Getting the Right Cells to the Array: Gene Expression Microarray Analysis of Cell Mixtures and Sorted Cells

Peter Szaniszlo,[1] Nan Wang,[2] Mala Sinha,[3,5] Lisa M. Reece,[2] James W. Van Hook,[4] Bruce A. Luxon,[3,5] and James F. Leary[1–3,5]

[1]Department of Microbiology and Immunology, University of Texas Medical Branch at Galveston, Galveston, Texas
[2]Department of Internal Medicine, University of Texas Medical Branch at Galveston, Galveston, Texas
[3]Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch at Galveston, Galveston, Texas
[4]Department of Obstetrics and Gynecology, University of Texas Medical Branch at Galveston, Galveston, Texas
[5]Bioinformatics Program, University of Texas Medical Branch at Galveston, Galveston, Texas

Background: Most biological samples are cell mixtures. Some basic questions are still unanswered about analyzing these heterogeneous samples using gene expression microarray technology (MAT). How meaningful is a cell mixture's overall gene expression profile (GEP)? Is it necessary to purify the cells of interest before microarray analysis, and how much purity is needed? How much does the purification itself distort the GEP, and how well can the GEP of a small cell subset be recovered?

Methods: Model cell mixtures with different cell ratios were analyzed by both spotted and Affymetrix MAT. GEP distortion during cell purification and GEPs of purified cells were studied. CD34+ cord blood cells were purified and analyzed by MAT.

Results: GEPs for mixed cell populations were found to mirror the cell ratios in the mixture. Over 75% pure samples were indistinguishable from pure cells by their overall GEP. Cell purification preserved the GEP. The GEPs of small cell subsets could be accurately recovered by cell sorting both from model cell mixtures and from cord blood.

Conclusions: Purification of small cell subsets from a mixture prior to MAT is necessary for meaningful results. Even completely hidden GEPs of small cell subpopulations can be recovered by cell sorting.   © 2004 Wiley-Liss, Inc.

Key terms: microarray; gene expression profile; cell sorting; flow cytometry; immunomagnetic separation; cell fixation; scatterplot; stem/progenitor cells; genomics; cytomics

Gene expression microarray technology (MAT) is a rapidly developing analytical tool in basic and clinical research. The number of publications using microarrays (MAs) has been growing exponentially over the past four to five years (1,2). In the 1990s, MAT was considered to be one of the most promising new tools for the science of the 21st century, when fast, high-throughput technologies would be used to grasp the complexity of biological systems (3–5). Since MAT is potentially capable of looking at all of the cellular processes at the mRNA level at a given moment, it was expected to deliver not only a comprehensive quantity of data about the transcriptional level, but also to shed light on novel processes, pathways, and molecular interactions in the living cell. MAT data, providing "freeze-frame" views of the transcriptome, could help us understand the role and weight of known, and yet to be discovered, molecular mechanisms in the "big picture" (6–8).

In clinical research, MAT is expected to help us better understand the molecular basis of diseases, and the difference between healthy and diseased cells, tissues, and organs, giving us clues about potential treatment (9–11).

At the same time, microarrays (MAs) can be used to monitor the effects of these treatments, especially of new (and old) drugs (12–14). The unique view that MAT offers of the cells and tissues also generated high hopes in the field of pathology. It was expected to revolutionize and automate the analysis of tissue sections and the classification of diseases, disease subtypes, and disease stages. Surprisingly, the latter areas (especially the analysis and classification of different types, subtypes, and stages of tumors) are the ones in which MAT has proven to be immediately useful, delivering very promising and convincing results (15–17).

However, in basic research, initial MA studies often caused disappointment, failing to generate or confirm new hypotheses. On many occasions, MAT has generated more confusion than comprehension, and more questions than answers (14,18,19). Despite this fact, MAT has improved greatly in recent years. Most of the early problems (e.g., reproducibility, sensitivity, high background, standardization, preparation of samples, data analysis) have been addressed, and greatly improved; however, MAT still has problems breaking out of a relatively narrow field of applicability. One of the main sources of the remaining problems is that while MA analysis requires 0.5–5.0 million cells per sample, biological systems (tissues, organs) with this amount of cells are almost always mixtures of several different cell types (20)—all of which may behave differently in a given experiment. The two main exceptions are cell lines and tumors, in which one can have more than enough cells of the same type for MA analysis. Not surprisingly, these two are the main fields of successful MA studies (21–23). However, tumors only represent a narrow segment of pathological processes in humans, and immortalized cell lines have been repeatedly shown to differ significantly from in vivo cells, even the ones from which they originated (24,25). In many other MA studies of unsorted cells, success could be achieved because most of the cells in the sample behaved similarly, thus their reaction to experimental condition changes could be detected (26–28). One would suspect, though, that even in these cases, many subtle effects might have remained undetected, overshadowed by the nonreacting cells. Even major changes in gene expression levels of a minor cell subset of the mixture might have been lost in the background of the more numerous unchanged cells.

