Cross-platform integration of transcriptomics data

Georgia Tsiliki, Marina Ioannou and Dimitris Kafetzopoulos

Abstract An increasing number of studies have profiled gene expressions in tumor specimens using distinct microarray platforms and analysis techniques. With the accumulating amount of microarray data, one of the most challenging tasks is to develop robust statistical models to integrate the findings. This article reviews some recent studies on the field. We also study the intensity similarities between data sets derived from various platforms, after appropriate rescaling of the measurements. We found that intensity and fold-change variability similarities between different platform measurements can assist the analysis of independent data sets and can produce comparable results with those obtained for the independent data set alone.

1 Introduction

With the increasing availability of published microarray data sets there is a need to develop approaches for validating and integrating results across multiple studies. The overlap of gene expression signatures of various studies is very small, for example between the "Amsterdam" signature [23] and the "Rotterdam" signature [24], mainly due to the small sample sizes of individual studies and error measurements. A major concern in the "meta-analysis" of DNA microarrays is the lack of a single standard experimental platform for data generation. The microarray technologies

Georgia Tsiliki
FORTH, Institute of Molecular Biology & Biotechnology, P.O. Box 1385, 711 10, Heraklion, Greece, e-mail: tsiliki@imbb.forth.gr

Marina Ioannou
FORTH, Institute of Molecular Biology & Biotechnology, P.O. Box 1385, 711 10, Heraklion, Greece, e-mail: mioannou@imbb.forth.gr

Dimitris Kafetzopoulos
FORTH, Institute of Molecular Biology & Biotechnology, P.O. Box 1385, 711 10, Heraklion, Greece, e-mail: kafetzo@imbb.forth.gr

currently in use differ in how DNA sequences are laid on the array, the length of these sequences, splicing variations and the number of samples measured in each hybridization. As a result, an important source of technological variability in gene expression measurements is the platform used.

The increasing number and availability of large-scale gene expression studies of human and other organisms provide strong motivation for cross-study analyses that combine existing and/or new data sets. In a cross-study analysis, the data, relevant test statistics or conclusions of several studies are combined. Several studies have compared measurements across platforms [9] and reported their findings in terms of reproducibility of results, power increase of studies, validation of gene signature results [10] [8] [11] [25]. The MAQC Quality Control Consortium, the FDAs Critical Path Initiative, NCIs caBIG and others are implementing procedures that will broadly enhance data quality. The MAQC consortium have reported that proper sample preparation is sufficient to dramatically enhance multi-lab and multi-platform correlations [16].

However, combining data from different expression studies and possibly different gene expression platforms poses a number of statistical difficulties due to the different processing facilities. As a consequence, measurements from different platforms cannot be directly combined. Identifying and removing such systematic effects is the primary statistical challenge in cross-study analysis. We note that technological differences between studies may be confounded with biological differences arising from the choice of patient cohorts (e.g. age, gender or ethnicity). In many cases, technological artifacts are dominant, though care should be taken to verify this, and one can hope to remove them while leaving biological information intact.

Here we briefly review some recent techniques to minimize error measurements and safely combine results of studies which address the same biological questions. Furthermore, we evaluate how the direct use of intensity data from independent data sets and platforms, can facilitate the statistical analysis of other microarray studies. An advantage of such an approach is that the same methodology can be used and the measurement errors can be controlled in the same way for all data sets. Our scope is to demonstrate that the power of any statistical conclusions can be retained when the data is enhanced with external data from various platforms. For that purpose our working example is the classification of ER samples in a breast cancer data set.

1.1 Recent literature review

In general, it will make sense to combine data sets of studies which address the same questions, or, experiments with some sufficiently similar aspects so that one can hope to make better inference from the whole than from the experiment separately. However, in order to compare experiments that are performed on different gene expression platforms, the first thing one should look at is how to link oligonucleotide probe sets, spotted sequences, and other microarray features. Typically, a sequence-specific identifier (GenBank accession number) serves as a reference to the array probe sequences. Thus, the first step in a cross-study analysis would be to identify a subset of genes which are consistently measured across platforms. The next step would be to derive for each individual data set numerically comparable quantities from the expression values of genes in the common list by applying specific data transformation and normalization methods.

