Robust Cross-generation and Cross-laboratory Predictions of Affymetrix Microarrays by Rank-based Methods

Hsi-Che Liu
Dept. of Pediatrics, Mackay Memorial Hospital, Taipei, Taiwan
College of Medicine, Taipei Medical University, Taipei, Taiwan
hsiche@ms1.mmh.org.tw

Chien-Yu Chen
Dept. of Bio-industrial Mechatronics Engineering, National Taiwan University, Taipei, Taiwan
cychen@mars.csie.ntu.edu.tw

Yu-Ting Liu and Cheng-Bang Chu
Graduate School of Biotechnology and Bioinformatics, Yuan Ze University, Chung-Li, Taiwan

Der-Cherng Liang
Department of Pediatrics, Mackay Memorial Hospital, Taipei, Taiwan

Lee-Yung Shih
Division of Hematology-Oncology, Chang Gung Memorial Hospital, Taoyuan, Taiwan
College of Medicine, Chang Gung University, Taoyuan, Taiwan

Chih-Jen Lin
Dept. of Computer Science and Information Engineering, National Taiwan University, Taipei,Taiwan
cjlin@csie.ntu.edu.tw

ABSTRACT

Past experiments of the popular Affymetrix (Affy) microarrays have accumulated a huge amount of public data sets. To apply them for more wide studies, the comparability across generations and experimental environments is an important research topic. This paper particularly investigates the issue of cross-generation/laboratory predictions. That is, whether models built upon data of one generation (laboratory) can differentiate data of another. We consider eight public sets of three cancers. They are from different laboratories and are across various generations of Affy human microarrays. Each cancer has certain subtypes, and we investigate if a model trained from one set correctly differentiates another. We propose a simple rank-based approach to make data from different sources more comparable. Results show that it leads to higher prediction accuracy than using expression values. We further investigate normalization issues in preparing training/testing data.

1: INTRODUCTION

Gene expression profiling by DNA microarrays is a useful tool in biological and clinical research. Superior to traditional biological experiments, it compares thousands of genes simultaneously. With fast and systematic analysis of expression values, one can quickly identify significant genes for certain diseases or build models for patient diagnosis/tumor classification.

Though the microarray technology is popular, not many institutions can conduct enough experiments for effective analysis due to the lack of patient samples or the high cost. Studies in recent years have accumulated a huge amount of microarray samples in public databases. If data experimented under similar conditions can be combined together, not only any laboratory can directly apply microarray technology in practical use, but also more extensive and reliable studies are possible. A microarray experiment from raw samples to expression values is a complicated procedure. Expression values from various sources are not easily comparable. Many recent studies explore the cross-platform comparability between cDNA and oligonucleotide arrays, but so far contradictive results have been reported. Even using the same samples, some papers (e.g., [14, 17]) conclude that measurements from the two platforms are poorly correlated. Though recent studies (e.g., [2,11,15]) give more promising results, they still consider that the reproducibility across platforms is not easily available.

For the same type of arrays, comparability issues also occur. In particular, whether results from various generations of popular Affymetrix (Affy) human oligonucleotide arrays can be used together is an issue. Though the same platform tends to produce more consistent expression values, cross-generation and

cross-laboratory use of Affy arrays remains a challenging task. This paper intends to have a detailed investigation on this subject. Existing papers of this topic mainly study the following three issues:

1. Whether differentially expressed genes identified across two generations (laboratories) are similar or related.
2. Whether the same samples lead to similar expression values across two generations (laboratories).
3. Whether models built upon data of one generation (laboratory) can differentiate data of another.

This paper focuses on studying the third issue.

Most work studying issue 1 concludes that genes identified across generations (laboratories) are related (e.g., [21]). In contrast, the other two issues are less settled. For issue 2, one of the first studies is [18]. Using the same samples on two generations, it reports that better similarity of the probe sets leads to higher correlation between expression values. References [10, 13] further strengthen this finding by showing that considering only probes with overlapping sequences gives excellent comparability. However, even with these studies, some still doubt the reproducibility of expression values across generations, so several papers propose more sophisticated techniques. By calculating expression changes at the probe-level, Elo et al. [7] report that such information gives better comparability. Bhattacharya and Mariani [3] propose regression models, which reflect the relationship between expression values of two generations.

