Cross-platform classification in microarray-based leukemia diagnostics

Björn Nilsson
Anna Andersson
Mikael Johansson
Thoas Fioretos

Gene expression profiling is a powerful technique for classifying hematologic malignancies. Its clinical use is, however, currently hindered by the need to collect large sets of expression profiles at each diagnostic facility. To overcome this limitation, we introduced cross-platform classification, allowing classifier construction using pre-existing microarray datasets. As proof-of-principle, we performed cross-platform classification of acute myeloid leukemia and childhood acute lymphoblastic leukemia using expression data from four different facilities. We show that cross-platform classification of these disorders is achievable, and, strikingly, that the diagnostic accuracy can be retained. We conclude that cross-platform classification constitutes an effective and convenient way to implement microarray diagnostics.

Key words: acute leukemia, gene expression profiling, diagnostic classification, cross-platform analysis.

Haematologica 2006; 91:821-824

From the Department of Clinical Genetics, Lund University Hospital, Lund, Sweden (BN, AA, TF); Department of Signals, Systems and Sensors, Royal Institute of Technology, Stockholm, Sweden (MJ).

Correspondence:
Björn Nilsson, Department of Clinical Genetics, Lund University Hospital, 221 85 Lund, Sweden.
E-mail: bjorn.nilsson@med.lu.se

Gene expression profiling using microarrays has rapidly evolved into a powerful technique for investigating hematologic malignancies. In particular, many leukemia subtypes display characteristic expression patterns that admit diagnostic classification.[1-7] Despite these advances, the clinical use of array-based diagnostics is late in coming. A major cause of this delay is that, to be effective, the advanced computational methods employed in array-based classification (e.g., support-vector machines or *k*-nearest neighbors) must be calibrated (trained) with large sets of example gene expression profiles (training data). The production of these data involves profiling tissue samples from a substantial number of patients, which can be overwhelmingly resource-consuming. Therefore, an attractive approach would be to re-use pre-existing sets of expression profiles as training data. In fact, several leukemia datasets are available in public databases. This approach means that the training set is generated on a remote microarray set-up, different from the one used locally for analyzing new samples for classification and can, therefore, be referred to as cross-platform classification, as opposed to the current within-platform approach, which utilizes the same platform throughout. The difficulty with the cross-platform approach is that data from different microarray set-ups cannot be compared directly. However, notwithstanding initial observations indicating that results may be highly discordant,[8-10] recent studies show that - with cautious data processing - cross-platform consistency and reproducibility may be better than previously believed.[11-13] Herein, we propose classification across platforms to facilitate array-based diagnostics. Specifically, we classify cases of childhood acute lymphoblastic leukemias (ALL) and adult acute myeloid leukemias (AML) using training and test data from distinct facilities.

Design and Methods

Data normalization

A prerequisite for classification is that all data are represented in a common numerical format. In cross-platform classification, this may not be the case because two platforms, with different characteristics, are involved. To achieve acceptable results, all data must be standardized (normalized). We consider four normalization methods: mean centering (MC), mean centering with unit variance (MCUV), relative ranks (RR), and relative ranks with unit variance (RRUV) (*see Technical Supplement, online only*).

Classifier design and performance assessment

To evaluate the cross-platform approach, we computed the overall cross-platform classification correctness using within-platform classification as the state-of-the-art control method. Throughout, we used *k*-nearest neighbors (*k*-NN) classifiers. To verify the robustness of the results, a wide range of classifier parameters was used (*see Technical Supplement*). All datasets used are described in Table 1 and in the Technical Supplement. Intuitively, a correctness rate is the number of correctly classified cases divided by the total number of cases. The exact computations are, however, more

Table 1. Overview of the datasets used for evaluation of the methods.

	Oligonucleotide Ross et al.	*cDNA Andersson et al.*
B-cell ALL, hyperdiploid	16	29
B-cell ALL with *ETV6/RUNX1* gene fusion	17	20
B-cell ALL with *TCF3/PBX1* gene fusion	15	6
B-cell ALL with *BCR/ABL1* gene fusion	15	3
B-cell ALL with *MLL* gene rearrangement	20	2
T-cell ALL	13	11
	Valk et al.	*Bullinger et al.*
AML with normal karyotype	116	45
AML with t(8;21)	22	11
AML with inv(16)	19	15
AML with t(15;17)	18	12
AML with monosomy 7	15	11
AML with 11q23 rearrangement	17	8
AML with trisomy 8	16	4

ALL: acute lymphoblastic leukemia; AML: acute myeloid leukemia. The table elements indicate the numbers of patients with each leukemia subtype, as defined by cytogenetic or molecular genetic criteria.

complicated because care must be taken to ascertain that the within- and cross-platform correctness rates can be meaningfully compared (*see Technic0al Supplement*). Nevertheless, if these measures are equal or nearly equal, the accuracy of the cross-platform approach is on par with that of the within-platform approach, i.e. the diagnostic precision is not impaired.