The obvious solution to the problem of cell mixtures showing mixed gene expression profiles (GEPs) is cell sorting. Delivering sorted, more homogenous cell samples to MAs is expected to produce much clearer results than studying cell mixtures. Several recent studies have taken this approach and many of them have shown promising results (29–31). However, it is still not well understood how different cell sorting methods and sample handling protocols affect the GEP (18,32,33). It is also not well known how much effect different cell types have on each other's GEPs when they are mixed together. On the other hand, in the real world of both basic and clinical applications, 100% purity of a given cell type is not always achievable or feasible. So the question is, how pure is pure enough?

To address these very basic questions, we studied the overall GEP of defined cell mixtures to model heterogeneous biological samples. For the "overall GEP," we used the unprocessed microarray readout of the cell sample with no values excluded. We evaluated the effects of cell labeling, fixation, and sorting on the overall GEP. We also analyzed how well we could recover the GEP of a pure cell type by sorting these cells from a mixture. Since different types of microarrays do not necessarily produce the same data (7,34,35), we used both spotted (Clontech, Palo Alto, CA) and short-oligonucleotide arrays (Affymetrix, Santa Clara, CA) to compare the results of the same experiments on these different platforms.

Microarray data analysis is not an obvious exercise; in fact it has developed into its own new field, with quite a few competing methods. The complexity of these methods often creates a communicational gap between the data-producing biologists and the data-analyzing mathematicians and biostatisticians (7,8,36). To avoid this gap, and to keep the presentation of our results as directly connected to the samples they represent as possible, we have deliberately used very simple, straightforward methods to compare the overall GEP of different samples, rather than any of the more sophisticated software packages existing today.

This article demonstrates why "cytomics" approaches, as discussed by Valet et al. (37) are important. Without the ability to analyze and purify cell subpopulations using cytomics technologies, much of the power of MAT is compromised or even lost.

MATERIALS AND METHODS
Cell Cultures

CEM. A human, CD4+ T-cell line (acute lymphoblastoid leukemia, ALL); obtained through the AIDS Research and Reference Reagent Program, Division of AIDS, NIAID, NIH: CEM-T4 from Dr. J.P. Jacobs. CEM cells were cultured using RPMI 1640 medium with 2 mM L-glutamine and 10% fetal bovine serum in the presence of 5% CO_2 at 37°C.

A2780. A human, CD4- epithelial cell line (human ovarian carcinoma, ECECC 93112519); kindly provided by Dr. Istvan Boldogh (Department of Microbiology and Immunology, University of Texas Medical Branch, Galveston, Texas). A2780 cells were cultured using RPMI 1640 medium with 2 mM L-glutamine and 10% fetal bovine serum in the presence of 5% CO_2 at 37°C.

KG-1a. A human, CD34+ stem-cell cell line (human bone marrow acute myelogenous leukemia, ATCC CCL-246.1); kindly provided by Dr. Brian R. Davis (Sealy Center for Molecular Hematology and Oncology, University of Texas Medical Branch, Galveston, Texas). KG-1a cells were cultured using Iscove's modified Dulbecco's medium with 4 mM L-glutamine and 20% fetal bovine serum in the presence of 5% CO_2 at 37°C.

Cell culture medium, serum, and glutamine were purchased from Gibco BRL (Grand Island, NY).

Model Cell Mixtures

Cultured CEM and A2780 cells were counted and tested for viability. Calculated volumes for 3.0×10^7 cells each were pelleted and resuspended in PBS (Gibco BRL, Grand Island, NY). Both cell suspensions were recounted and appropriate volumes of each cell type for each planned mixture were calculated. The cell mixtures were then prepared by mixing the calculated volumes of each cell suspension.

Cord Blood Cells

Human cord blood was obtained under informed consent from HIV-negative normal donors under IRB-approved protocols at the Department of Obstetrics and Gynecology, Maternal-Fetal Medicine, University of Texas Medical Branch, Galveston, Texas. Cord blood was drawn into yellow-capped vacutainer tubes (Beckman Coulter, Inc. Fullerton, CA), containing acid citrate dextrose (ACD) anticoagulant. Two to four blood samples were pooled and cord blood mononuclear cells (CBMCs) were isolated using Ficoll-Paque density gradient (Pharmacia Biotech, Piscataway, NJ) following manufacturer recommended protocols. CD34+ stem/progenitor cells were purified from CBMCs by magnetic sorting using MACS CD34 Progenitor Cell Isolation Kit (Miltenyi Biotec, Bergisch Gladbach, Germany) following the manufacturer's recommendations.