Compared to issue 2, issue 3 concerns more about the practical use of data from different sources. Many applications such as cancer diagnosis and tumor classification are of this type. If data of other laboratories can be used, an institution can classify its patient samples without huge initial experiments/costs. Some have checked issue 3: Bloom et al. [4] collect samples of 21 tumor types across different laboratories and two Affy generations. They normalize expression values of various sources and apply an Artificial Neural Networks (ANN) model. High prediction accuracy (88%) is reported. Jiang et al. [12] consider lung cancer data sets across two generations. They develop special data transformation and report high prediction accuracy. Xu et al. [33] study prostate cancer samples across different laboratories but under the same Affy generation. Using a classifier based on only two genes, a model built on data from three laboratories successfully separates a test set from another laboratory to normal or cancer.

In this paper, we investigate whether expression values are reliable for the prediction tasks. In cross-platform analysis (e.g., cDNA and oligonucleotide), quite a few (e.g., [26, 31]) observe inconsistent expression values, so they use information less dependent on the scale of values (e.g., rank levels). While expression values seem to be more consistent if only Affy arrays are considered, it is essential to check which way is better. We propose a rank-based approach and compare it with using expression values.

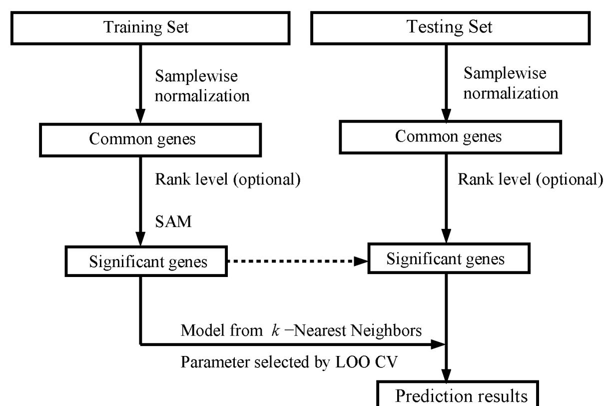

Figure 1. Workflow of the analysis. The width of each box reflects the number of genes. Details are in the beginning of Section 2.

2: METHODS

Figure 1 outlines our approach. In each experiment two sets of the same cancer type are used as training/prediction sets. The details of our methods are described in the following subsections.

2.1: MICROARRAY DATA COLLECTION AND PREPROCESSING

To conduct the integrated, cross-generation and cross-laboratory predictions of Affy human arrays, we select public data of three cancer types. The first type, acute lymphoblastic leukemia (ALL), includes two data sets of different generations published from the same laboratory. The data set of Yeoh et al. [34], from HG-U95Av2 array, is denoted as ALL-95. Another data set of Ross et al. [23], from HG-U133A array, is called ALL-133. The 132 samples of ALL-133 were chosen from the original 335 HG-U95Av2 data obtained by Yeoh et al. To generate an independent source, ALL-95 includes only 203 non-replicated cases. The subtypes to be predicted are ALLs with defined recurrent chromosomal aberrations: t(12;21), t(1;19) and hyperdiploid with more than 50 chromosomes (HD>50). The second cancer type is acute myeloid leukemia (AML). Three data sets are generated by three different institutions using HG-U133A chips: The set Ross et al. [22], abbreviated as AML-1, is a childhood study. The other two studies, AML-2 (Valk et al. [29]) and AML-3 (Guti´errez et al. [8]), involve adult samples. The predicted subtypes are AMLs with t(8;21), inv(16) and t(15;17). These biologically distinct subtypes are identical both in pediatric and adult AMLs. The last group is breast cancer. Three data sets across three generations of chips (HuGeneFL, HGU95Av2 and HG-U133A) are collected from three different institutions. They are denoted as Breast-FL (West et al. [32]), Breast-95 (Huang et al. [9]) and Breast-133 (Wang et al. [30]), according to the generation of chip used in each individual study. The estrogen receptor (ER) status (positive or negative) is the variable to be predicted. The summary of key characteristics and URL addresses of all

the eight microarray data sets can be found at http://biominer.bime.ntu.edu.tw/~cychen/datasets.pdf.

After expression values are downloaded from the referred public websites, values of each array are rescaled by setting the 2% trimmed mean of all the genes in an array to be 500, as suggested in the Affy Microarray Suite 5.0 (MAS 5.0) program.

