Available online at www.sciencedirect.com
SCIENCE DIRECT
Genomics 87 (2006) 653 – 664
GENOMICS
www.elsevier.com/locate/ygeno

A highly standardized, robust, and cost-effective method for genome-wide transcriptome analysis of peripheral blood applicable to large-scale clinical trials

Svenja Debey [a,1], Thomas Zander [a,1], Benedikt Brors [b], Alexey Popov [a], Roland Eils [b], Joachim L. Schultze [a,*]

[a] *Molecular Tumor Biology and Tumor Immunology, Clinic I for Internal Medicine, University of Cologne, Kerpenerstrasse 62, 50937 Cologne, Germany*
[b] *Division for Theoretical Bioinformatics, German Cancer Research Center, Heidelberg, Germany*

Received 18 August 2005; accepted 19 November 2005
Available online 4 January 2006

Abstract

The use of peripheral blood mononuclear cells (PBMC) for transcriptome analysis has already been proven valuable for assessing disease-associated and drug-response-related gene signatures. While these proof-of-principle studies have been critically important, the instability of RNA within PBMC prohibits their use in large-scale multicenter trials for which samples have to be transported for a prolonged time prior to RNA isolation. Therefore, a prerequisite for transcriptome analysis of peripheral blood in clinical trials will be a standardized and valid method to stabilize the RNA profile immediately after blood withdrawal. Moreover, to be able to perform such large-scale clinical studies routinely in several hundred patients more cost-effective array technologies are required. To address these critical issues, we have combined a whole-blood RNA stabilization technology with a method to reduce globin mRNA, followed by genome-wide transcriptome analysis using a newly introduced BeadChip oligonucleotide technology. We demonstrate that the globin mRNA reduction method results in significantly improved data quality of stabilized RNA samples with low intragroup variance and a detection rate of expressed genes similar to that in PBMC. More important, even small differences in gene expression such as are observed between females and males were detected and sufficient to predict gender in whole-blood samples. We therefore propose the combination of globin mRNA reduction after whole-blood RNA stabilization with a newly introduced cost-effective BeadChip array as the preferred approach for large-scale multicenter trials, especially when establishing predictive markers for disease and treatment outcome in peripheral blood.

Keywords: Peripheral blood; Peripheral blood mononuclear cells; Gene expression profiling; Transcriptome

Microarray-based transcriptome analysis is the first mature genome-wide technology to be considered for routine clinical use [1,2]. In fact, numerous landmark studies, particularly in cancer research, have highlighted the power of this technology to improve molecular diagnostics [3–6], prediction of prognosis at diagnosis [7–11], and even prediction of drug efficacy [12–14]. These landmark studies led to the acceptance of the technology and its significantly increasing use, especially in clinical medicine. More recently, peripheral blood mononuclear cells (PBMC) have been used to study transcriptome profiles in systemic lupus erythematosus [15–17], acute myeloid leukemia [11], renal cell carcinoma (RCC), or drug response in RCC [18].

Guidelines for reporting, annotation, and data analysis of microarray data have been set (MIAME [19]), and recent reports have focused on the impact of different microarray platforms and lab-to-lab variability in reproducibility and comparability of microarray results [20–22], suggesting the use of highly standardized protocols from RNA amplification to data analysis. While the latter findings are of great importance, still too little attention has been paid to sample procurement, storage, and preparation, particularly when using peripheral blood or PBMC for transcriptome analysis. We and others have demonstrated that sample handling and prolonged transportation significantly influence gene expression profiles [23,24]. Particularly concerning is the development of a

* Corresponding author. Fax: +49 221 478 86095.
E-mail address: Joachim.Schultze@uk-koeln.de (J.L. Schultze).
[1] These authors contributed equally to this work.

0888-7543/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.ygeno.2005.11.010

hypoxia-associated gene signature in PBMC when processed after delayed sample handling [23], since such influences seem to have a significant impact on expression signatures associated with, e.g., malignant diseases (T. Zander, unpublished observation). For multicenter trials, it has been demonstrated that significant effort will be necessary to achieve a high grade of standardization leading to good quality gene expression data when using PBMC [25]. To overcome these limitations, new approaches have been developed to stabilize gene expression profiles, especially in peripheral blood. One of the most promising techniques for whole-blood sample handling in clinical trials is the PAXgene Blood RNA System, which allows standardized blood collection. While analysis of individual genes by RT-PCR in PAXgene-stabilized blood samples suggested stability of mRNA expression even after prolonged sample storage [24], we have recently demonstrated that blood samples prepared with the PAXgene system demonstrate a significant increase in overall variability and reduction in present call rates when applying Affymetrix microarray technology, limiting its application for large-scale expression profiling studies. It has been speculated that the reduced sensitivity and also the variability might be due to the predominant amount of globin mRNA transcripts in whole-blood samples [23,26].

Because peripheral blood is most likely to become the prime tissue for prediction of treatment and disease outcome or diagnosis of systemic diseases, a method is critically needed that addresses the current limitations of gene expression profiling of peripheral blood. For large-scale clinical trials, a substantial cost reduction in currently available gene expression technology will also be of great importance. We have addressed these limitations by evaluating a method that combines stabilization of expression profiles in whole-blood-derived RNA samples with a globin mRNA reduction approach. This method results in improved microarray results such as increased present calls and reduced variance. More important, we demonstrate for the first time that this method increases the sensitivity toward overall small differences in gene expression profiles between two groups, as assessed by the establishment of more robust classifiers for gender prediction than with PAXgene samples that were not subjected to globin reduction. Furthermore, we demonstrate that whole-blood-derived RNA samples prepared by this method can be successfully applied to a newly introduced and very cost effective bead-based oligonucleotide microarray (Illumina BeadChip) platform, thus providing a reliable, robust and cost-effective method allowing genome-wide gene expression profiling of peripheral blood in large clinical trials.

Results

Improvement of array performance of whole-blood samples by globin mRNA reduction

To apply transcriptome analysis to large multicenter clinical trials, RNA derived from peripheral blood needs to be stabilized directly after blood withdrawal to prevent ex vivo transcriptional changes [23]. PAXgene (PAX) is a commercially available system stabilizing RNA profiles in blood samples. However, we previously observed a large intragroup variance and low rates of genes detected as present ("present calls") in whole-blood samples prepared by PAX [23]. Therefore, the use of whole-blood samples in clinical trials depends on the improvement of both present call rate and variance. The high abundance of globin mRNA has been suggested to account for these limitations [23,26]. We therefore tested the impact of a methodology to reduce globin mRNA from total RNA prepared with PAXgene prior to array analysis (PAX-GRP) on present call rate and variance. Gene expression profiles of PAX-stabilized blood samples either with or without reduction of globin mRNA from 14 healthy individuals including 7 females and 7 males were assessed on the Affymetrix platform (U133A). As demonstrated in Fig. 1A, samples prepared with PAX-GRP showed a reduced variance compared to PAX samples, suggesting that the globin mRNA reduction is beneficial to the quality of the dataset. Furthermore, present call rates for PAX-GRP samples significantly ($p < 0.001$) increased from 40.98 ± 3.27% for PAX samples to 49.99 ± 3.27% for PAX-GRP samples (Fig. 1B). In comparison to a historical control of PBMC samples (56.7 ± 2.8%), the present call rate of PAX-GRP samples was now almost in the same range as the PBMC samples. Successful removal of globin transcripts in PAX-GRP samples was also demonstrated by agarose gel electrophoresis of the generated cRNA products (Fig. 1C). The initial discrete cRNA bands derived from highly abundant globin mRNA transcripts in PAX samples changed to the typical cRNA smear pattern after globin reduction. These findings are in concordance with a significant ($p < 0.001$) decrease in the signal intensities for the α1 (37,38%; probe set 211699_x_at), α2 (39,93%; probe set 211745_x_at), and β hemoglobin transcripts (48,13%; probe set 211696_x_at) (Fig. 1D).