Magnetic Cell Sorting

CD34+ KG-1a cells and CD34+ cord blood stem/progenitor cells were purified from CBMCs with MACS CD34 Progenitor Cell Isolation Kit (Miltenyi Biotec, Bergisch Gladbach, Germany) following manufacturer's recommendations. CD4+ CEM cells were purified from model cell mixtures with MACS CD4 Microbeads (Miltenyi Biotec, Bergisch Gladbach, Germany) following manufacturer's recommendations.

After three consecutive rounds of column purification, the purity (typically around 95–99%) of eluded cells was analyzed by subsequent flow-cytometric analysis of an aliquot labeled with anti-CD34-PE (for CD34+ cord blood cells and KG-1a cells) or anti-CD4-PE (for CD4+ CEM cells) antibodies (Caltag Laboratories, Burlingame, CA).

Flow Cytometry Analysis and Cell Sorting

CD34+ cord blood cells and KG-1a cells were labeled with phycoerythrin (PE)-conjugated, murine, anti-CD34 antibody (Caltag Laboratories. Burlingame, CA) using factory recommended protocols. CEM cells and CEM/A2780 cell mixtures were similarly labeled with PE-conjugated, murine, anti-CD4 antibody (Caltag Laboratories. Burlingame, CA). Cells were analyzed and sorted on our custom-built High Resolution Cell Sorter (HiReCS) system (38) set up for standard fluorescence analysis. A tunable argon-ion laser tuned to a 488-nm wavelength was used in all analyses, with optical filters that were optimal for PE excitation and emission. Samples were acquired on three parameters: PE, FSC, and SSC and stored as listmode data in flow cytometry standard (FCS 2.0) file format for subsequent analysis. The program WinList 5.0™ (Verity Software House, Topsham, ME) was used for flow-cytometric data analysis.

A CEM/A2780 cell mixture containing 10% CEM cells was presorted by one round of magnetic sort using MACS CD4 Microbeads (Miltenyi Biotec, Bergisch Gladbach, Germany), as described above. The resulting CEM-enriched cell mixture was labeled with PE-conjugated, murine, anti-CD4 antibody (Caltag Laboratories, Burlingame, CA) and flow-sorted for PE-positive cells. This sort enhanced CEM purity from 70 to 95% as shown by subsequent flow-cytometric analysis of an aliquot. The resulting 95% pure sample was analyzed by Affymetrix microarray analysis as purified CEM cells.

Cell Fixation

Fluorescent antibody-labeled cells were washed once in PBS (Gibco BRL, Grand Island, NY) and resuspended in 100–200 µl PBS. The sample was mixed with 500–1,000 µl (5× the volume of PBS), –20°C-cold methanol (Sigma, St. Louis, MO) and incubated at –20°C for 5 min in the dark. Cells then were pelleted and resuspended in PBS for further processing.

Preparation of Labeled Probes and Microarray Analysis

Total RNA was isolated from all cell samples using RNAqueous™-4PCR RNA isolation kit (Ambion, Austin, TX), following manufacturer's recommendations. Within each experiment, each sample was normalized by the amount of isolated RNA. For spotted microarray analysis of 82 genes, an Atlas™ Array Trial Kit (Clontech, Palo Alto, CA) was used, following factory recommended protocols. For signal detection, a Storm 860 phosphorimager (Molecular Dynamics, Sunnyvale, CA) was used. The array images created were analyzed by Scanalyze software (Stanford University, Stanford, CA) to quantitate microarray data and produce the raw data files. For short-oligonucleotide microarray analysis of over 12,000 genes, a Gene-Chip® Human Genome U95Av2 (Affymetrix, Santa Clara, CA) was used. RNA labeling, hybridization, and scanning to produce the raw microarray data files were done by the Molecular Genomics Core Facility of the University of Texas Medical Branch at Galveston, following factory recommended protocols.