2.2: GENE MAPPING (COMMON PROBE SETS IDENTIFICATION)

According to their launch time, the three array generations can be aligned as the order of HuGeneFL, HG-U95Av2, and HG-U133A. Because of multiple design advances used to produce newer Affy human arrays, many probe sets differ between generations of arrays. For a comparative analysis, it is critical to identify a subset of common genes. One approach is to match the UniGene IDs among genes. Each UniGene ID corresponds to a cluster containing sequences that represent a unique gene and its related information [19]. However, different UniGene Builds are used for the three Affy generations. Some UniGene IDs can not be exactly tracked between two Builds. An alternative method considers LocusLink (currently implemented as Entrez Gene [16]), and it does not suffer from the same problem as much. Another popular method uses matching tables provided by Affymetrix. The matching between two generations of arrays is based on the similarity of sequence information of probe sets (http://www.affymetrix.com/support/technical/comparison_spreadsheets.affx). According to different constructions, there are two mappings called "Good Match" and "Best Match." The latter, obtained under a more stringent criterion than the former, chooses probe sets with the greatest likelihood of representing the same gene across generations [1].

Hwang et al. [10] test methods of UniGene IDs, LocusLink IDs and Best Match to match genes between HG-U95Av2 and HG-U133A arrays. They experimented with 14 samples on both generations of arrays. Correlation coefficients indicate that Best Match demonstrates higher reproducibility than the other two methods. We thus adopt the method of Best Match in the study. Since Best Match is not available between HuGeneFL and HG-U95Av2, to generate matched probe sets, we follow similar procedures in previous studies [3, 12, 18]. Most probe sets between HG-U95Av2 and HG-U133A have one-to-one correspondence. For few multiple (HG-U95Av2)-to-one (HG-U133A) mappings, we select the first-cited probe set on HG-U95Av2 to make them one-to-one. The situation for HuGeneFL and HG-U95Av2 is more complicated as multiple-to-multiple relations occur. For any given HuGeneFL probe set, from its corresponding ones in HG-U95Av2, we select the one with the highest overlap. This procedure leads to multiple (HuGeneFL)-to-one (HGU95Av2) matchings. We then apply the same process for HG-U95Av2 and HG-U133A to obtain one-to-one relationships. Finally, we get a list of 5,979 common probe sets between HuGeneFL and HG-U95Av2, and a list of 9,530 probe sets between HG-U95Av2 and HG-U133A. The intersection of the above two lists gives 5,045 probe sets, which are shared across three generations.

2.3: RANK-BASED NORMALIZATION

Previous work has shown that considering a gene's rank within a chip instead of using its expression value better eliminates systematic biases and thus improves the classification accuracy [25, 31, 33]. There are some variants of the rank-based normalization. The simplest one replaces the expression value of a gene by its rank among expression values of a single chip [24, 27]. Quantile normalization is another rank-based approach [5, 20]. It calculates a value for each rank level by taking the average of the expression values of that particular rank in available arrays, and then replaces the expression value of each gene by the associated reference value of its rank. Median rank scores is also a rank-based approach. This variant derives the median of each gene among the available arrays and sorts those medians as the reference value of a particular rank [25, 31]. Tsodikov et al. [27] show that replacing expression values by ranks performs well in terms of selecting differentially expressed genes. Qiu et al. [20] also reveal that this simple scheme outperforms the quantile normalization method in reduction of between-gene dependence and identification of differential genes. Thus in this work we investigate if a direct replacement of expressions by ranks is effective in cross-generation and cross-laboratory comparisons. Below we provide details of the adopted procedure.

First of all, we obtain common genes from data of each cancer type. The rank-based normalization method then replaces the value of each probe set with its rank in the set of common genes. Next, gene selection is performed with SAM (Significance Analysis of Microarrays) [28] to identify the list of differentially expressed genes based on the training data set of each experiment. The FDR (false discovery rate) is set as 5%. As far as the testing data is concerned, the same procedure of replacing expression values with ranks is applied. After that, the list of differential probe sets selected based on the training data set filters out unwanted probe sets in the testing data set

2.4: PREDICTIONS

In each experiment, one data set is for training and another data set (across generations or laboratories) is for testing. The k-Nearest Neighbors (KNN) [6] is employed in the prediction task. For any instance in the test set, KNN predicts its class by the majority class of its k closest neighbors in the training set. The distance between any two data instances is by the Euclidean metric. Since the performance of KNN depends on the parameter k, in data classification one usually implements a validation procedure to select it. Here we consider leave-one-out cross-validation (LOO CV). For any given k, LOO CV sequentially singles one training

instance out for validation. That is, KNN predicts this instance by checking its neighbors in the remaining set. The value k with the best LOO CV accuracy is then applied to predict the independent testing data. In our experiments, we consider odd integers from 1 to 17 to search for the best k. Values beyond this range do not give better LOO CV.