A limitation of the PAX-GRP method could be a nonspecific deletion of transcripts other than globin due to the RNase H treatment. In this case signal intensities for such transcripts should be higher in PAX samples compared to PAX-GRP samples despite the overall reduction of present calls in PAX samples. To assess this issue, we performed a comparison analysis of the two experimental groups (PAX vs. PAX-GRP) with fold change >2 and difference of mean >50 as filtering criteria. In total we found 415 transcripts with higher mean signal intensities in PAX samples (1.86% of all transcripts assessed). To assess whether the differences in mean signal intensities were statistically significant for each probe set and all sample pairs (PAX vs. PAX-GRP), we performed two-sample t tests ($p < 0.05$) for each of the 415 transcripts, each represented by 11 perfect-match probes on the U133A arrays, in the 14 corresponding sample pairs, resulting in 5810 t tests. Most surprising, in 90.7% (5267) of the t tests performed, no significant differences between PAX and PAX-GRP were revealed. For only 9 of the 415 transcripts all 14 comparisons revealed statistically significant differences. These genes have been listed in Supplemental Table 2. We hypothesized that the

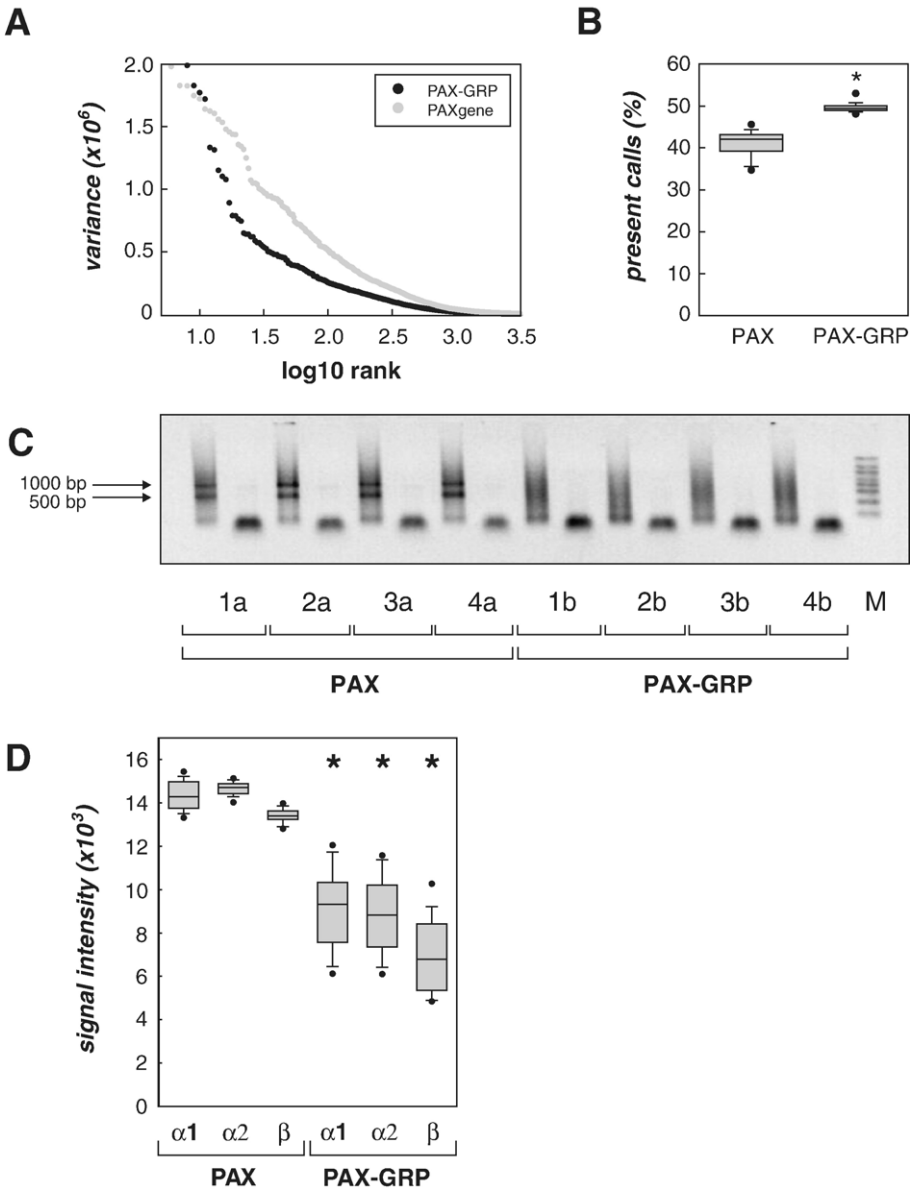


Fig. 1. Comparison of present call rates and variance in PAX and PAX-GRP samples. Gene expression data were generated using the Affymetrix U133A array. (A) The variance for all probe sets was calculated within each group (PAX, $n = 14$; PAX-GRP, $n = 14$), ordered by rank, and plotted against the decade logarithm of the rank. (B) The present calls were assessed with dCHIP1.3 software. The boundary of the box closest to 0 indicates the 25th percentile, the line within the box marks the median, and the boundary of the box farthest from 0 indicates the 75th percentile. Whiskers above and below the box indicate the 90th and 10th percentiles. Outliers are plotted as dots. *Significant vs. PAX, $p < 0.001$. (C) Original gel electrophoresis picture of cRNA derived from whole-blood samples prepared with PAXgene (PAX; lanes 1a–4a) or cRNA generated from the same RNA samples after globin reduction (PAX-GRP; lanes 1b–4b). M, molecular weight marker. The lanes between the cRNA samples are the respective fragmented cRNAs. (D) Signal intensities for hemoglobin without (PAX) or with removal of hemoglobin transcripts (PAX-GRP). The boundary of the box closest to 0 indicates the 25th percentile, the line within the box marks the median, and the boundary of the box farthest from 0 indicates the 75th percentile. Whiskers above and below the box indicate the 90th and 10th percentiles. Outliers are plotted as dots. *Significant vs. PAX, $p < 0.001$.

higher mean intensities of the 415 transcripts in PAX samples might be due mainly to an increased frequency of outliers in the PAX samples. This can be assessed by determining the variability of the signal intensities for the perfect-match probes. As a measure of variability we calculated the relative standard deviations (SD/mean) of the 11 perfect-match probes for the 415 transcripts in all 28 arrays and used these to build the ratios between PAX and PAX-GRP samples ((SD/mean [PAX])/(SD/mean [PAX-GRP])). We found that 98% of these ratios were >1, indicating higher variance of signal intensity in PAX samples compared to PAX-GRP samples due to outliers in the perfect-match probes. Therefore, most differences between PAX and PAX-GRP samples are not due to nonspecific decrease in signals by the GRP method, but due to the higher variability of signal intensity in PAX samples.

In summary, our data strongly suggest that qualitative limitations in present call rates and variability of PAXgene samples can be overcome by using PAXgene-based RNA stabilization combined with reduction of globin mRNA transcripts (PAX-GRP).