Microarray Data Analysis

Images of Affymetrix arrays were generated directly from the raw image files. The original array images were viewed, magnified, pseudocolored, and cropped using the Affymetrix Microarray Suite 5.0 software (Affymetrix, Santa Clara, CA). For array-to-array image comparison, the same segment of the Affymetrix array was always selected, inspected, and compared. Images of Clontech arrays, generated by the Storm 860 phosphorimager (Molecular Dynamics, Sunnyvale, CA) were directly used for visual array-to-array image comparison. Bar graphs, scatterplots, and regression analysis results were generated from the raw
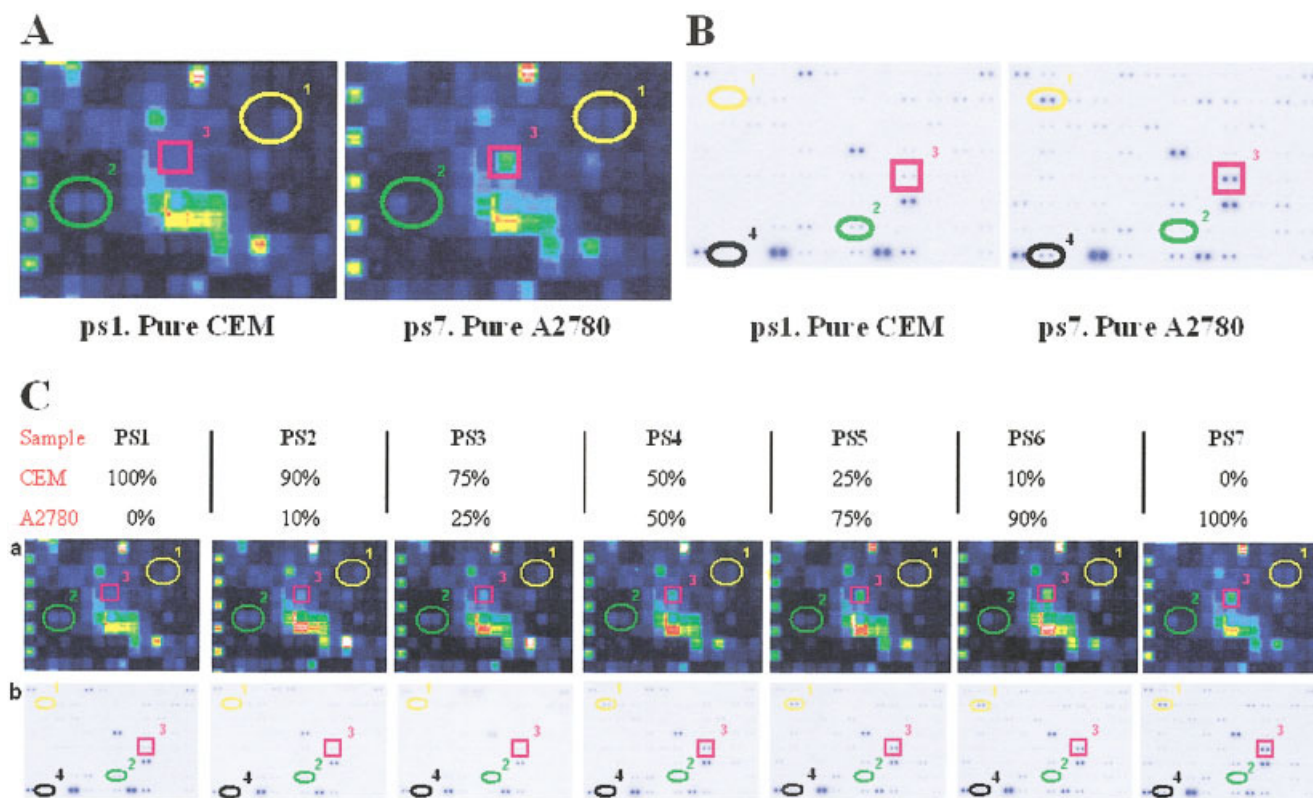
FIG. 1. Microarray images of cell mixtures. Genes expressed in only one of the cell types are in circles 1, 2, and 4. Differentially-expressed genes are in square 3. **A:** Pure CEM and A2780 cells. Cropped segments of Affymetrix images. **B:** Pure CEM and A2780 cells. Spotted array images. **C:** Corresponding Affymetrix and spotted array images of cell mixtures with different CEM/A2780 cell ratios.

data files using Microsoft Excel 2000 software (Microsoft, Redmond, WA). Trellis plots were generated using S-Plus 6.2 software (Insightful Corp., Seattle, WA). Hierarchical clustering was generated using Spotfire 7.2 data analysis software (Spotfire Inc., Cambridge, MA). For generating the heat map, raw Affymetrix data files of samples PS301–PS304 were trimmed by excluding the 2% of all genes with the highest readings and the 2% with the lowest intensity readings. The array data from the four samples were normalized by trimmed mean overall expression level. Genes that were excluded were: 1) those that were called "absent" in all four arrays by the Affymetrix software; and 2) all genes with less than two-fold expression level difference in all the possible pairwise comparisons among the four arrays. The remaining 4,848 genes were organized into a heat map by hierarchical clustering, based on gene expression levels.