3: RESULTS AND DISCUSSION

This section first compares prediction results under two ways of processing arrays: one directly uses gene expression values, and the other transforms expression values to rank levels. The experimental procedure has several variants of normalization, so subsequently we check their respective effects.

3.1: A COMPARISON BETWEEN USING EXPRESSION VALUES AND RANK INFORMATION

To perform cross-generation/laboratory predictions, one can prepare training and testing data by directly using expression values of significant genes. However, the scale of each gene may vary due to different chip generations or experimental environments. We can instead use each gene's rank in the same subset of significant genes. Experiments below compare which method is better.

For any cancer type, each experiment considers one subtype as the target prediction label. A data set of this cancer type is used as the training data of two classes: whether an array is associated with the specified subtype or not. For each array in another set (called testing data), we then predict its class label and calculate the accuracy. This procedure is repeated for every two sets of the same cancer type. One exception is that AML-3 has no t(8:21) arrays, so for this subtype AML-3 is not used as the testing set.

Except the difference on using expression values or rank levels, all other settings follow the procedure in Figure 1. Table 1 gives results of the comparison. Table 1 clearly shows that using the ranks of the selected genes within an array consistently produces better results than using the original expression values. For ALL and AML, the prediction by using rank levels is excellent. One exception is to predict inv(16) by using AML-3 as the training set. Since AML-3 contains only four inv(16) arrays, there is no enough information to discriminate this subtype from others.

Both methods give worse accuracy in predicting breast cancer subtypes. As indicated earlier, three data sets of this cancer type are the most heterogeneous. They are cross-generation as well as cross-laboratory, but ALL sets are cross-generation only and AML sets are cross-laboratory only. Training Breast-95 to classify the other two sets gives much lower accuracy than other cases. We suspect the reason is that Breast-95 is the most unbalanced (74 ER+ and 15 ER− arrays) among the three breast cancer sets.

Table 1. A comparison of cross-generation(or laboratory) predictions: using rank levels and expression values (Exp. val.). In each row, we boldface the value which gives higher accuracy.

Training→Testing	Cancer subtype	Accuracy (%)	
		Rank	Exp. val.
Acute lymphoblastic leukemia (ALL)			
	t(12;21)	**96.2**	68.1
ALL-95→ALL-133	HD>50	91.6	**92.4**
	t(1;19)	**100**	97.7
	t(12;21)	**97**	93.1
ALL-133→ALL-95	HD>50	**95**	87.1
	t(1;19)	**99.5**	98.5
Acute myeloid leukemia (AML)			
AML-1→AML-2	t(15;17)	**99.2**	**99.2**
	t(8;21)	**100**	99.6
	inv(16)	**97.8**	91.2
AML-1→AML-3	t(15;17)	**100**	95.3
	inv(16)	**97.6**	88.3
AML-2→AML-1	t(15;17)	**100**	**100**
	t(8;21)	**99.2**	96.9
	inv(16)	**99.2**	90.7
AML-2→AML-3	t(15;17)	**97.6**	93
	inv(16)	**97.6**	95.3
AML-3→AML-1	t(15;17)	99.2	**100**
	inv(16)	**90**	89.2
AML-3→AML-2	t(15;17)	**99.6**	97.1
	inv(16)	**98.2**	93.3
Breast cancer			
Breast-FL→Breast-95	ER(+/−)	**86.5**	83.1
Breast-FL→Breast-133		**89.2**	81.1
Breast-95→Breast-FL		**87.8**	53.1
Breast-95→Breast-133		**86.4**	75.2
Breast-133→Breast-FL		**85.7**	75.5
Breast-133→Breast-95		**86.5**	83.1