The most simple approach to integrate data would be to sample standardize and gene median center each available data set, and then combine data sets. More systematic approaches have been proposed for integration of findings from multiple studies using different array technologies. Particularly, according to [14], there are several potential approaches to cross-study analysis, depending on what information is being synthesized. Existing studies either combine information from primary statistics (such as t-statistics or p-values) [13] [19] or secondary statistics (such as gene lists) that are derived from the individual studies [3]. Additionally, other approaches to meta-analysis of gene-expression data are considered by [4] [15] [12], which directly integrate the data and then proceed with the analysis.

[22] proposed optimization methods for cross-laboratory and cross-platform microarray expression data, based on three simple and often employed techniques to identify discrepancy in expression data sets. They created an experimental design that compared three functionally different normal tissues: human liver, lung and spleen. Particularly, they reported that when precision, biological interpretation and multiple platform data sets were considered together, they allowed for better selection of genes with respect to a particular outcome. They considered precision and sensitivity measurements which were useful in finding the minimal detectable fold-change and raw performance values for an array platform. Also, Gene Ontology and pathway analyses were considered, which were thought to be a valuable way of examining and comparing the actual biological interpretation. Differences in pathways indicated consistency problems which could be quantified by counting the differentially expressed genes between platforms that moved in different directions.

Along these lines, [25] integrated three independent microarray gene expression data sets for breast cancer and identified a structured prognostic signature consisting of 112 genes organized into 80 pair-wise expression comparisons. The method used for integration of data sets was based on the ranks of the expression values within each sample first introduced in [5]. Since the features were rank-based, data normalization was not necessary before data integration.

A cross-study normalization method called *XPN* was suggested by [14], which based on identifying homogeneous groups of genes and samples in the combined data. Specifically, they employed k-means clustering independently to genes and samples of the combined data to identify blocks (or clusters) in the data. Then, each gene expression value was a scaled and shifted block mean plus noise. Their model assumed that the samples of each available study fall roughly into one of the statistically homogenous sample groups identified, and that each group was defined by an associated gene profile that was constant within each of the estimated gene groups. They examined three existing breast cancer data sets and reported that XPN successfully preserved biological information according to ER prediction error rates while removing systematic differences between platforms.

The reliability of gene expression across three previously published breast cancer studies was evaluated by [4]. They compared the strength of evidence of gene to phenotype associations across studies and combined effects across studies. Their methods are implemented by [2] on an `R` package `(www.r-project.org)` library called `MergeMaid` `(http://www.bioconductor.org/packages/2.2/bioc/html/MergeMaid.html)`. They defined a reliability score and set a threshold via permutations to distinguish which were the "reliable" genes in two study experiments, i.e. the genes consistently measured in all studies. For multi-study experiments they considered an alternative interclass correlation coefficient per gene. Finally, they used a between studies combined effect based on the first eigenvector of a *principal component analysis* (PCA) of each study, to determine the genes that are associated with the phenotype.

In order to account for inter-study variation, [3] suggested an "effect size" model for multiple microarray studies. They defined effect sizes as standardized indexes measuring the magnitude of a treatment or covariate effect. They suggested the use of a fixed-effects model (FEM) or a random-effects model (REM) (or alternatively a hierarchical Bayesian model) depending on the homogeneity of study effects. Finally, they measured the statistical significance of their combined results by permutation tests and FDR calculations. Many of their methods are implemented in `GeneMeta` R package library (`http://www.bioconductor.org/packages/2.2/bioc/html/GeneMeta.html`).

Finally, an interesting approach is that by [15] who applied a two-stage Bayesian mixture modeling strategy to analyze four independent breast cancer microarray studies derived from different microarray platforms (spotted cDNAs, Affymetrix GeneChip, and inkjet oligonucleotides). They derived an inter-study validated 90-gene "meta-signature" predictive of breast cancer recurrence. Their analysis was based on the signed conditional probabilities of differential expression as introduced in [12]. Particularly, [12] proposed a Bayesian mixture model transformation of DNA microarray data with potential features applicable to meta-analysis of microarray studies, although they employed them in the context of molecular classification. The basic idea was to estimate the platform independent probability of over-expression, under-expression or baseline expression for gene sample combinations given the observed expression measurements. Along these lines, [15] reported that the use of the specific probability measures increased the power of statistical analysis by increasing the sample size.