Results and Discussion

To evaluate the cross-platform approach (Figure 1) in a diagnostic setting, we created a childhood ALL classifier that was trained using the dataset produced by Ross *et al*.[3] (Table 1). This classifier was used for subtype prediction in an independent series of childhood ALL cases from our laboratory[4] (Table 1). Because the datasets are generated on oligonucleotide and cDNA arrays, respectively, the classification is performed across platforms. As shown in Table 2 (*and Supplementary Figure 1A and Supplementary Table 1*), the ALL classifier performed well. The differences between the overall cross- and within-platform correctness rates were small (average difference -5.0% to –0.6%), and all four normalization methods yielded excellent results.

Mean centering with unit variance (1.8% to -0.9%) and relative ranks with unit variance (-1.8% to -0.6%) performed better than mean centering (-3.9% to -2.8%) and relative ranks (-5.0% to -3.4%). Throughout, most errors were caused by misclassified cases of ALL with *BCR/ABL1* (*Supplementary Table 2*). In conclusion, cross-platform classification of childhood ALL is achievable with retained accuracy. Next, we created an adult AML classifier using the oligonucleotide-based dataset published by Valk *et al*.[5] (Table 1) as training data. Again, to challenge the cross-platform approach, the classifier

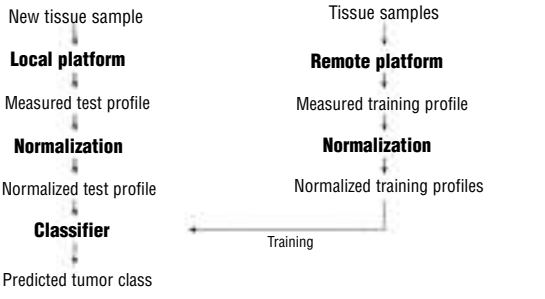

Figure 1. Outline of the cross-platform classification concept. First, a suitable existing microarray data set containing gene expression profiles of the leukemia subtypes of interest is obtained, e.g. from a public domain data repository. Second, the data are normalized. Third, the classifier is calibrated (trained) using the normalized data. Fourth, new profiles from incoming samples (unknown leukemic subtype) are normalized and then classified. Hence, in contrast to current strategies, classification is performed without prior data collection at the local diagnostic facility.

was applied to an independent series from an unrelated microarray set-up, in this case the cDNA-based series by Bullinger *et al*.[6] (Table 1). This experiment corresponds to constructing an AML classifier at the Bullinger facility using externally produced training data. As shown in Table 2 (and *Supplementary Figure 1B)*, the differences between the diagnostic precisions were small. Mean centering with unit variable and relative ranks with unit variance yielded the best results (-1.0% to +1.9% and -1.5% to 0.0%). Relative ranks performed almost as well (-4.0% to +0.7%), whereas mean centering yielded relatively poor results (-15.2% to -8.2%). Together, these data support that AML classification across platforms is feasible, and that normalization using mean centering with unit variable or relative ranks with unit variance may be advantageous. To test cross-platform classification with cDNA training data and oligonucleotide test data, we interchanged the roles of the Valk and Bullinger sets, mimicking establishment of array-based AML diagnostics at the Valk facility by importing the Bullinger dataset. As shown in Supplementary Figure 1C, Table 2 and *Supplementary Table 1*, the cross-platform and within- platform classification results were comparable. In all AML experiments, most errors were caused by misclassifications among four subtypes: normal karyotype, +8, -7 and 11q23 (*Supplementary Table 2*). The ALL experiment was not be reversed because the low numbers of cases of the *TCF3/PBX1*, *BCR/ABL1* and *MLL* subtypes in the Andersson dataset prevent adequate classifier training. For completeness, we performed cross-platform classification without normalization. As expected, poor results were obtained (cross-platform correctness ~15-40%), confirming that normalization is necessary.

