Booser D, Theriault RL, Buzdar AU, Dempsey PJ, Rouzier R, Sneige N, Ross JS, Vidaurre T, Gómez HL, Hortobagyi GN, Puztai L (2006) Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* 24(26):4236–4244. doi:10.1200/JCO.2006.05.6861
29. Farmer P, Bonnefoi H, Becette V, Tubiana-Hulin M, Fumoleau P, Larsimont D, Macgrogan G, Bergh J, Cameron D, Goldstein D, Duss S, Nicolaz AL, Brisken C, Fiche M, Delorenzi M, Iggo R (2005) Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* 24(29):4660–4671. doi:10.1038/sj.onc.1208561
30. Miller WR, Larionov AA, Renshaw L, Anderson TJ, White S, Murray J, Murray E, Hampton G, Walker JR, Ho S, Krause A, Evans DB, Dixon JM (2007) Changes in breast cancer transcriptional profiles after treatment with the aromatase inhibitor, letrozole. *Pharmacogenet Genom* 17(10):813–826. doi:10.1097/FPC.0b013e32820b853a
31. The International Genomics Consortium (IGC) The expO project (Expression Project For Oncology) <http://www.intgen.org/>
32. Yu K, Ganesan K, Tan LK, Laban M, Wu J, Zhao XD, Li H, Leung CH, Zhu Y, Wei CL, Hooi SC, Miller L, Tan P (2008) A precisely regulated gene expression cassette potentially modulates metastasis and survival in multiple solid cancers. *PLoS Genet* 4(7):e1000129. doi:10.1371/journal.pgen.1000129
33. Januario T, Lackner MR (2009) Gene expression profiling of 30 human breast cancers. *Gene Expression Omnibus Series GSE12763*. <http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12763>
34. Richardson AL, Wang ZC, De Nicolo A, Lu X, Brown M, Miron A, Liao X, Iglehart JD, Livingston DM, Ganesan S (2006) X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell* 9(2):121–132. doi:10.1016/j.ccr.2006.01.013
35. Klein A, Wessel R, Graessmann M, Jürgens M, Petersen I, Schmutzler R, Niederacher D, Arnold N, Meindl A, Scherneck S, Seitz S, Graessmann A (2007) Comparison of gene expression data from human and mouse breast cancers: identification of a conserved breast tumor gene set. *Int J Cancer* 121(3):683–688. doi:10.1002/ijc.22630
36. Marty B, Maire V, Gravier E, Rigai G, Vincent-Salomon A, Kappler M, Lebigot I, Djelti F, Tourdès A, Gestraud P, Hupé P, Barillot E, Cruzalegui F, Tucker GC, Stern MH, Thiery JP, Hickman JA, Dubois T (2008) Frequent PTEN genomic alterations and activated phosphatidylinositol 3-kinase pathway in basal-like breast cancer cells. *Breast Cancer Res* 10(6):R101. doi:10.1186/bcr2204
37. Chen DT, Nasir A, Culhane A, Venkataramu C, Fulp W, Rubio R, Wang T, Agrawal D, McCarthy SM, Gruidl M, Bloom G, Anderson T, White J, Quackenbush J, Yeatman T (2009) Proliferative genes dominate malignancy-risk gene signature in histologically-normal breast tissue. *Breast Cancer Res Treat*. doi:10.1007/s10549-009-0344-y
38. Affymetrix (2001) Statistical algorithms reference guide, technical report. Affymetrix
39. Gautier L, Cope L, Bolstad BM, Irizarry RA (2004) affy-analysis of Affymetrix GeneChip data at the probe level. *Bioinformatics* 20(3):307–315. doi:10.1093/bioinformatics/btg405
40. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smith C, Smyth G, Tierney L, Yang JY, Zhang J (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 5(10):R80. doi:10.1186/gb-2004-5-10-r80