There is a great challenge to compare and integrate results across independent microarray studies. Meta-analysis studies sometimes produce comparable results even under different logics. Although all approaches, normalization or combination of secondary results, have their merits, here we proceed with studying the effects of scaling existing measurements from various platforms as that was suggested by [12]. An important selection criterion for data integration is the measurement correlations between platforms [18]. Nonetheless, a large number of genes might be lost when looking at the correlation due to different levels of noise between platforms. We find that rescaling of measurements should be able to prevent that.

2 Integrate findings

Here we suggest some characteristics of the data that need to be accounted for when assimilating results from different studies, and evaluate them in independent data sets. Particularly, we consider the "translation" procedure of values as that was first suggested by [12] and it was employed by [15] on the same content. They estimated probabilities of over-expression, under-expression and baseline expression, and translated the intensity measurements into a probability of differential expression. The new probability scale can make comparisons between platforms on a unified scale rather than using gene-specific summaries. For an analytic description of the method see [15].

We use the four data sets also considered by [15], namely the [20], [21], the [23], [7] data sets. The first two studies are cDNA microarray studies, the third is an injekt oligonucleotide array study and the fourth an Affymetrix GeneChip study. The data sets consist of 305 breast cancer samples in total and $2,555$ common genes. The study-specific breast cancer prognosis signatures have been previously reported to have a small overlap. [15] suggested that combination of the four in a probability scale derives a 90 gene meta-signature which is strongly associated with survival in breast cancer patients. We study their approach in terms of the sample's ER status categorization. Furthermore, we suggest a few modifications which seem to strengthen our results in an independent data set produced with homemade two-colour spotted arrays from Qiagen V3 human library. All results presented here are with respect to that independent data set which consists of $34,772$ 70mer probes and 29 samples (18 ER+ and 11 ER− samples). We refer to that data set as $Data_1$ from here onwards.

Measurements for all four data sets [20] [21] [23] [7] considered here are on the so called "poe" scale [15] and vary in the interval $[0,1]$. Our scope is to measure the accuracy of sample classification with respect to their ER status by using simple statistical measures. For that reason, we only consider t-test calculations and Ward's hierarchical clustering with euclidean distance. We avoid comparing our results with those derived when studies are considered individually, since those finding are based on a more advanced statistical methodology. Thus, our scope is to compare the ER classification outcome in $Data_1$ samples when it is assisted by external data and under the same statistical methodology.

2.1 ER signatures when combining data sets

If we consider all 304 samples (one sample from [23] data set had an unknown ER status and was excluded from further analysis), we find a set of 272 genes adequate to distinguish the two classes (ER+, ER−). From those we found 75 common with $Data_1$. There are some common genes with those reported by [23], for example, for ER categorization. Particularly, [23] reported a set of 550 genes, from which 223 are common with $Data_1$. However, only 12 genes are common between the two list and can be found in $Data_1$. In Figure 1 we can see the two ER signatures. An interesting observation is that both appear to have two mis-calssification errors. We apply agglomerative hierarchical clustering algorithm to expression ratios using Euclidean distance metric and Ward clustering algorithm [13].

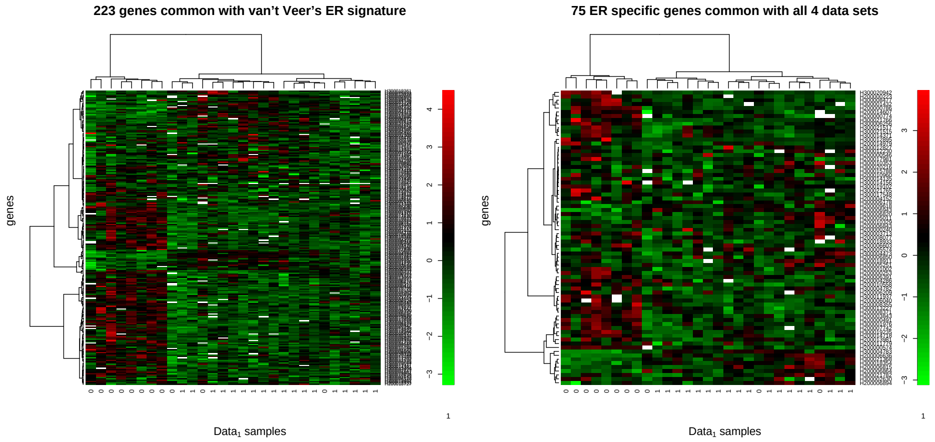

Fig. 1 The ER statistically significant genes reported by van't Veer and those found when we considered the combined four data sets. Results are shown on $Data_1$.