Classification of hematologic malignancies by gene expression profiling is a potentially valuable diagnostic tool. However, the clinical use of this technology is fundamentally limited by the fact that a large set of gene expression profiles of the leukemic subtypes of interest must be provided. With current classification strategies, such a set must be collected at each diagnostic facility.

Table 2. **Summary of the classification results (k=10).**

	Correctness (%)		
	Cross-platform	*Within-platform*	*Difference*
ALL study			
Mean centering	93.3	96.3	-3.1
Mean centering, unit variance	94.7	96.3	-1.6
Relative ranks	92.6	96.3	-3.8
Relative ranks, unit variance	95.4	96.3	-0.9
AML study			
Mean centering	67.8	78.0	-10.2
Mean centering, unit variance	78.1	78.0	+0.1
Relative ranks	77.6	78.0	-0.4
Relative ranks, unit variance	77.9	78.0	-0.1
Reversed AML study			
Mean centering	76.0	77.5	-1.5
Mean centering, unit variance	75.1	77.5	-2.5
Relative ranks	69.6	77.5	-7.9
Relative ranks, unit variance	72.9	77.5	-4.6

Average cross-platform and within-platform correctness over 10 to 1000 discriminatory genes. The results shown were obtained with a k-NN classifier (k= 10). Comparable cross- and within-platform correctness rates were obtained for k=5 to 15 (Supplementary Table 1).

Herein, we propose cross-platform classification to increase the accessibility of array-based classification. The important advantage of this approach is that it bypasses data collection by allowing diagnostic classifiers to be trained with imported, externally produced sets of expression profiles. In particular, datasets may be obtained from public domain repositories, e.g. the NCBI Gene Expression Omnibus (GEO) or Oncomine[14] databases. Hence, in principle, new locally profiled samples can be classified directly without extensive prior reference sample profiling.

As proof-of-principle, we performed subtype classification of childhood ALL and adult AML. We demonstrated that the diagnostic precision of the cross-platform approach reaches that of the current within-platform approach. Hence, we conclude that cross-platform classification of childhood ALL and adult AML is feasible, and, remarkably, that the diagnostic accuracy can be retained. Overall, the classification accuracy was higher for ALL than for AML. To some extent, this could be explained by the more heterogeneous nature of AML expression patterns,[5,6] whereas ALL subtypes have been shown to exhibit strong signatures.[3,4] Further, the misclassification patterns obtained within and across platforms were similar. For ALL, most errors were caused by misclassified *BCR/ABL1* cases, consistent with previous findings indicating that the expression patterns of this subtype may be heterogeneous.[4,15] For AML, most errors were caused by confusions among the normal karyotype, 11q23 rearrangement, -7 and +8 subtypes. This could be explained by subtype heterogeneity[5,6] or by less distinct expression signatures.[16]

Previous work on cross-platform classification was reported by Bloom *et al*.,[17] who created a well-performing multiple-site classifier, mainly for epithelial tumors. Stec *et al*.[18] studied cross-platform classification of breast cancer using relatively few discriminatory genes, but observed significant performance drops. Recently, Warnat *et al*.[19] performed cross-platform classification of prostate cancer, breast cancer and AML, reporting good results for prostate cancer and breast cancer but poor results (cross-platform correctness below 40%) for the same AML sets studied here. The reason for the lower accuracies in their study is unclear, but could be attributed to the combination of normalization, gene selection, and classification methods used. Moreover, we systematically compared four different methods for performing the necessary data normalization. Our data indicate that the methods of mean centering with unit variance and relative ranks with unit variance yield better results than plain mean centering or plain relative ranks. Most likely, this is explained by the fact that the two former strive to normalize the signal amplitude, which varies between set-ups.

Our findings are consistent with those of other recent studies indicating that the reproducibility of microarray results between platforms and laboratories may be better than initially believed.[12–14] Cross-platform comparisons depend on the accurate identification of genes across platforms, implying that imperfect array annotation, cross-hybridization pattern discrepancies, and detection of different splice-variants may constitute error sources not present in the within-platform method. Our results indicate that, although most likely present, errors introduced by these sources have little impact on the final classification result, probably because measurements from several genes are integrated, thereby diluting the effects of occasional inconsistencies.

In conclusion, array-based classification across platforms constitutes an effective and convenient method that should facilitate the implementation and further development of clinical microarray diagnostics.