41. Spyrtos F, Ferrero-Poüs M, Trassard M, Hacène K, Phillips E, Tubiana-Hulin M, Le Doussal V (2002) Correlation between MIB-1 and other proliferation markers: clinical implications of the MIB-1 cutoff value. *Cancer* 94(8):2151–2159. doi:10.1002/ncr.10458
42. de Azambuja E, Cardoso F, de Castro G Jr, Colozza M, Mano MS, Durbecq V, Sotiriou C, Larsimont D, Piccart-Gebhart MJ, Paesmans M (2007) Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12, 155 patients. *Br J Cancer* 96(10):1504–1513. doi:10.1038/sj.bjc.6603756

43. Venables WN, Ripley BD (2002) *Modern Applied Statistics with S*, chap 16.3, 4th edn. Springer. ISBN 0-387-95457-0
44. Anderson WF, Chen BE, Jatoi I, Rosenberg PS (2006) Effects of estrogen receptor expression and histopathology on annual hazard rates of death from breast cancer. *Breast Cancer Res Treat* 100(1):121–126. doi:[10.1007/s10549-006-9231-y](https://doi.org/10.1007/s10549-006-9231-y)
45. Jatoi I, Chen BE, Anderson WF, Rosenberg PS (2007) Breast cancer mortality trends in the United States according to estrogen receptor status and age at diagnosis. *J Clin Oncol* 25(13):1683–1690. doi:[10.1200/JCO.2006.09.2106](https://doi.org/10.1200/JCO.2006.09.2106)
46. Polyak K (2007) Breast cancer: origins and evolution. *J Clin Invest* 117(11):3155–3163. doi:[10.1172/JCI33295](https://doi.org/10.1172/JCI33295)
47. Collins LC, Botero ML, Schnitt SJ (2005) Bimodal frequency distribution of estrogen receptor immunohistochemical staining results in breast cancer: an analysis of 825 cases. *Am J Clin Pathol* 123(1):16–20. doi:[10.1309/HCF035N9WK40ETJ0](https://doi.org/10.1309/HCF035N9WK40ETJ0)
48. Nadji M, Gomez-Fernandez C, Ganjei-Azar P, Morales AR (2005) Immunohistochemistry of estrogen and progesterone receptors reconsidered: experience with 5, 993 breast cancers. *Am J Clin Pathol* 123(1):21–27. doi:[10.1309/4WV79N2GHJ3X1841](https://doi.org/10.1309/4WV79N2GHJ3X1841)
49. Barnes DM, Millis RR, Beex LV, Thorpe SM, Leake RE (1998) Increased use of immunohistochemistry for oestrogen receptor measurement in mammary carcinoma: the need for quality assurance. *Eur J Cancer* 34(11):1677–1682. doi:[10.1016/S0959-8049\(98\)00149-X](https://doi.org/10.1016/S0959-8049(98)00149-X)
50. Ross JS, Symmans WF, Pusztai L, Hortobagyi GN (2007) Standardizing slide-based assays in breast cancer: hormone receptors, HER2, and sentinel lymph nodes. *Clin Cancer Res* 13(10):2831–2835. doi:[10.1158/1078-0432.CCR-06-2522](https://doi.org/10.1158/1078-0432.CCR-06-2522)
51. Badve SS, Baehner FL, Gray RP, Childs BH, Maddala T, Liu ML, Rowley SC, Shak S, Perez EA, Shulman LJ, Martino S, Davidson NE, Sledge GW, Goldstein LJ, Sparano JA (2008) Estrogen- and progesterone-receptor status in ECOG 2197: comparison of immunohistochemistry by local and central laboratories and quantitative reverse transcription polymerase chain reaction by central laboratory. *J Clin Oncol* 26(15):2433–2435. doi:[10.1200/JCO.2007.13.6424](https://doi.org/10.1200/JCO.2007.13.6424)