Alternatively, if we consider the whole of $Data_1$ and apply the same methodology as before, we find 279 genes able to statistically distinguish ER. We refer to those results as the *Intrinsic Model* results. However, it would be interesting to consider only the genes of $Data_1$ which are common with the [20] [21] [23] [7] data sets. In this case, 120 statistically significant genes are able to distinguish the two ER classes. Those results are refer to as the *Starting Model* results.

2.2 Intensity and fold-change similarities

Many times the intensity measurements vary between platforms for their common probes. That variability could indicate platform specific effects, or even random noise due to experiment conditions. In this subsection, we study how that variability can affect an ER derived signature which is based on many platforms. For that reason, we consider only probes that appear to have "similar" values across the four data sets in terms of magnitude on the "poe" scale. Particularly, since we are interested on ER classification, we search for genes with similar intensity behaviour in separately ER+ and ER− samples.

We employ Kruskal-Wallis rank sum tests [6, p.115] per gene, to test the null hypothesis that the location parameters of the distribution of ER+ and ER− samples are the same in each of the four data sets. The alternative is that they differ in at least one. We consider only genes with high p-values for both ER+ and ER− samples, which based on the test give evidence for accepting the null. The left hand-side plot of Figure 2 shows the 44 genes that appear to have the same location distribution parameters for both ER+ and ER− samples across the four data sets. For the right hand-side plot we consider 100 permutations per gene and finally report only 65 genes with significant permutation based empirical p-values with respect to ER status. We can observe that the mis-classification errors are three in both cases, however, permutation procedure is inferior in terms of the number of genes included.

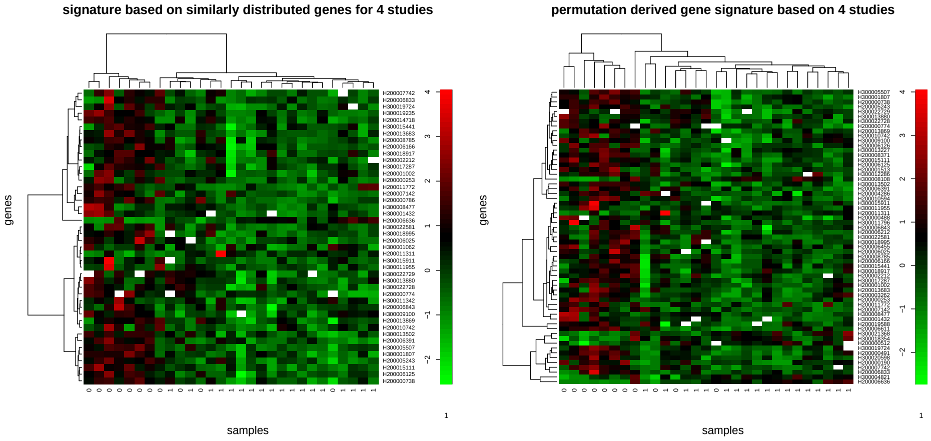

Fig. 2 The ER statistically significant genes among those with same location distribution parameters for the four data sets, as reported by Kruskal-Wallis sum rank test without or with data permutation techniques. Results are shown on $Data_1$.

Another characteristic of the data is the fold-change behaviour between the ER+ and ER− samples. When we consider genes with the same amount of fold-change variability across the four data sets, we find that 24 genes, common for the four data sets and $Data_1$, could distinguish the two ER classes. The genes were selected to have the same fold-change levels for the four platform measurements examined here. In Figure 3 we can see that the two ER classes can be well distinguished and in this case.