3.2: ADDITIONAL NORMALIZATION FOR EXPRESSION VALUES AND RANK LEVELS

While Table 1 indicates that rank levels are better than expression values, we investigate if the same conclusion stands after slight changes of the experimental procedure. One issue we intend to study is the effect of gene-wise normalization. That is, after selecting significant genes, for each gene we normalize ranks or expression values in all training arrays to have mean zero and standard deviation one. In data classification such a procedure is called feature scaling (normalization). The purpose is to avoid the possible dominance of genes having large values. The same scaling factors are then employed to normalize the testing data. Table 2 lists accuracy with and without gene-wise normalization. Using expression values, the accuracy with normalization is slightly improved for ALL and AML, but is worse for some cases of breast cancer. Thus one cannot conclude that this normalization is always helpful. For rank levels, the accuracy with/without gene-wise normalization is very similar. It

consistently outperforms expression values no matter the gene-wise normalization is performed or not. Overall this normalization has a bigger effect on using expression values than rank levels.

4: CONCLUSIONS

We conduct a detailed study on cross-generation and cross-laboratory predictions of Affy microarray data. A focus is on investigating if using expression values is suitable. Experiments show that an alternative way of using simple rank levels gives more stable prediction results.

The framework proposed in this paper is rather simple. As more studies involve such cross-generation and cross-laboratory predictions, we expect our approach to be very useful. For example, existing data can be trained to predict arrays from a new generation of Affymetrix human oligonucleotide array, Plus 2.0. Future work includes experiments on more cancer types or future Affy generations.

5: ACKNOWLEDGEMENTS

The study was supported by a grant from Mackay Memorial Hospital (MMH-E-95009).

Table 2. A comparison showing the effect of genewise normalization. In each row, we boldface the value which gives the highest accuracy.

Genewise normalization		Y		Y	
Training→Testing	Subtype	Rank		Exp. val.	
Acute lymphoblastic leukemia(ALL)					
	t(12;21)	96.9	96.2	**98.4**	68.1
ALL-95→ALL-133	HD>50	93.1	91.6	**97.7**	92.4
	t(1;19)	**100**	**100**	99.2	97.7
	t(12;21)	**97.5**	97	**97.5**	93.1
ALL-133→ALL-95	HD>50	94.5	**95**	78.3	87.1
	t(1;19)	**99.5**	**99.5**	**99.5**	98.5
Acute myeloid leukemia (AML)					
	t(15;17)	**99.6**	99.2	98.9	99.2
AML-1→AML-2	t(8;21)	**100**	**100**	**100**	99.6
	inv(16)	98.2	97.8	**98.5**	91.2
AML-1→AML-3	t(15;17)	97.6	**100**	95.3	95.3
	inv(16)	97.6	97.6	**100**	88.3
	t(15;17)	**100**	**100**	**100**	**100**
AML-2→AML-1	t(8;21)	**99.2**	**99.2**	98.4	96.9
	inv(16)	96.9	**99.2**	98.4	90.7
AML-2→AML-3	t(15;17)	**97.6**	**97.6**	95.3	93
	inv(16)	97.6	97.6	**100**	95.3
Breast cancer, subtype: ER(+/–)					
Breast-FL→Breast-95		84.3	**86.5**	84.3	83.1
Breast-FL→Breast-133		85.7	**89.2**	85	81.1
Breast-95->Breast-FL		81.6	**87.8**	51	53.1
Breast-95→Breast-133		80.1	**86.4**	72.7	75.2
Breast-133→Breast-FL		**87.8**	85.7	53.1	75.5
Breast-133→Breast-95		82	**86.5**	82	83.1

REFERENCES

[1] Affymetrix. User's guide to product comparison spreadsheets, 2003.

[2] T. Bammler, R. Beyer, S. Bhattacharya, G. Boorman, A. Boyles, B. Bradford, R. Bumgarner, P. Bushel, K. Chaturvedi, and D. Choi. Standardizing global gene expression analysis between laboratories and across platforms. Nature Methods, 2(5):351–356, May 2005.

[3] S. Bhattacharya and T. J. Mariani. Transformation of expression intensities across generations of affymetrix microarrays using sequence matching and regression modeling. Nucleic Acids Research, 33(18):e157, 2005.

[4] G. Bloom, I. V. Yang, D. Boulware, K. Y. Kwong, D. Coppola, S. Eschrich, J. Quackenbush, and T. J. Yeatman. Multi-platform, multi-site, microarray-based human tumor classification. American Journal of Pathology, 164:9–16, 2004.