Establishing class predictors for gender differences in PBMC samples

A typical approach when using transcriptome data to determine disease or treatment outcome would be to build classifiers to predict responders or nonresponders, patients with respectively good or bad prognosis [7–10]. Several different methods such as supervised learning models [3] or prediction analysis of microarrays (PAM) [27] have been previously applied to identify such predictors. For two reasons, we initially chose to establish a class predictor for gender in a dataset derived from PBMC samples. First, in previous studies only PBMC and not whole-blood-derived RNA was used to assess class predictors in patients with renal cell carcinoma, lupus erythematosus, or juvenile arthritis. Second, we wanted to use two distinct groups with preferably very small gene expression differences to be able to determine whether more clinically applicable methods of RNA preparation are also usable to detect these minor differences of gene expression in peripheral blood. Small differences between females and males in peripheral blood were recently reported by Whitney et al. [28].

In our study using PBMC samples for class comparison analysis we also revealed very small gender-associated differences (Fig. 2). Even when applying filter criteria of low stringency (fold change >2, difference of mean >50; $p < 0.05$), only eight transcripts were found to be differentially expressed between the two groups (Fig. 2A). Six of the eight genes are located on the Y chromosome and two genes are located on the X chromosome. These differences are considerably smaller than described, for example, for healthy controls and trauma patients [25]. Reanalyzing this previously reported dataset with the above described filter criteria revealed 308 differentially expressed genes (data not shown). The rather small differences between females and males in our dataset were also confirmed by applying unsupervised methods such as hierarchical clustering. Hierarchical clustering performed with genes selected by large variation across samples (0.5 < standard deviation/mean < 10) did not result in separation of female and male subjects (Fig. 2B), indicating that differences in the gene

A

Gene Symbol	Accession Number	Affymetrix probe set	Mean Expression		Difference of Means	Chromosome
			female	male		
RPS4Y1	NM_001008	201909_at	414.61	3380.68	2966.08	Y
DDX3Y	NM_004660	20500_at	26.76	618.11	591.35	Y
SMCY	NM_004653	206700_s_at	86.03	439.89	353.86	Y
EIF1AY	BC005248	204409_s_at	53.05	350.55	297.51	Y
CYorf15B	AL049280	214131_at	50.89	165.61	114.72	Y
USP9Y	NM_004654	206624_at	53.80	109.44	55.65	Y
	AA628440	221728_x_at	862.08	65.84	-796.23	X
	AV699347	214218_s_at	905.75	34.48	-871.28	X

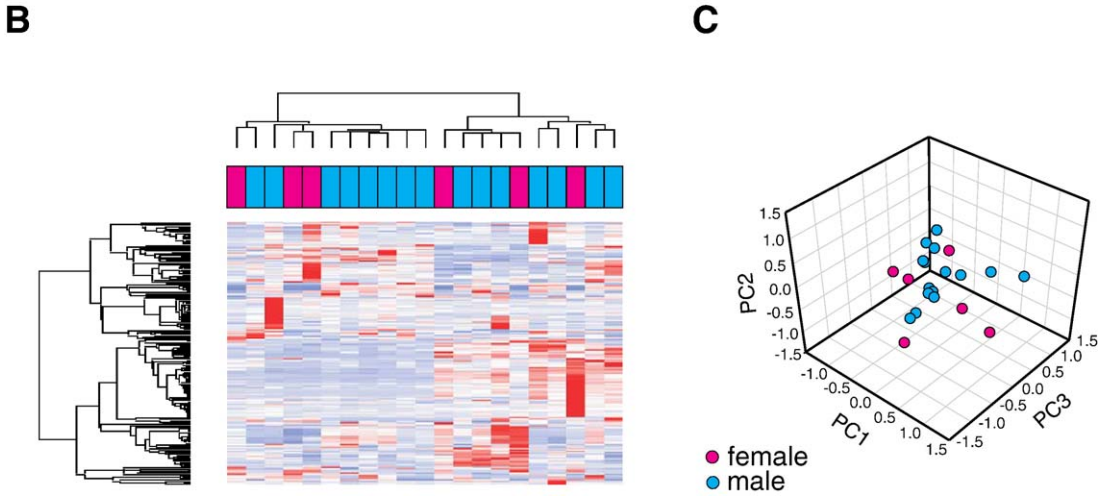


Fig. 2. Unsupervised analyses of gender differences in PBMC. (A) Genes differentially expressed in PBMC samples between females and males. Filter criteria were fold change >2, difference of mean >50, and $p < 0.05$. Analysis was performed in dCHIP1.3 on data normalized also with dCHIP. (B) Cluster analysis of 21 PBMC samples. Hierarchical clustering of samples was performed with dCHIP1.3 software using precalculated distances (1 − correlation) and centroid linkage method. Samples were clustered with genes selected by large variation across samples (0.5 < standard deviation/mean < 10). Pink indicates females, blue indicates males. (C) Principal components analysis with 21 PBMC samples. Principal components analysis was performed using R software. Scores for each case and component were calculated and plotted.

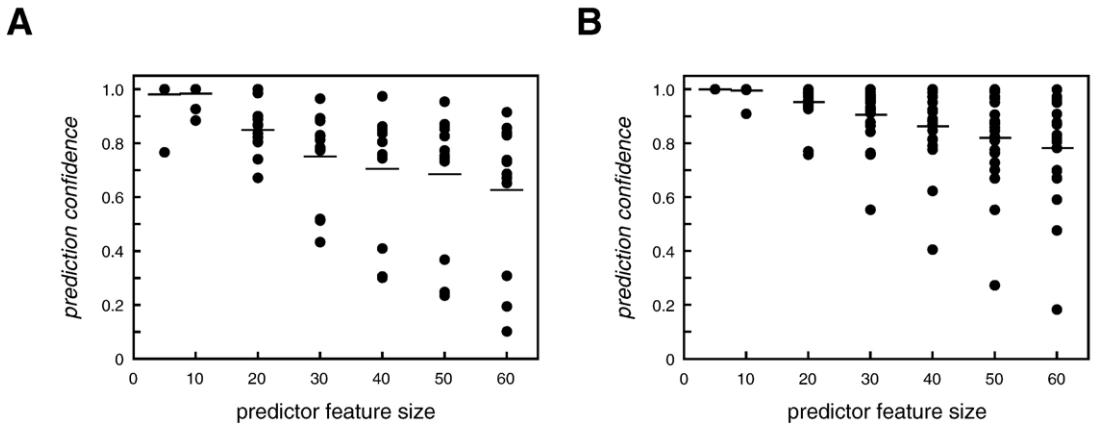
signatures of the two groups are indeed very small. Using principal components analysis (PCA) we demonstrated that variation within healthy individuals is not assigned to clearly different subgroups. In summary the first three principal components accounted for 57% of the overall variation within the group of PBMC samples but no clear separation on gender basis was detected (Fig. 2C).

Next we tested whether these small differences are sufficient to build class predictors for gender. We used a dataset of PBMC samples derived from 6 females and 6 males and applied a supervised learning model using leave-one-out cross-validation [3]. For this approach 5 to 60 features (genes used to build a predictor) were used to predict the class assignment. As shown in Fig. 3A, despite the small sample size of the dataset, correct class assignment of all 12 samples was reached for all predictors, although prediction confidence was low for some individuals. As expected for two groups with small differences in gene expression and a small sample size, mean values of prediction confidence were high with predictors of small feature sizes (5–10 features), but steadily decreased with increasing features size. Including PBMC samples from an additional 9 males in the analysis did not change the class assignment. However, the larger sample size increased the mean prediction confidence for predictors of larger feature size. Furthermore, the previously poor performing samples demonstrated also an increased prediction confidence, suggesting that a larger dataset can lead to an even more robust classifier (Fig. 3B). In contrast to class prediction of gender, classifiers for trauma patients were very robust even using large predictor feature sizes (data not shown), further demonstrating that the chosen model of gender prediction is most suited to interrogating novel approaches such as PAX-GRP for quality, since very small differences in gene expression between two groups can be assessed.