2.3 Results

In order to evaluate the approaches suggested before and account for statistical sampling error, we employ *multiclass bootstrap resampling* techniques and estimate via probabilistic measures whether clusters of the original data found by hierarchical

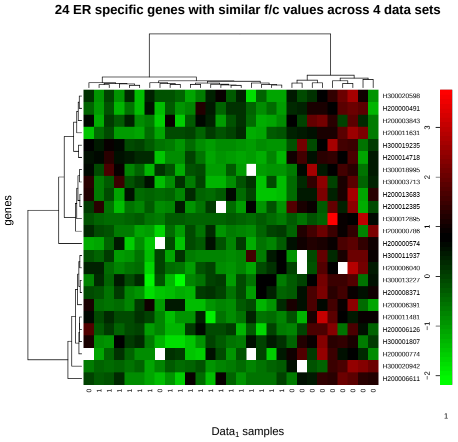

Fig. 3 The ER statistically significant genes with similar fold-change levels among the four data sets examined. Results are shown on $Data_1$.

clustering are strongly supported by the data. For that reason we calculate two types of p-values as they are defined in [17]; the *approximately unbiased* (AU) p-value and the *bootstrap probability* (BP) value. AU p-value is computed by multiscale bootstrap resampling and is thought to be a better approximation to unbiased p-value than BP value which is computed by normal bootstrap resampling. However, the AU p-values themselves include sampling error, since they are also computed by a limited number of bootstrap samples. The null hypothesis in this case is that the clusters of the data are observed by chance. Clusters with AU p-values higher than 95% are strongly supported by data, i.e. those clusters do not seem to be caused by sampling error, but may stably be observed if we increase the number of observations.

[17] suggested that 10 sample sizes for each data set should be examined. Along these lines, we consider sample sizes equal to the $r' = \{0.49, 0.6, 0.69, 0.8, 0.89, 1.0, 1.09, 1.2, 1.29, 1.4\}$ percentages of the original sample size. For each sample size we generate $10,000$ bootstrap samples. For each bootstrap sample, we apply hierarchical clustering to obtain the sets of bootstrap replications of dendrograms and compute the BP for observing each cluster. Finally, we estimate AU p-values by fitting a regression model to the BP values calculated for each cluster and each sample. For an analytic description of the method see [17].

In Table 1 we report the AU and BP values for the approaches already mentioned for the two major clusters of the data C_0 and C_1, where C_0 mostly contains ER$-$ samples and C_1 mostly contains ER$+$ samples. We also report the frequency of misclassified samples in C_0 and C_1, and the number of statistically significant genes with respect to ER status. Note the decrease in the number of significant genes because of mapping when information from combined data is used. The results in the first row of the table (Intrinsic Model) correspond to clustering results after t-test calculations are directly employed to $Data_1$, whereas, results in the second raw (Starting Model) refer to $Data_1$ but only to its common genes with the four data sets. We consider those values as a baseline for comparison with other approaches suggested here. The results in the third row (Simple Model) correspond to clustering results derived from the four merged data sets. Particularly, we found the significant genes, with respect to the ER status, when the four data sets were considered together and after *Benjamini-Hochberg* correction was applied, and applied our finding to $Data_1$. The Fold-change variability, Kruskal-Wallis (K-W), K-W with Permutations results correspond to methods presented in section 2.2.

Table 1 We report the AU and BP values from bootstraping, the frequency of mis-classified samples and the number of statistically significant genes with respect to ER status. For each variable the two values corresponds to clusters C_0 and C_1, respectively. K-W corresponds to Kruskal-Wallis method. Results are reported for $Data_1$.

Approach	AU (%)	BP (%)	Mis-classifications Freq.	Num. Genes
Intrinsic Model	88 - 83	69 - 52	0 - 0.182	279
Starting Model	89 - 88	38 - 34	0 - 0.182	120
Simple Model	75 - 79	10 - 8	0.111 - 0.15	75
Fold-change Variability	93 - 95	34 - 34	0 - 0.25	24
K-W	76 - 79	27 - 10	0 - 0.182	44
K-W with Permutations	86 - 78	12 - 7	0 - 0.182	65

We can observe that the K-W and Simple Model results have similar AU p-values, although the number of significant genes is higher in the second case. However, they both have smaller AU p-values compared to the Starting Model. Better results in terms of AU p-values and number of genes, can be observed in the case of permutation sampling with Kruskal-Wallis tests. The number of genes increases from 44 to 65 and the AU p-values are elevated supporting the alternative hypothesis that C_0 and C_1 clusters are not observed by chance. However, Fold-change Variability results exhibit the highest AU p-values, although the number of significant genes is small compared to that of the other approaches. The mis-classification frequency is relatively small in all cases, whereas the BP values are variable compared to the AU.