[5] B. Bolstad, R. Irizarry, M. Astrand, and T. Speed. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics, 19(2):185–93, Jan 2003.

[6] T. Cover and P. Hart. Nearest neighbor pattern classification. IEEE Trans Information Theory, 13:21–27, 1967.

[7] L. L. Elo, L. Lahti, H. Skottman, M. Kylaniemi, R. Lahesmaa, and T. Aittokallio. Integrating probe-level expression changes across generations of affymetrix arrays. Nucleic Acids Research, 33(22):e193, 2005.

[8] N. C. Guti´errez, R. L´opez-P´erez, J. M. Hern´andez, I. Isidro, B. Gonz´alez, M. Delgado, E. Fermi˜n´an, J. L. Garc´ıa, L. V´azquez, M. Gonz´alez, and J. F. S. Miguel. Gene expression profile reveals deregulation of genes with relevant functions in the different subclasses of acute myeloid leukemia. Leukemia, 19:402–409, 2005.

[9] E. Huang, S. H. Cheng, H. Dressman, J. Pittman, M. H. Tsou, C. F. Horng, A. Bild, E. S. Iversen, M. Liao, C.M. Chen, M. West, J. R. Nevins, and A. T. Huang. Gene expression predictors of breast cancer outcomes. Lancet, 361(9369):1590–1596, 2003.

[10] K.-B. Hwang, S. W. Kong, S. A. Greenberg, and P. J. Park. Combining gene expression data from different generations of oligonucleotide arrays. BMC Bioinformatics, 5:159, 2004.

[11] R. A. Irizarry, D. Warren, F. Spencer, I. F. Kim, S. Biswal, B. C. Frank, E. Gabrielson, J. G. N. Garcia, J. Geoghegan, G. Germino, C. Griffin, S. C. Hilmer, E. Hoffman, A. E. Jedlicka, E. Kawasaki, F. Martinez-Murillo, L. Morsberger, H. Lee, D. Petersen, J. Quackenbush, A. Scott, M. Wilson, Y. Yang, S. Q. Ye, and W. Yu. Multiple-laboratory comparison of microarray platforms. Nature Methods, 2:345–350, 2005.

[12] H. Jiang, Y. Deng, H.-S. Chen, L. Tao, Q. Sha, J. Chen, C.-J. Tsai, and S. Zhang. Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. BMC Bioinformatics, 5:81, 2004.

[13] S. W. Kong, K. B. Hwang, R. D. Kim, B.-T. Zhang, S. A. Greenberg, I. S. Kohane, and P. J. Park. Crosschip: a system supporting comparative analysis of different generations of affymetrix arrays. Bioinformatics, 21(9):2116–2117, 2005.

[14] W. P. Kuo, T.-K. Jenssen, A. J. Butte, L. Ohno-Machado, and I. S. Kohane. Analysis of matched mRNA measurements from two different microarray technologies. Bioinformatics, 18(3):405–412, 2002.

[15] J. E. Larkin, B. C. Frank, H. Gavras, R. Sultana, and J. Quackenbush. Independence and reproducibility across

microarray platforms. Nature Methods, 2(5):337–344, April 2005.

[16] D. Maglott, J. Ostell, K. D. Pruitt, and T. Tatusova. Entrez gene: gene-centered information at ncbi. Nucleic Acids Research, 33:D54–D58, 2005.

[17] N. Mah, A. Thelin, T. L. S. Nikolaus, T. K¨uhbacher, Y. Gurbuz, H. Eickhoff, G. Kl¨oppel, H. Lehrach, B. Mellg°ard, C. M. Costello, and S. Schreiber. A comparison of oligonucleotide and cDNA-based microarray systems. Physiological Genomics, 16:361–370, 2004.

[18] A. Nimgaonkar, D. Sanoudou, A. J. Butte1, J. N. Haslett, L. M. Kunkel, A. H. Beggs, and I. S. Kohane. Reproducibility of gene expression across generations of affymetrix microarrays. BMC Bioinformatics, 4:27, 2003.

[19] J. Pontius, L. Wagner, and G. Schuler. Unigene: a unified view of the transcriptome. In NCBI Handbook. Bethesda (MD): National Center for Biotechnology Information, 2003.