To corroborate the data derived from leave-one-out cross-validation we used a second approach for class prediction of gender—the estimation of misclassification errors by cross-validation (MCR estimation)—which combines nested cross-validation together with the PAM algorithm [27]. Starting with the comparison of 6 females and 6 males this approach also resulted in a correct class prediction for each individual (Fig. 3C) and this was also true when increasing the number of individuals in the male sample group from 6 to 15 (Fig. 3D). Taken together, despite the small number of significantly changed genes, gender can already be predicted very accurately in a rather small set of PBMC samples.

Globin reduction (PAX-GRP) increases accuracy and confidence of prediction in datasets from PAX samples

Since we demonstrated that class predictors could be established in a dataset derived from PBMC samples, we next evaluated whether similarly small differences between genders can be detected and whether class predictors can be generated in data derived from the 14 PAX and PAX-GRP samples, as described above. Calculating differences in gene expression by class comparison analysis (fold change >2, difference of mean >50; $p < 0.05$) revealed only four transcripts to be differentially



C

MCR estimate: 10-fold cross-validation, 3-fold

	cases	prediction		
		male	female	class error
male	6	6	0	0
female	6	0	6	0

D

MCR estimate: 10-fold cross-validation, 3-fold

	cases	prediction		
		male	female	class error
male	15	15	0	0
female	6	0	6	0

Fig. 3. Gender prediction in PBMC samples using leave-one-out cross-validation. (A) The scatter plot shows the gender prediction results within a dataset derived from 12 PBMC samples (6 females and 6 males) generated using Affymetrix U133A arrays. Accurately predicted samples are denoted by closed circles, incorrectly predicted samples are denoted by open circles. Horizontal lines denote the mean prediction confidence. (B) An analysis similar to that shown in A. However, instead of 12 PBMC samples, 21 PBMC samples (6 females and 15 males) were used for the analysis. (C) MCR estimation based on PAM classification [27] as implemented in the R package MCRestimate. The misclassification rate was estimated by running three reiterations of 10-fold cross-validation. (D) MCR estimation based on PAM classification using the 21 sample set.

expressed in PAX samples (Fig. 4A) and nine transcripts in PAX-GRP samples (Fig. 4B). Three of the gender-specific transcripts identified in PBMC samples were also identified in PAX samples (Fig. 4A). The transcript of the granulysin gene (*GNLY*) was the only one passing our filtering criteria in PAX samples but not in PBMC samples. As we had already previously observed, the gene expression values in PAX samples assessed on the Affymetrix platform were generally lower than for PBMC samples (compare Figs. 2A and 4A), which might also be one of the reasons even fewer genes passed our filtering criteria in PAX samples.

Class comparison analysis of PAX-GRP samples using the same filtering criteria revealed nine transcripts to be differentially expressed between females and males of which four transcripts were also identified in PBMC samples (Fig. 4B). These transcripts were located either on the X or the Y chromosome, while the remaining five transcripts detected in PAX-GRP samples were coded on chromosomes 2, 6, and 11 (Fig. 4B).

Next we applied the same methods described for PBMC samples to generate class predictors for gender in PAX samples. In contrast to PBMC samples we were unable to establish a class predictor without misclassification of one sample when applying the supervised learning model (Fig. 4C). Misclassification was independent of feature size; even at very low feature size misclassification was observed. Moreover, the prediction confidence was greatly reduced compared to PBMC. When applying MCR estimation algorithms we corroborated

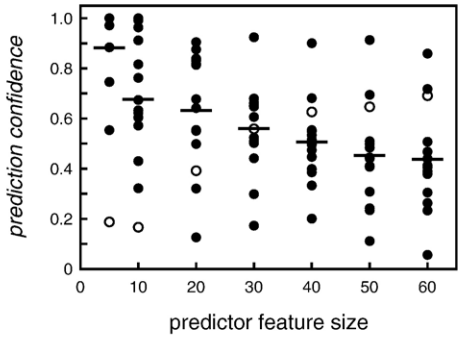
A

Gene Symbol	Accession Number	Affymetrix probe set	mean expression		difference of means	Chromosome
			female	male		
RPS4Y1	NM_001008	201909_at	33.5	507.43	473.93	Y
GNLY	NM_006433	205495_s_at	278.43	576.22	297.79	2
	AA628440	221728_x_at	100.96	16.67	-84.29	X
	AV699347	214218_s_at	94.62	6.41	-88.21	X

B

Gene Symbol	Accession Number	Affymetrix probe set	mean expression		difference of means	Chromosome
			female	male		
RPS4Y1	NM_001008	201909_at	45.63	636.06	590.43	Y
GNLY	M85276	37145_at	199.59	490.15	290.56	2
GNLY	NM_006433	205495_s_at	181.78	413.95	232.17	2
HLA-DPA1	M27487	211991_s_at	109.51	234.01	124.5	6
DDX3Y	NM_004660	205000_at	-1.12	57.36	58.48	X
ATXN1	AW235612	203231_s_at	119.12	47.25	-71.86	6
ADM	NM_001124	202912_at	179.6	80.95	-98.66	11
	AA628440	221728_x_at	115.99	6.12	-109.87	X
	AV699347	214218_s_at	125.51	1.53	-123.98	X

C

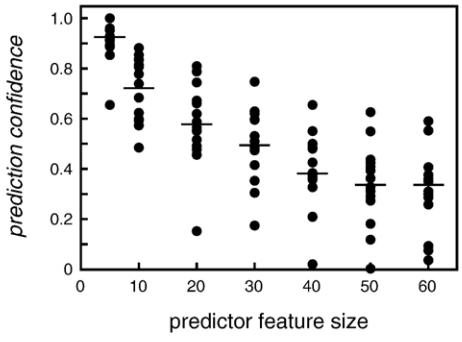


D

MCR estimate: 10-fold cross-validation, 3-fold

	cases	prediction		
		male	female	class error
male	7	7	0	0
female	7	**1**	6	**0.143**

E



F

MCR estimate: 10-fold cross-validation, 3-fold

	cases	prediction		
		male	female	class error
male	7	7	0	0
female	7	0	7	0

Fig. 4. Class comparison analysis and gender prediction in PAX and PAX-GRP samples. (A) Genes differentially expressed in PAX samples between 7 females and 7 males. Analysis was performed using the same criteria described for PBMC samples (FC >2, difference of mean >50, and $p < 0.05$) using dCHIP1.3. (B) An analysis similar to that shown in A. However, instead of 14 PAX samples, 14 PAX-GRP samples (7 females and 7 males) were used for analysis. (C) The scatter plot shows the gender prediction results within a dataset derived from 14 PAX samples (7 males and 7 females) generated using the Affymetrix U133A arrays. Accurately predicted samples are denoted by closed circles, incorrectly predicted samples are denoted by open circles. Horizontal lines denote the mean prediction confidence. (D) MCR estimation of PAX samples based on PAM classification. The misclassification rate was estimated by running three reiterations of 10-fold cross-validation. (E) An analysis similar to that shown in C. However, instead of 14 PAX samples, 14 PAX-GRP samples (7 females and 7 males) were used for analysis. (F) MCR estimation based on PAM classification using the PAX-GRP sample set.

these findings, again leading to misclassification of one sample (Fig. 4D). These findings indicate that the overall quality of expression data is decreased in PAX samples and might therefore not be applicable to detecting slight differences in gene expression or to establish valid class predictors in samples with small differences in gene expression.