To prove the power of a high number of independent data sets used, in Table 2 we focus on the fold-change variability results but for only three data sets ([23] [20] [21]) and two data sets ([23] [21]) chosen at random from the four. We can observe that our results benefit in terms of AU p-values when information from more data sets is used.

Table 2 We report the AU and BP values from bootstraping, the frequency of mis-classified samples and the number of statistically significant genes with respect to ER status. For each variable the two values corresponds to clusters C_0 and C_1, respectively. K-W corresponds to Kruskal-Wallis method. Results are reported for $Data_1$.

Approach	AU (%)	BP (%)	Mis-classifications Freq.	Num. Genes
Fold-change Variability	93 - 95	34 - 34	0 - 0.25	24
F-c V 3 data sets	70 - 75	25 - 20	0 - 0.182	29
F-c V 2 data sets	63 - 69	20 - 19	0 - 0.143	31

3 Conclusions

We considered how information from studies using various platforms can facilitate the search for significant genes with respect to the categorization of ER samples. Our analysis focused on ER status classification although other parameters, binary or continuous such as breast cancer prognosis, could be studied. An obvious limitation of such approaches is the restriction of the study to only annotated common probes.

We studied the effect of rescaling measurements from four platforms to a common scale and use the information obtained by that data. We employed resampling techniques to minimize sampling error and variability introduced by the different platforms. Our results were compared to those obtained from direct analysis of data, and were thought to be able to describe properties of independent data sets. Particularly, we found that an important property in such kind of analyses is the fold-change variability of common probes across various studies. The performance of K-W analysis was also comparable to that of direct analysis, when data was enhanced with permutations. In all cases, gain in terms of AU p-values resulted in loss of some genes. Overall, we showed that knowledge from numerous data sets produced under the same biological question, can greatly assist the statistical analysis of independent data sets.

References

1. Cope, L., Garrett-Mayer, E.S., Gabrielson, E., Parmigiani, G.: The Integrative Correlation Coefficient: a Measure of Cross-study Reproducibility for Gene Expression Array Data. Working paper, Johns Hopkins University, Dept. of Biostatistics (2007)
2. Cope, L., Zhong, X., Garrett-Mayer, E.S., Parmigiani, G.: MergeMaid: R Tools for Merging and Cross-Study Validation of Gene Expression Data. Working paper, Johns Hopkins University, Dept. of Biostatistics (2004)
3. Choi, J.K., Yu, U., Kim, S., Yoo, O.J.: Combining multiple microarray studies and modeling interstudy variation. Bioinformatics **19**, i84–i90 (2003)
4. Garrett-Mayer, E., Parmigiani, G., Zhong, X., Cope, L., Gabrielson, E.: Cross-study validation and combined analysis of gene expression microarray data. Biostatistics **9(2)**, 333–354 (2008)