RESULTS
Microarray Images of Cell Mixtures

We selected two cell lines (CEM and A2780) with characteristic similarities and differences in their GEPs, to be able to monitor their contributions to the overall GEP of each sample. First, we looked at the direct microarray images of the samples to gain an overall impression of the situation. Figure 1A shows the Affymetrix microarray im-

ages of each cell type. Throughout this study, we always inspected the same selected segment of the Affymetrix array for each sample, to visually confirm our findings. The original array images were magnified, pseudocolored, and cropped using the Affymetrix Microarray Suite software. Each small square (feature) represents a gene sequence with millions of identical 25-mer oligonucleotides attached to the array surface. The segment we used for visual analysis contained a total of 165 features. Different colors represent different, increasing, expression levels, in the following order: black, violet, blue, green, yellow, orange, red, and white. The two images in Figure 1A exhibit very similar patterns (both samples being human cell lines), but several differences can be found between them. Some sequences are only expressed in one or the other (circles 1 and 2), and some are expressed in both, but at different levels (square 3). Triplicate arrays of the same sample showed virtually identical images (not shown).

Figure 1B displays the spotted microarray images of the same two cell lines (CEM and A2780). Each pair of spots represents a gene sequence, 96 sequences in all on each spotted array. Larger and darker spots represent higher gene expression levels. Again, within the very similar gene expression patterns, there are several characteristic differences, with some genes expressed only in one of the cell

lines (circles 1, 2, and 4), and others showing differential gene expression (square 3). Spotted array images of replicate samples were virtually identical (not shown). These observations confirmed that the GEPs of the two chosen cell lines are indeed different for many selected genes, and the differences between those selected gene expression levels can be visually detected in the images of both array types.

From these two cell lines, we prepared a set of cell mixtures with gradually changing cell ratios. As Figure 1C shows, PS1 contained pure CEM cells. The CEM/A2780 ratio in PS2 was 90:10, and it was 75:25 in PS3, 50:50 in PS4, 25:75 in PS5, and 10:90 in PS6. PS7 had only A2780 cells. These gradual changes in the mixture ratio could be detected as gradually appearing/disappearing (circles 1, 2, and 4), or strengthening/weakening (square 3) genes in both the consecutive Affymetrix images and the spotted array images. These observations suggested that the GEP of a cell mixture quite accurately mirrors the ratios of the participating cell populations.

Scatterplot Analysis

After the "intuitive approach" described above, in which we visually inspected expression levels of a few single genes throughout all samples, we examined the "overall GEP" of the same samples. For a comprehensive overall comparison of two GEPs, we used simple scatterplots, in which each sample's GEP was represented on one axis, and each gene (small squares) was positioned using its expression levels in the two samples as coordinates. The two samples' GEPs were considered to be more similar when the genes lined up closer along a linear trend line (multiple linear regression analysis). To calculate this GEP similarity for each scatterplot, we used, as is commonly done, the R^2-value (multiple correlation coefficient) as a measure of "goodness-of-fit." A higher R^2-value was considered to indicate higher similarity between two GEPs. Figure 2A shows pairwise scatterplots of triplicate samples (pure A2780 cells). GEPs were produced by Affymetrix arrays. As expected, triplicate GEPs of over 12,000 genes demonstrated high similarity. Repeated triplicate experiments consistently produced R^2-values of 0.97–0.995. Replicate spotted array results were similar (not shown). Based on these results, we considered samples with R^2-values over 0.97 identical.

After establishing the approximate range of experimental error, we examined the overall GEPs of our model cell mixtures. Figure 2B compares samples PS1 and PS2, a pure CEM sample to a 90% pure one. With an R^2-value of 0.9842, the overall GEPs of these two samples seemed to be just as identical as if they were the same sample. Similarly, the R^2-value for comparing pure A2780 cells to 90% pure ones (PS7 versus PS6) was found to be 0.9826 (not shown). Obviously, this does not mean that the expression levels of each individual gene in these samples were practically the same, but it does suggest that for the vast majority of genes a 90% pure sample could be used as a "pure sample" in terms of an "overall" GEP.

Comparing the pure CEM sample to the pure A2780 sample (Fig. 2C) produces a scatterplot with an R^2-value of 0.7532. Although we selected the two model cell lines to be very different (CEM is a human T-lymphoblastoid cell line while A2780 cells are derived from human ovarian carcinoma cells), both are human tumor cells with significant similarities in their GEPs. In our model, this R^2-value of 0.7532 is "as bad as it gets." This value probably corresponds to the degree of similarity one sees with very different human cell types, due to the necessary expression of housekeeping and fundamental genes needed for routine growth and survival.