Based on the increased quality of the data generated from PAX-GRP samples, we postulated that class prediction in these samples compared to PAX samples might also be improved. Similar to PBMC and PAX samples we first used the supervised learning model [3]. Leave-one-out cross-validation with predictors comprising 5 to 60 features was established to predict gender in PAX-GRP samples. In contrast to PAX samples and similar to PBMC samples, PAX-GRP samples were always assigned correctly with sufficient mean prediction confidence (Fig. 4E). Furthermore, MCR estimation also resulted in correct class prediction for each individual (Fig. 4F). Taken together, gene expression analysis of whole-blood samples prepared by the PAX-GRP method can be successfully used to build class predictors to differentiate two classes of even small differences in overall gene expression.

Comparability of PAX-GRP samples assessed on Affymetrix and Illumina arrays

In multicenter trials, large numbers of samples have to be analyzed. Therefore, inexpensive technology is a prerequisite for general clinical use. The newly launched Illumina gene expression analysis platform [29] allows higher throughput with significantly lower costs (~50% less) in comparison to other currently used microarray systems. A further critical point is the amount of total RNA required for microarray analysis in a clinical setting. For example, in comparison to the Affymetrix technology for which a minimum of 5 μg of total RNA is needed for target cRNA preparation with a single round of in vitro transcription, only 50–500 ng total RNA is required for the Illumina platform. Since the overall performance of microarray technology is also dependent on the quality of the sample preparation, we tested whether the Illumina technology could be combined with RNA samples prepared by the PAXgene method followed by globin mRNA reduction. For comparison between the results obtained on the Affymetrix platform and the Illumina platform, we assessed gene expression on Illumina Sentrix humanref-8 BeadChip arrays targeting 24,136 transcripts in the same PAX-GRP samples that were used for the analysis on the Affymetrix platform (U133A array). Next, platforms were cross-annotated using gene symbols for the respective genes. The Illumina Sentrix humanref-8 BeadChip harbors 24,136 probes representative of 18,957 annotated genes in comparison to the U133A Affymetrix array interrogating 12,846 well-annotated genes using 22,215 probes. Commonly represented on both platforms were 11,398 genes as defined by an existing gene symbol. Overall, the large majority of well-annotated genes on the U133A array is also represented on the newly introduced Illumina Sentrix humanref-8 BeadChip array.

Applying stringent criteria to determine present calls on the Illumina arrays, an average of 46.6 ± 3.7% of the genes demonstrated a significant hybridization signal compared to background signal within the 14 PAX-GRP samples. This is similar to the average present call rate observed on the Affymetrix platform. For comparison of the two platforms, a gene was defined as “expressed” when at least one probe (Illumina) or one probe set (Affymetrix) representing a corresponding gene was called present in all 14 PAX-GRP samples. Accordingly, within the shared gene space 5930 genes were expressed on the Sentrix humanref-8 BeadChip arrays, while only 4438 genes were expressed on Affymetrix U133A arrays (Fig. 5A). A majority of genes called expressed in PAX-GRP samples were detected on both platforms (3979 genes). To investigate whether one or both techniques introduce variability due to technical reasons, we compared the coefficient of variation of the data assessed by Affymetrix arrays versus those assessed by Illumina arrays. When calculating the coefficient of variation of all well-annotated genes represented on both platforms there was a slightly lower coefficient of variation in samples analyzed on the Illumina platform (Fig. 5B). This decreased variability may be due to the higher stringency of the 50-mer-based Illumina BeadChip array compared to the 25-mer-based Affymetrix array.

To visualize the impact of the array platform on the expression profile we performed dimension reduction using PCA. To compare both sample groups directly we set the maximum score for each sample to 1. When performing PCA with the 14 PAX-GRP samples analyzed by Illumina and Affymetrix microarrays no distinction between the two platforms was found, indicating that gene expression data generated on both platforms exhibit good comparability (Fig. 5C).

Correct gender prediction in PAX-GRP samples analyzed on high-density BeadChip arrays

For comparison analysis and establishment of class predictors we primarily used the complete gene set (18,957 genes) interrogated by the Sentrix humanref-8 BeadChip array. Class comparison analysis using the same filtering criteria as described above revealed a total of 10 transcripts to be differentially expressed (Fig. 6A). Five of these transcripts (*RPS4Y1, DDX3Y, SMCY, EIF1AY, CYorf15B*) were also detected as differentially expressed between females and males in PBMC samples (see Fig. 2A). Probes to detect transcripts derived from the genes *PHACS*and *RPS4Y2* are not represented on the Affymetrix U133A array, while two not yet annotated sequences (*AA628440, AV699347*) are not represented on the Illumina Sentrix humanref-8 bead array. Whereas granulysin (*GNLY*) was detected to be differentially expressed only in PAX-GRP samples (irrespective of the platform used), *USP9Y* (ubiquitin-specific protease 9, Y chromosome) was detected to be differentially expressed only in PBMC samples.

Next we analyzed gender predictability in this dataset using the same algorithms as for the Affymetrix dataset. When applying leave-one-out cross-validation, predictors with up to 20 features classified all samples correctly (Fig. 6B). Only with

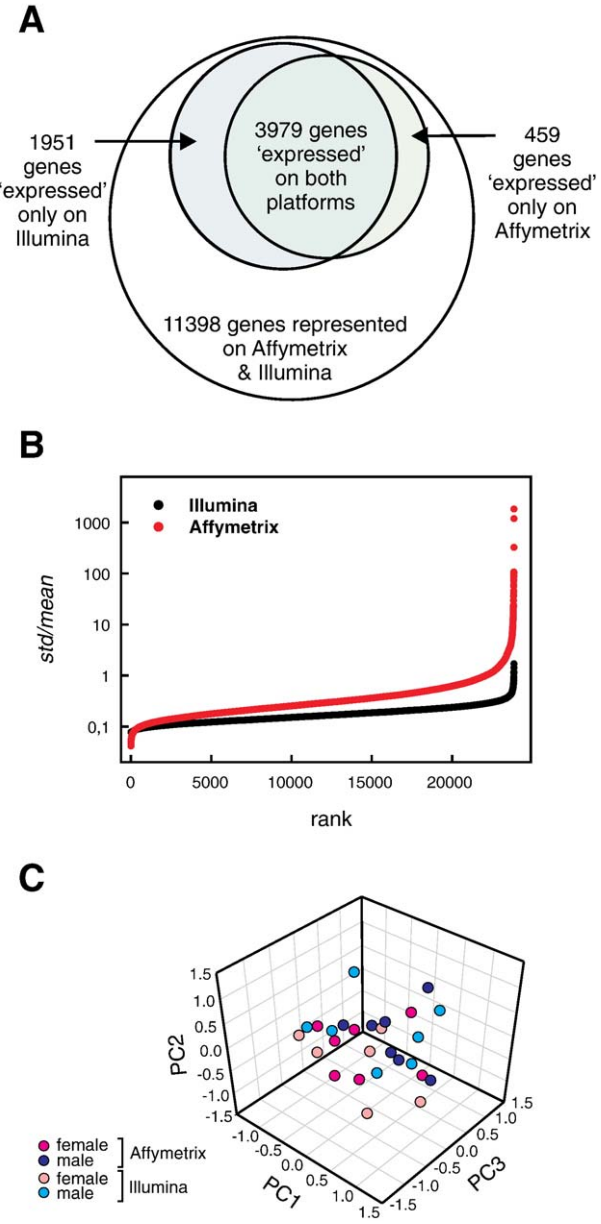


Fig. 5. Platform comparison for PAX-GRP samples using Affymetrix or Illumina arrays. (A) Venn diagram illustrating the gene space (11,398 genes) shared by the Affymetrix U133A array and the Illumina Sentrix humanref-8 bead array. Shown are the genes expressed on either both or only one platform within the common gene space. (B) All well-annotated genes represented on both platforms were taken to calculate the coefficient of variation. Standard deviation of the intensity of each probe (Illumina arrays) or probe set (Affymetrix arrays) was calculated and divided by the mean intensity. Resulting coefficients of variation were ranked and plotted. A probe set may be plotted twice as matching with two different bead probe sequences. (C) Principal components analysis with 14 PAX-GRP samples analyzed on the Illumina and the Affymetrix platform. Analysis was performed using R software. Scores for each case were calculated and the first three components were plotted. Different numbers of probes were used for PCA on the Illumina and Affymetrix platform. Therefore, scores were normalized to a maximum of 1 to compare both groups.