5. Geman, D., d'Avignon, C., Naiman, D.Q., Winslow, R.L.: Classifying Gene Expression Profiles from Pairwise mRNA Comparisons. Statistical Applications in Genetics and Molecular Biology **3**, 19 (2004)
6. Hollander, M., Wolfe, D.A.: Nonparametric Statistical Methods. New York: John Wiley & Sons (1973)
7. Huang, E., Cheng, S.H., Dressman, H., Pittman, J., Tsou, M.H., Horng, C.F., Bild, A., Iversen, E.S., Liao, M., Chen, C.M., West, M., Nevins, J.R., Huang, A.T.: Gene expression predictors of breast cancer outcomes. Lancet **361**, 1590-1596 (2003)
8. Irizarry, R.A., Warren, D., Spencer, F., Kim, I.F., Biswal, S., Frank, B.C., Gabrielson, E., Garcia, J.G., Geoghegan, J., Germino, G., Griffin, C., Hilmer, S.C., Hoffman, E., Jedlicka, A.E., Kawasaki, E., Martinez-Murillo, F., Morsberger, L., Lee, H., Petersen, D., Quackenbush, J., Scott, A., Wilson, M., Yang, Y., Ye, S.Q., Yu, W.: Multiple-laboratory comparison of microarray platforms. Nature Methods **2(5)**, 329–330 (2005)
9. Patterson, T.A., Lobenhofer, E.K., Fulmer-Smentek, S.B., Collins, P.J., Chu, T.M., Bao, W., Fang, H., Kawasaki, E.S., Hager, J., Tikhonova, I.R., Walker, S.J., Zhang, L., Hurban, P., de Longueville, F., Fuscoe, J.C., Tong, W., Shi, L., Wolfinger, R.D.: Performance comparison of one-color and two-color platforms within the MicroArray Quality Control (MAQC) project. Nature Biotechnology **24(9)**, 1140–1150 (2006)
10. Miron, M., Woody, O.Z., Marcil, A., Murie, C., Sladek, R., Nadon, R.: A methodology for global validation of microarray experiments. BMC Bioinformatics **7**, 333–352 (2006)
11. Members of Toxicogenomics Research Consortium: Standardizing global gene expression analysis between laboratories and across platforms. Nature Methods **2(5)**, (2005)
12. Parmigiani, G., Garrett, E.S., Anbazhagan, R., Gabrielson, E.: A statistical framework for expression-based molecular classification in cancer. J. R. Statist. Soc. B **64(4)**, 717–736 (2002)
13. Rhodes, D.R., Barrette, T.R., Rubin, M.A., Ghosh, D., Chinnaiyan, A.M.: Meta-Analysis of Microarrays: Interstudy Validation of Gene Expression Profiles Reveals Pathway Dysregulation in Prostate Cancer. Cancer Research **62**, 4427–4433 (2002)
14. Shabalin, A.A., Tjelmeland, H., Fan, C., Perou, C.M., Nobel, A.B.: Merging two gene-expression studies via cross-platform normalization. Bioinformatics **24(9)**, 1154–1160 (2008)
15. Shen, R., Ghosh, D., Chinnaiyan, A.M.: Prognostic meta-signature of breast cancer developed by two-stage mixture modeling of microarray data. BMC Genomics **5**, 94 (2004)
16. Shi, L., Reid, L.H., Jones, W.D., Shippy, R., Warrington, J.A., Baker, S.C., Collins, P.J., de Longueville, F., Kawasaki, E.S. et al.: The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. Nature Biotechnology **24**, 1151–1161 (2006)
17. Shimodaira, H.: Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. The Annals of Statistics **32 (6)**, 2616–2641 (2004)
18. Shippy, R., Sendera, T.J., Lockner, R., Palaniappan, C., Kaysser-Kranich, T., Watts, G., Alsobrook, J.: Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations. BMC Genomics **5**, 61 (2004)
19. Smith, D.D., Sætrom, P., Snøve Jr, O., Lundberg, C., Rivas, G.E., Glackin, C., Larson, G.P.: Meta-analysis of breast cancer microarray studies in conjunction with conserved cis-elements suggest patterns for coordinate regulation. BMC Bioinformatics **9**, 63 (2008)
20. Sørlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Eystein Lonning, P., Borresen-Dale, A.L.: Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Pnas **98 (19)**, 10869–10874 (2001)
21. Sotiriou, C., Neo, S.Y., McShane, L.M., Korn, E.L., Long, P.M., Jazaeri, A., Martiat, P., Fox, S.B., Harris, A.L., Liu, E.T.: Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Pnas **100 (18)**, 10393–10398 (2003)
22. Stafford, P., Brun, M.: Three methods for optimization of cross-laboratory and cross-platform microarray expression data. Nucleic Acids Research **35**, 10 (2007)

23. 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.: Gene expression profiling predicts clinical outcome of breast cancer. Letters to Nature **415**, 530–536 (2002)
24. Wang, Y., Klijn, J.G.M., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J., Jatkoe, T., Berns, E.M.J.J., Atkins,D., Foekens, J.A.: Gene expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet **365**, 671–679 (2005)
25. Xu, L., Tan, A.C., Winslow, R.L., Geman, D.: Merging microarray data from separate breast cancer studies provides a robust prognostic test. BMC Bioinformatics **9**, 125 (2008)