Figure 2D shows all possible pairwise scatterplots of the seven samples in a Trellis plot format. Each scatterplot is positioned in the row and column of the two samples from which it was made (e.g., the second scatterplot in the fourth column compares samples PS2 and PS4). By following along a given row or column, it is visually easy to see the effect of the purity of cell mixtures on GEPs. The overall pattern was strikingly similar to the one observed in Figure 1C. When the R^2-values of all these scatterplots were calculated (data not shown), we found that they also followed the same tendency; the farther two samples were from each other in the Trellis plot, the lower R^2-value they produced when compared. Interestingly, the 75% mix compared to the pure sample (PS3 versus PS1 and PS5 versus PS7) produced R^2-values in the 0.96–0.975 range, which was established earlier as the estimated borderline for this method to be able to "tell the difference" between two samples. However, a very small number of genes was found to fall farther away from the regression line. These outlier genes showed significantly different expression levels within the otherwise almost identical samples. All scatterplots in which the cell ratios of the two samples compared were more than 25% different had a much lower R^2-value. These results suggest that a sample needs to be over 75% pure to truly represent the GEP of the given cell population.

Modeling Genes in "Real" Samples

After assessing the overall GEPs of cell mixtures, we modeled a possible "worst case scenario in real biological samples," in which the background cells would strongly express the investigated genes. We selected a set of individual genes that were moderately expressed in CEM cells (target cells) and strongly expressed in A2780 cells (background cells). As shown in Figure 2E, the expression level of these genes followed the ratio of A2780 cells in the mixtures. Although CEM cells also expressed these genes, their effect on the sum of the expression levels was totally washed out by the background cells. To monitor changes in the expression levels of these genes in CEM cells, even a 90% pure sample is not pure enough.

A possible "best case scenario" was modeled in Figure 2F. Again, the selected genes were moderately expressed in CEM cells, but A2780 cells did not express these genes at all. Now it was possible to monitor the expression levels of these genes in the target cells, even in the presence of nine times more background cells. The charts
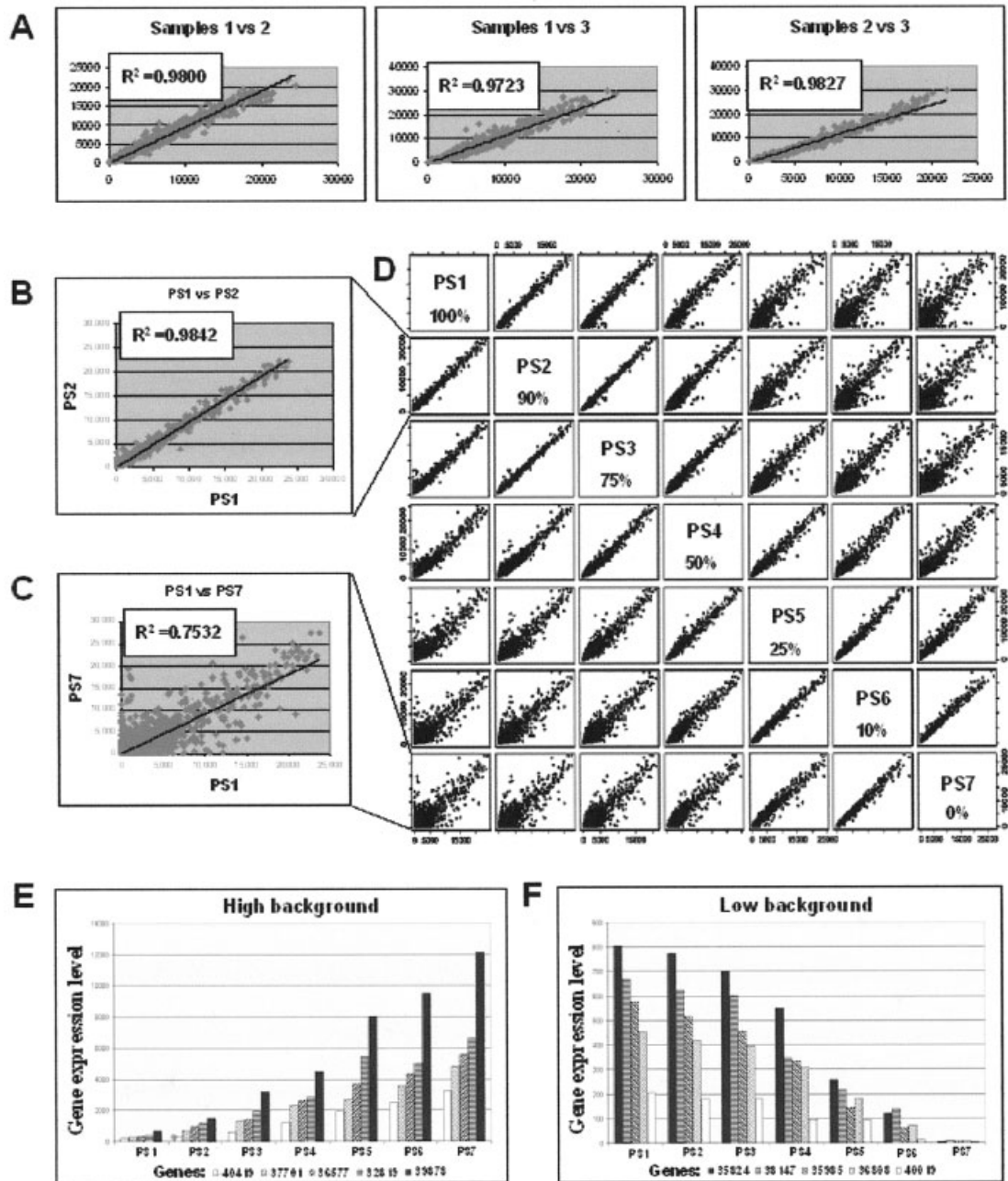
FIG. 2. Gene expression profile comparisons of cell mixtures and pure samples. **A:** Pairwise scatterplot analysis of triplicate A2780 samples. **B:** Scatterplot analysis comparing pure CEM cells and a 90% mixture of CEM/A2780 cells. **C:** Scatterplot analysis comparing pure CEM and pure A2780 cells. **D:** Trellis plot of all pairwise scatterplots, comparing samples PS1–PS7. Each scatterplot is positioned in the row and column of the two samples from which it is made. **E:** Expression levels of genes that are moderately expressed in CEM cells and strongly expressed in A2780 cells. **F:** Expression levels of genes that are moderately expressed in CEM cells and not expressed in A2780 cells.