[20] X. Qiu, A. I. Brooks, L. Klebanov, and A. Yakovlev. The effects of normalization on the correlation structure of microarray data. BMC Bioinformatics, 6:120, 2005.

[21] D. R. Rhodes, J. Yu, K. Shanker, N. Deshpande, R. Varambally, D. Ghosh, T. Barrette, A. Pandey, and A. M. Chinnaiyan. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proceedings of the National Academy of Sciences of the United States of America, 101(25):9309–9314, 2004.

[22] M. E. Ross, R. Mahfouz, M. Onciu, H.-C. Liu, X. Zhou, G. Song, S. A. Shurtleff, S. Pounds, C. C. J. Ma, R. C. Ribeiro, J. E. Rubnitz, K. Girtman,W. K.Williams, S. C. Raimondi, D.-C. Liang, L.-Y. Shih, C.-H. Pui, and J. R. Downing. Gene expression profiling of pediatric acute myelogenous leukemia. Blood, 104:3679–3687, 2004.

[23] M. E. Ross, X. Zhou, G. Song, S. A. Shurtleff, K. Girtman, W. K.Williams, H.-C. Liu, R. Mahfouz, S. C. Raimondi, N. Lenny, A. Patel, and J. R. Downing. Classification of pediatric acute lymphoblastic leukemia by gene expression profile. Blood, 102:2951–2959, 2003.

[24] A. Szabo, K. Boucher, W. Carroll, L. Klebanov, A. Tsodikov, and A. Yakovlev. Variable selection and pattern recognition with gene expression data generated by the microarray technology. Mathematical Biosciences, 176(1):71–98, 2002.

[25] J. T¨odling and R. Spang. Assessment of five microarray experiments on gene expression profiling of breast cancer. Poster Presentation RECOMB, 2003.

[26] R. W. Tothill, A. Kowalczyk, and D. Rischin. An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin. Cancer Research, 65(10):4031–4040, 2005.

[27] A. Tsodikov, A. Szabo, and D. Jones. Adjustments and measures of differential expression for microarray data. Bioinformatics, 18(2):261–60, Feb 2002.

[28] V. G. Tusher, R. Tibshirani, and G. Chu. Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America, 98:5116–5121, 2001.

[29] P. J. Valk, R. G. Verhaak, M. A. Beijen, C. A. Erpelinck, S. B. van Waalwijk van Doorn-Khosrovani, J. M. Boer, H. B. Beverloo, M. J. Moorhouse, P. J. van der Spek, B. Lowenberg, and R. Delwel. Prognostically useful gene-expression profiles in acute myeloid leukemia. New England Journal of Medicine, 350:1617–1628, 2004.

[30] Y. Wang, J. G. M. Klijn, Y. Zhang, A. M. Sieuwerts, M. P. Look, F. Yang, D. Talantov, M. Timmermans, M. E. M. van Gelder, J. Yu, T. Jatkoe, E. M. J. J. Berns, D. Atkins, and J. A. Foekens. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet, 365(9460):671–679, 2005.

[31] P. Warnat, R. Eils, and B. Brors. Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. BMC Bioinformatics, 6:265, 2005.

[32] M. West, C. Blanchette, H. Dressman, E. Huang, S. Ishida, R. Spang, H. Zuzan, J. John A. Olson, J. R. Marks, and J. R. Nevins. Predicting the clinical status of human breast cancer by using gene expression profiles. Proceedings of the National Academy of Sciences of the United States of America, 98(20):11462–11467, 2001.

[33] L. Xu, A. C. Tan, D. Q. Naiman, D. Geman, and R. L. Winslow. Robust prostate cancer marker genes emerge from direct integration of inter-study microarray data. Bioinformatics, 21(20):3905–3911, 2005.

[34] E.-J. Yeoh, M. E. Ross, S. A. Shurtleff, W. K. Williams, D. Patel, R. Mahfouz, F. G. Behm, S. C. Raimondi, M. V. Relling, A. Patel, C. Cheng, D. Campana, D. Wilkins, X. Zhou, J. L. H. Liu, C.-H. Pui, W. E. Evans, C. Naeve, L. Wong, and J. R. Downing. Pediatric lymphoblastic leukemia by gene expression profiling. Cancer Cell, 1:133–143, 2002.