higher feature sizes was one sample misclassified. While PAX-GRP samples analyzed using Affymetrix U133A arrays showed no misclassification at higher feature sizes, these differences were not significant as assessed with Fisher's exact test (significance level for comparison Illumina vs. Affymetrix $p = 0.2$). In contrast, when MCR estimation [27] was used, a perfect gender prediction was performed in all cases analyzed on the Illumina platform (Fig. 6C).

Performing class prediction within the shared gene space of 11,398 genes, misclassification was observed in PAX-GRP samples assessed on Affymetrix arrays only when using predictors with high feature size. In contrast, no misclassification was observed with predictors of up to 60 features when assessing the same PAX-GRP samples on Illumina arrays within the shared gene space (Supplemental Fig. 1). Again, these differences were not statistically significant.

Taken together, the overall performance of PAX-GRP samples analyzed on the Illumina platform was comparable to that of the Affymetrix platform. The Illumina system shows adequate sensitivity to discriminate between females and males, therefore allowing the assessment of very small biological differences between two groups. These findings together with the merits of low cost per sample and very low requirements of total RNA for microarray analysis reveal the Illumina platform as an attractive tool to enable cost-effective blood-based genomics studies.

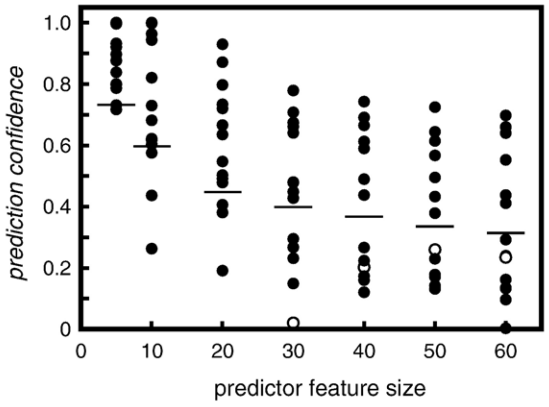
Discussion

Transcriptome analysis of peripheral blood is a very promising approach to determining disease and treatment outcome [11,15–18,23,28]. Unfortunately, current approaches based on the assessment of gene signatures in PBMC are applicable to multicenter clinical trials or even routine medical use only when PBMC are isolated as soon as possible after blood withdrawal at the clinical site, which demands a substantial organizational effort [25]. Furthermore, we have demonstrated recently that samples that cannot be processed immediately for gene expression profiling experiments yield significantly reduced data quality [23]. To circumvent such pitfalls, stabilization techniques for gene expression profiles in whole-blood samples have been introduced [24]. However, these techniques are associated with other technical problems of genome-wide transcriptome analysis such as significantly reduced overall present call rates, inferior data quality as assessed by increased intragroup variance [23], or decreased concordance in gene expression compared to expression profiles derived from PBMC samples [25]. Another important hurdle prohibiting large-scale clinical trials are the still significant costs of genome-wide transcriptome analyses [30]. Therefore, a method that could be used in large-scale multicenter trials or even for routine use would require the following: (1) a highly standardized and simple method for sample collection and processing with minimal hands-on time and without introducing ex vivo changes, (2) a method for genome-wide transcriptome analysis that can discriminate very small expression differences between two classes, (3) a method for which only small amounts of RNA are needed to minimize the volume of blood required for analysis, and (4) a cost-effective transcriptome technology allowing analysis of higher sample numbers at reasonable cost. Here we have addressed

A

Gene Symbol	Accession Number	Illumina probe ID	mean expression		difference of means	Chromosome
			female	male		
RPS4Y	NM_001008.2	GI_17981706-S	38.87	3812.42	3773.55	Y
RPS4Y2	NM_138963.1	GI_20373176-S	334.39	2822.12	2487.72	Y
GNLY	NM_006433.2	GI_7108343-A	112.97	2807.93	1694.95	2
GNLY	NM_006433.2	GI_7108343-I	465.61	1062.09	596.48	2
PHACS	NM_032592.1	GI_14211920-S	198.70	407.58	208.88	11
EIF1AY	NM_004681.2	GI_33356162-S	38.54	244.26	205.71	Y
GNLY	NM_012483.1	GI_7108345-I	139.71	296.00	156.29	2
SMCY	NM_004653.2	GI_33356559-S	67.11	220.13	153.02	Y
CYorf15B	NM_032576.1	GI_14211892-S	51.82	188.76	136.94	Y
DDX3Y	NM_004660.2	GI_13514808-S	50.87	135.64	84.76	X

B



C

MCR estimate: 10-fold cross-validation, 3-fold

	cases	prediction		
		male	female	class error
male	7	7	0	0
female	7	0	7	0

Fig. 6. Gender prediction in PAX-GRP samples using Sentrix humanref-8 bead arrays. (A) Genes differentially expressed in PAX-GRP samples between 7 females and 7 males as assessed using the Illumina Sentrix humanref-8 bead arrays. Data were primarily assessed using Illumina software. Date were imported into dCHIP1.3 and normalized and class comparison analysis was performed (FC >2, difference of mean >50, and $p < 0.05$). Genes significantly different according to the filtering criteria are listed. (B) The scatter plot shows the gender prediction results within a dataset derived from 14 PAX-GRP samples (7 females and 7 males) generated using the Illumina platform. Accurately predicted samples are denoted by closed circles, incorrectly predicted samples are denoted by open circles. Horizontal lines denote the mean prediction confidence. (C) MCR estimation based on PAM classification. The misclassification rate was estimated by running three reiterations of 10-fold cross-validation.

these issues and provide a new approach combining (1) RNA stabilization using PAXgene, (2) globin mRNA reduction, and (3) the newly introduced cost-effective Illumina Sentrix BeadChip array for genome-wide transcriptome analysis of small whole-blood-derived RNA samples (as small as 100 ng/sample). We demonstrate that this methodology leads to high-quality data with low intragroup variance and a present call rate comparable to currently used methodology assessing PBMC. More important, this method can be successfully applied to build class predictors even for classes of very small differences in gene expression such as gender–associated expression differences in blood cells [28]. We therefore propose this method to be applied to peripheral blood in large-scale multicenter clinical trials assessing disease or treatment outcome.