in Figures 2E–F demonstrate Affymetrix data; very similar charts could be created from spotted array results (not shown).

Effects of Sample Handling

The results displayed in Figures 1 and 2 confirmed the notion that, in order to investigate the GEP of a cell subset in a mixture, these cells need to be purified first. The next question was, how much distortion would the purification process itself (including cell fixation, labeling, and sorting) introduce into the studied GEP? We tested methanol-fixation, since alcohols are known to preserve nucleic acids better than cross-linking agents (39). Figure 3A shows a GEP scatterplot of live, unlabeled CEM cells and
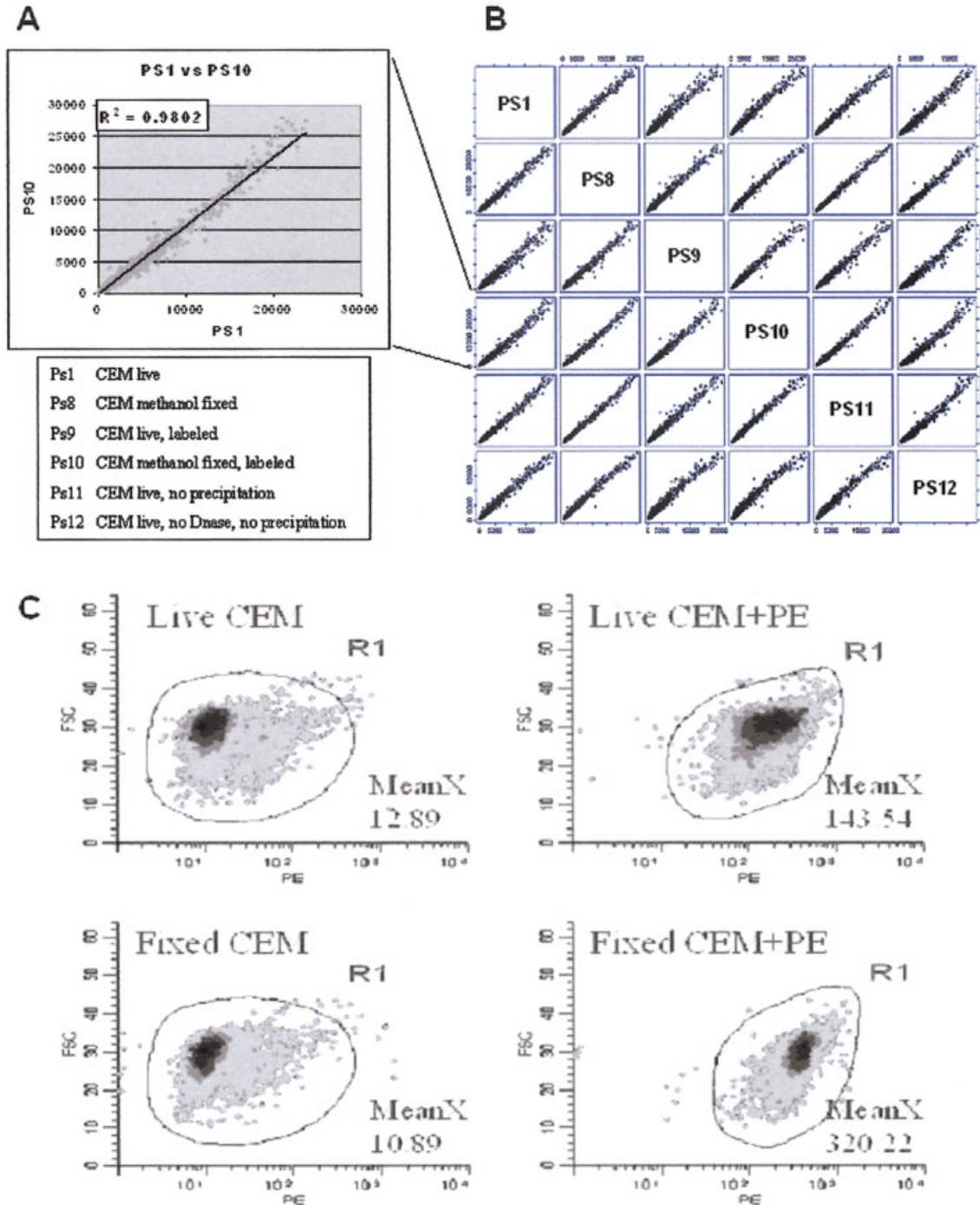
FIG. 3. Effects of sample processing on the gene expression profile. **A:** Scatterplot comparing the GEPs of live CEM cells and antibody labeled, methanol-fixed CEM cells. **B:** Trellis plot of all pairwise scatterplots, comparing samples PS1, and PS8–PS12. Each scatterplot is positioned in the row and column of the two samples from which it is made. **C:** Flow cytometry scattergrams comparing live (labeled and unlabeled) CEM cells to methanol fixed ones.

antibody-labeled, methanol-postfixed CEM cells. With an R^2-value of 0.982, we considered the overall GEP of the fixed and labeled sample unaltered.

Figure 3B is a Trellis plot of CEM cell samples that were all handled somewhat differently prior to microarray anal-

ysis. Two steps in the RNA isolation protocol, DNase digestion and purification of the isolated RNA by ethanol-precipitation, are generally thought to be important to generate good quality RNA. We omitted these steps in handling two samples, to test their importance. The ef-