BN designed the project and conducted the research. AA contributed one of the microarray datasets. MJ and TF supervised the project and oversaw the experiments.This work was supported by research grants from the Swedish Cancer Society, the Swedish Children's Cancer Foundation, and the Medical Faculty of Lund University. The authors declare that they have no potential conflicts of interest.

Manuscript received January 11, 2006. Accepted March 29, 2006.

References

1. Golub T, Slonim D, Tamayo P, Huard C, Gaasenbeek M, Mesirov J, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531-7.
2. Yeoh E, Ross M, Shurtleff S, Williams W, Patel D, Mahfouz R, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukaemia by gene expression profiling. Cancer Cell 2002;1:133-43.
3. Ross M, Zhou X, Song G, Shurtleff S, Girtman K, Williams W, et al. Classification of pediatric acute lymphoblastic leukemia by gene expression profiling. Blood 2003; 102:2951-9.
4. Andersson A, Olofsson T, Lindgren D, Nilsson B, Ritz C, Eden P, et al. Molecular signatures in childhood acute leukemia and their correlations to expression patterns in normal hematopoietic subpopulations. Proc Natl Acad Sci USA 2005;102:19069-74.
5. Valk P, Verhaak R, Beijen M, Erpelinck C, van Waalwijk van Doorn-Khosrovani S. B, Boer J, et al. Prognostically useful gene-expression profiles in acute myeloid leukemia. N Engl J Med 2004;350:1617-28.
6. Bullinger L, Dohner K, Bair E, Frohling S, Schlenk R, Tibshirani R, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. N Engl J Med 2004; 350:1605-16.
7. Haferlach T, Kohlmann A, Schnittger S, Dugas M, Hiddemann W, Kern W, et al. A global approach to the diagnosis of leukemia using gene expression profiling. Blood 2005;106:1189-98.
8. Kuo W, Jenssen T, Butte A, Ohno-Machado L, Kohane I. Analysis of matched mRNA measurements from two different microarray technologies. Bioinformatics 2002;18:405-12.
9. Rogojina A, Orr W, Song B, Geisert E. Comparing the use of Affymetrix to spotted oligonucleotide microarrays using two retinal pigment epithelium cell lines. Mol Vis 2003;9:482-96.
10. Woo Y, Affourtit J, Daigle S, Viale A, Johnson K, Naggert J et al. A comparison of cDNA, oligonucleotide and affymetrix GeneChip gene expression microarray platforms. J Biomol Tech 2004;15:276-84.
11. Irizarry R, Warren D, Spencer F, Kim I, Biswal S, Frank B et al. Multiple-laboratory comparison of microarray platforms. Nat Methods 2005;2:345-50.
12. Larkin J, Frank B, Gavras H, Sultana R, Quackenbush J. Independence and reproducibility across microarray platforms. Nat Methods 2005;2:337-44.
13 Bammler T, Beyer R, Bhattacharya S, Boorman G, Boyles A, Bradford B, et al. Standardizing global gene expression analysis between laboratories and across platforms. Nat Methods 2005; 2:351-6.
14. Rhodes D, Yu J, Shanker K, Deshpande N, Varambally R, Ghosh D et al. ONCOMINE: A Cancer Microarray Database and Data-Mining Platform. Neoplasia 2004;6:1-6.
15. Fine B, Stanulla M, Schrappe M, Ho M, Viehmann S, Harbott J, Boxer L. Gene expression patterns associated with recurrent chromosomal translocations in acute lymphoblastic leukemia. Blood 2004; 103: 1043-9.
16. Virtaneva K, Wright F, Tanner S, Yuan B, Lemon W, Caliguri M, Bloomfield C, de La Chapelle A, Krahe R. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proc Natl Acad Sci USA 2001;98:1124-9.
17. Bloom G, Yang I, Boulware D, Kwong K, Coppola D, Eschrich S, et al. Multi-platform, multi-site, microarray-based human tumor classification. Am J Pathol 2004;164:9-16.
18. Stec J, Wang J, Coombes K, Ayers M, Hoersch S, Gold D, et al. Comparison of the predictive accuracy of DNA array-based multigene classifiers across cDNA arrays and Affymetrix GeneChips. J Mol Diagn 2005;7:357-67.
19. Warnat P, Eils R, Brors B. Cross-platform analysis of cancer microarray data improves gene expression based classification of phenotypes. BMC Bioinformatics 2005;6:265.