We and others have previously observed that RNA derived from whole-blood samples exhibits a reduced data quality after RNA stabilization leading to high variability and reduced sensitivity toward leukocyte-derived transcripts [23,26]. The abundant expression of RNAs derived from globin transcripts was the most likely explanation for poor data quality of whole-blood samples prepared with PAXgene. In fact, when reducing globin mRNA transcripts prior in vitro transcription (PAX-GRP samples) both intragroup variance and present call rate were greatly improved. While we used RNase H digestion of globin oligonucleotides/globin mRNA hybrids in our study an

alternative method to remove globin transcripts has been introduced recently. Current experiments are addressing whether this method (1) allows removal of globin transcripts from smaller amounts of initial total RNA input and (2) results in data of similar quality with low intragroup variance, sufficient present call rate, and higher sensitivity for class predictor development (S. Debey, unpublished results).

The still rather large amount of RNA required for genome-wide transcriptome analysis using currently commercially available microarray technologies and their still high expenses have so far prohibited applying these technologies to large-scale clinical trials for the development of predictive signatures for disease or treatment outcome. The newly introduced Illumina BeadChip array platform is significantly more cost-effective (~50% less) than other currently available technologies. Moreover, the small amount of 100 ng per sample of RNA needed to perform a genome-wide expression analysis on this platform is of great benefit for clinical trials since it allows one to reduce significantly the blood sample size for each patient at any given time point during a clinical study. As shown here for gender differences, stabilized RNA from peripheral blood combined with globin mRNA reduction and use of the Illumina BeadChip technology is most appropriate to detect small differences between two defined groups.

The importance of standardization of blood sample handling and processing procedures as provided here can be further illustrated by reassessing previously published and publicly available data derived from PBMC. For this purpose we have established a predictor for delayed sample handling of PBMC samples (T. Zander, unpublished results). Many genes associated with this predictor are related to hypoxia [23]. When applying this predictor to previously published datasets, a large number of these samples was classified similar to samples assessed after prolonged handling time (T. Zander, unpublished results). Under these circumstances signatures related to real pathophysiology and technically induced changes cannot be distinguished any more, making it difficult to interpret such datasets correctly.

We conclude that the prerequisites—reproducibility, feasibility, and affordability—for using gene expression profiling of blood in large clinical multicenter trials can be met by combining PAXgene technology to stabilize expression profiles during sample processing and transportation, reducing globin mRNA to increase quality of expression data, using only a single round of in vitro transcription and employing affordable microarray technology, which still complies with high industry standards. Having these techniques in place will now allow the widespread use of gene expression profiling in validation studies and further clinical applications.

Material and methods

Sample collection and RNA preparation

Blood samples from apparently healthy female or male blood donors were collected in PAXgene collection tubes (Qiagen, Hilden, Germany) after written informed consent had been obtained and following approval by the institutional review board. Age, body mass index, and time of blood draw showed no significant differences between females and males. Alternatively, PBMC were prepared following previously published methods [23]. RNA was prepared with the PAXgene Blood RNA Kit (Qiagen) according to the manufacturer's instruction including an optional DNase digestion step. In the case of PBMC, RNA was prepared by the Trizol method as previously described [23].

Globin reduction protocol

Globin reduction was performed according to an Affymetrix technical note [31]. To reduce globin mRNA 8 μg of total RNA prepared with the PAXgene Blood RNA Kit was incubated in 10 mM Tris–HCl, pH 7.6, 20 mM KCl (total volume 10 μl), with the globin-specific oligos α1, 5′-TGCAGGAAGGGGAG-GAGGGGCTG-3′ (1.5 μM); α2, 5′-TGCAAGGAGGGGAGGAGGGCCCG-3′ (1.5 μM); β, 5′-CCCCAGTTTAGTAGTTGGACTTAGGG-3′ (4 μM) for 5 min at 70°C and then cooled to 4°C. Globin mRNA–DNA hybrids were removed by adding 10 μl of RNase H digestion mix (2 units RNase H, 20 mM Tris–HCl, pH 7.6, 2 mM DTT, 4 mM $MgCl_2$) and incubating for 10 min at 37°C. Samples were cooled to 4°C and reactions were stopped by adding 1 μl of 0.5 M EDTA. RNase H-treated RNA samples were then purified with IVT cRNA Cleanup spin columns from the Affymetrix GeneChip Sample Cleanup Module according to the manufacturer's protocol and globin-depleted RNA was eluted with 14 μl H_2O. The integrity of the RNase H-treated RNA samples was assessed by Agilent 2100 bioanalyzer analysis prior to in vitro transcription (IVT) and revealed RNA of normal quality.

Microarray procedure using Affymetrix U133A arrays

Three to five micrograms of globin-depleted total RNA was used to generate double-stranded cDNA with a $T7(dT)_{24}$-oligonucleotide primer performed with a One Cycle cDNA Synthesis Kit (Affymetrix, Santa Clara, CA, USA). After purification with the Sample Cleanup Module (Affymetrix), cDNA served as a template to prepare biotinylated cRNA via IVT, using the IVT Labeling Kit (Affymetrix). The labeled cRNA transcripts were purified using the Sample Cleanup Module and assessed for quantity and quality by UV spectroscopy and electrophoresis. Fragmentation of cRNA transcripts, hybridization, and scanning of the high-density oligonucleotide microarrays (HG-U133A arrays; Affymetrix) were performed according to the manufacturer's GeneChip Expression Analysis Technical Manual (Affymetrix).

Microarray procedure using Illumina Sentrix humanref-8 arrays

One hundred nanograms of globin-depleted total RNA was used for first- and second-strand cDNA synthesis performed with a MessageAmp Kit (Ambion, Cambridgeshire, UK) using quarter volumes. Purification of cDNA was performed with a QiaQuick PCR Purification Plate on a QIAvac 96 vacuum manifold (Qiagen, Chatsworth, CA, USA). Biotin-labeling of cRNA was performed with biotin-16–UTP (Perkin–Elmer) and the MessageAmp Kit, and cRNA was purified using an RNeasy Plate (Qiagen) on a QIAvac 96 vacuum manifold. Hybridization and scanning of Sentrix humanref-8 expression bead arrays were performed according to the manufacturer's instructions (Illumina, San Diego, CA, USA).

Data extraction and statistics

For Affymetrix data collection and assessment, we used Affymetrix Microarray Analysis Suite version 5.0 (MAS5.0) and dCHIP 1.3. According to standard microarray analysis methods, for data analysis using MAS5.0 the fluorescence intensity of each array was scaled to an overall intensity of 100 to enable comparison of all hybridization data. In dCHIP Affymetrix CEL files were imported and normalized to the median intensity using the Perfect Match model [32]. Data extraction for Illumina was performed using the AnEx software and dCHIP 1.3. In dCHIP 1.3 Illumina raw expression values were imported as external data and normalized to the median intensity. Differentially expressed genes were assessed using dCHIP1.3. For variance analysis and Fisher's exact test, the R software was used. For comparison we also analyzed a previously

published dataset publicly available in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/; accession number GSE2328).

Comparison between the Affymetrix U133A array and the Illumina Sentrix humanref-8 bead arrays

For comparison of the two platforms we cross-annotated the Affymetrix probe set IDs as well as the Illumina oligonucleotide IDs to gene symbols. Data were retrieved from the most current annotation file from the Affymetrix NetAffx Web site (http://www.affymetrix.com/support/technical/byproduct.affx?product = hgu133) and the annotation file supplied by Illumina. Gene symbols and alternative variants of the gene symbol (called "synonyms" within the Illumina annotation file) were taken for cross-annotation. By using gene symbols for cross-annotation it is possible that differentially spliced transcripts that are interrogated differentially on both platforms are merged within one gene symbol. Cross-annotation was performed using a relational MS Access database that linked Affymetrix probe sets and Illumina oligonucleotide IDs with gene symbols.