52. Mann GB, Fahey VD, Feleppa F, Buchanan MR (2005) Reliance on hormone receptor assays of surgical specimens may compromise outcome in patients with breast cancer. *J Clin Oncol* 23(22):5148–5154. doi:[10.1200/JCO.2005.02.076](https://doi.org/10.1200/JCO.2005.02.076)
53. Ma XJ, Hilsenbeck SG, Wang W, Ding L, Sgroi DC, Bender RA, Osborne CK, Allred DC, Erlander MG (2006) The HOXB13:IL17BR expression index is a prognostic factor in early-stage breast cancer. *J Clin Oncol* 24(28):4611–4619. doi:[10.1200/JCO.2006.06.6944](https://doi.org/10.1200/JCO.2006.06.6944)
54. Sørlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Børresen-Dale AL (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 98(19):10869–10874. doi:[10.1073/pnas.191367098](https://doi.org/10.1073/pnas.191367098)
55. Gruvberger S, Ringnér M, Chen Y, Panavally S, Saal LH, Borg A, Fernö M, Peterson C, Meltzer PS (2001) Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res* 61(16):5979–5984
56. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415(6871):530–536. doi:[10.1038/415530a](https://doi.org/10.1038/415530a)
57. Lusa L, McShane LM, Reid JF, De Cecco L, Ambrogi F, Biganzoli E, Gariboldi M, Pierotti MA (2007) Challenges in projecting clustering results across gene expression-profiling datasets. *J Natl Cancer Inst* 99(22):1715–1723. doi:[10.1093/jnci/djm216](https://doi.org/10.1093/jnci/djm216)
58. Teschendorff AE, Naderi A, Barbosa-Morais NL, Caldas C (2006) PACK: profile analysis using clustering and kurtosis to find molecular classifiers in cancer. *Bioinformatics* 22(18):2269–2275. doi:[10.1093/bioinformatics/btl174](https://doi.org/10.1093/bioinformatics/btl174)
59. Ertel A, Tozeren A (2008) Switch-like genes populate cell communication pathways and are enriched for extracellular proteins. *BMC Genomics* 9:3. doi:[10.1186/1471-2164-9-3](https://doi.org/10.1186/1471-2164-9-3)
60. Gormley M, Tozeren A (2008) Expression profiles of switch-like genes accurately classify tissue and infectious disease phenotypes in model-based classification. *BMC Bioinformatics* 9:486. doi:[10.1186/1471-2105-9-486](https://doi.org/10.1186/1471-2105-9-486)
61. Ertel A, Tozeren A (2008) Human and mouse switch-like genes share common transcriptional regulatory mechanisms for bimodality. *BMC Genomics* 9(1):628. doi:[10.1186/1471-2164-9-628](https://doi.org/10.1186/1471-2164-9-628)
62. Sell S, Pierce GB (1994) Maturation arrest of stem cell differentiation is a common pathway for the cellular origin of teratocarcinomas and epithelial cancers. *Lab Invest* 70(1):6–22
63. Pardal R, Clarke MF, Morrison SJ (2003) Applying the principles of stem-cell biology to cancer. *Nat Rev Cancer* 3(12):895–902. doi:[10.1038/nrc1232](https://doi.org/10.1038/nrc1232)
64. Sanai N, Alvarez-Buylla A, Berger MS (2005) Neural stem cells and the origin of gliomas. *N Engl J Med* 353(8):811–822. doi:[10.1056/NEJMra043666](https://doi.org/10.1056/NEJMra043666)
65. Reya T, Morrison SJ, Clarke MF, Weissman IL (2001) Stem cells, cancer, and cancer stem cells. *Nature* 414(6859):105–111. doi:[10.1038/35102167](https://doi.org/10.1038/35102167)
66. Jordan CT, Guzman ML, Noble M (2006) Cancer stem cells. *N Engl J Med* 355(12):1253–1261. doi:[10.1056/NEJMra061808](https://doi.org/10.1056/NEJMra061808)