After cross-annotation the variance for each bead probe and the respective probe set was calculated for the Illumina or Affymetrix samples and divided by the mean expression value for the respective bead probe or probe set. The obtained coefficient of variation was sorted and ranked.

Principal components were calculated using R software and scores for each case were plotted. As size of the scores depends on the number of variables used for PCA, we normalized scores for each group (Illumina or Affymetrix) to a maximum of 1.

Calculation of present calls for Illumina data

Image analysis and data extraction were performed using Illumina AnEx software. Briefly, each sequence type is represented by an average of 30 beads on the array. Bead signals are computed with weighted averages of pixel intensities, and local background is subtracted [33]. Sequence-type signal is calculated by averaging corresponding bead signals with outliers removed (using median absolute deviation). Detection p values were computed using a model based on signals of negative controls as described by Fan et al. [34]. To correct for multiple testing a Bonferroni correction on a p value of 0.05 was applied. Consequently those signals differing from the background with a p value of 0.05/24,115 were called present.

Prediction of gender

For class predictions using several different algorithms expression values from the respective datasets were exported from dCHIP1.3. First GeneCluster2 software [3] was used and the data were preprocessed as follows: Signal intensities below 20 and above 16,000 (Affymetrix) or below 5 and above 5000 (Illumina) were set to the respective minimum and maximum values, and data were normalized to mean = 0 and variance = 1. In the supervised learning approach predictors were generated and tested by the leave-one-out cross-validation model using the weighted voting algorithm. Furthermore, nested cross-validation based on PAM classification [27] was used to predict the class membership [35] as implemented in the R package MCRestimate. Here the misclassification rate was estimated by running three reiterations of 10-fold cross-validation. Generation of class predictors for gender was also performed in the gene space shared between Affymetrix and Illumina arrays.

Acknowledgments

We thank our colleagues at the Division of Transfusion Medicine for their technical support and we are grateful to all the blood donors for donating blood for this study. We also thank Daniela Eggle for performing statistical analyses. This work was supported mainly by the Alexander von Humboldt Foundation via a Sofja-Kovalevskaja Award to J.L.S. T.Z. was supported by the Frauke-Weiskam-Christel Ruranski Foundation. J.L.S. is a member of the NGFN.

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ygeno.2005.11.010.

References

[1] M. Branca, Genetics and medicine: putting gene arrays to the test, Science 300 (2003) 238.
[2] C.M. Schubert, Microarray to be used as routine clinical screen, Nat. Med. 9 (2003) 9.
[3] T.R. Golub, et al., Molecular classification of cancer: class discovery and class prediction by gene expression monitoring, Science 286 (1999) 531–537.
[4] A.A. Alizadeh, et al., Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling, Nature 403 (2000) 503–511.
[5] M. Bittner, et al., Molecular classification of cutaneous malignant melanoma by gene expression profiling, Nature 406 (2000) 536–540.
[6] S.A. Armstrong, et al., MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia, Nat. Genet. 30 (2002) 41–47.
[7] M.J. van de Vijver, et al., A gene-expression signature as a predictor of survival in breast cancer, N. Engl. J. Med. 347 (2002) 1999–2009.
[8] M.A. Shipp, et al., Diffuse large B-cell lymphoma outcome prediction by gene-expression profiling and supervised machine learning, Nat. Med. 8 (2002) 68–74.
[9] A. Rosenwald, et al., The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma, N. Engl. J. Med. 346 (2002) 1937–1947.
[10] L. Bullinger, et al., Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia, N. Engl. J. Med. 350 (2004) 1605–1616.
[11] P.J. Valk, et al., Prognostically useful gene-expression profiles in acute myeloid leukemia, N. Engl. J. Med. 350 (2004) 1617–1628.
[12] D.L. Gerhold, R.V. Jensen, S.R. Gullans, Better therapeutics through microarrays, Nat. Genet. 32 (2002) 547–551 (Suppl).
[13] U. Scherf, et al., A gene expression database for the molecular pharmacology of cancer, Nat. Genet. 24 (2000) 236–244.
[14] E.C. Gunther, et al., Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro, Proc. Natl. Acad. Sci. USA 100 (2003) 9608–9613.
[15] L. Bennett, et al., Interferon and granulopoiesis signatures in systemic lupus erythematosus blood, J. Exp. Med. 197 (2003) 711–723.
[16] A.K. Palucka, et al., Cross-regulation of TNF and IFN-alpha in autoimmune diseases, Proc. Natl. Acad. Sci. USA 102 (2005) 3372–3377.
[17] V. Pascual, et al., Role of interleukin-1 (IL-1) in the pathogenesis of systemic onset juvenile idiopathic arthritis and clinical response to IL-1 blockade, J. Exp. Med. 201 (2005) 1479–1486.
[18] M.E. Burczynski, et al., Transcriptional profiles in peripheral blood mononuclear cells prognostic of clinical outcomes in patients with advanced renal cell carcinoma, Clin. Cancer Res. 11 (2005) 1181–1189.
[19] A. Brazma, et al., Minimum information about a microarray experiment (MIAME)—toward standards for microarray data, Nat. Genet. 29 (2001) 365–371.
[20] T. Bammler, et al., Standardizing global gene expression analysis between laboratories and across platforms, Nat. Methods 2 (2005) 351–356.
[21] R.A. Irizarry, et al., Multiple-laboratory comparison of microarray platforms, Nat. Methods 2 (2005) 345–350.
[22] J.E. Larkin, et al., Independence and reproducibility across microarray platforms, Nat. Methods 2 (2005) 337–344.
[23] S. Debey, et al., Comparison of different isolation techniques prior to gene expression profiling of blood derived cells: impact on physiological

responses, on overall expression and the role of different cell types, Pharmacogenom. J. 4 (2004) 193–207.
[24] L. Rainen, et al., Stabilization of mRNA expression in whole blood samples, Clin. Chem. 48 (2002) 1883–1890.
[25] J.P. Cobb, et al., Application of genome-wide expression analysis to human health and disease, Proc. Natl. Acad. Sci. USA 102 (2005) 4801–4806.
[26] R.J. Feezor, et al., Whole blood and leukocyte RNA isolation for gene expression analyses, Physiol. Genom. 19 (2004) 247–254.
[27] R. Tibshirani, et al., Diagnosis of multiple cancer types by shrunken centroids of gene expression, Proc. Natl. Acad. Sci. USA 99 (2002) 6567–6572.
[28] A.R. Whitney, et al., Individuality and variation in gene expression patterns in human blood, Proc. Natl. Acad. Sci. USA 100 (2003) 1896–1901.
[29] K. Kuhn, et al., A novel, high-performance random array platform for quantitative gene expression profiling, Genome Res. 14 (2004) 2347–2356.
[30] J. Bell, Predicting disease using genomics, 429 (2004) 453–456.
[31] Affymetrix Technical Note, Globin reduction protocol: a method for processing whole blood RNA samples for improved array results (2003).
[32] C. Li, W.H. Wong, Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection, Proc. Natl. Acad. Sci. USA 98 (2001) 31–36.
[33] V.L. Galinsky, Automatic registration of microarray images: II. Hexagonal grid, Bioinformatics 19 (2003) 1832–1836.
[34] J.B. Fan, et al., A versatile assay for high-throughput gene expression profiling on universal array matrices, Genome Res. 14 (2004) 878–885.
[35] M. Ruschhaupt, et al., A compendium to ensure computational reproducibility in high-dimensional classification tasks, Stat. Appl. Genet. Mol. Biol. 3 (2004) 37.