fects of methanol fixation and antibody labeling—separately and combined—were also tested. None of the scatterplots presented here had an R^2-value lower than 0.975 (not shown), and were practically identical. Similarly to Figure 2D, a very small number of outlier genes showed significantly different expression levels within the otherwise almost identical samples.

While the GEP might not be altered significantly by methanol fixation, we were concerned that this fixation might alter the cell subset separation process. However, as shown by the flow cytometric results of Figure 3C, the fluorescent intensity of CEM cells labeled with anti-CD4, R-phycoerythrin-conjugated (PE) antibody after methanol fixation was found to be comparable to that of live, labeled cells. Similar results were obtained for fluorescein isothiocyanate (FITC)-conjugated antibodies. Autofluorescence of methanol-fixed cells was actually somewhat lower than that of live cells, allowing improved separation of labeled and unlabeled cells after fixation.

Recovering Gene Expression Profiles by Sorting

The next questions we addressed were: was it possible to recover the pure GEP of a cell subset after sorting these cells out of a mixture, and how close would the recovered profile be to the original, especially when the target cells were only a minor cell population in a mixture? Figure 4A displays the flow cytometry scattergrams of pure CEM cells, pure A2780 cells, their 1:10 mixture, and the CEM cells sorted from this mixture after anti-CD4-PE labeling. We tested the effects of both Miltenyi magnetic bead–based cell sorting and flow cytometry/cell sorting on the recovered GEP.

All four samples were analyzed by both spotted microarray technology (Fig. 4B) and short-oligonucleotide arrays (images not shown). On the spotted array images, we observed that the GEP of the mixture looked very similar to that of the A2780 cells, and the characteristics of the CEM cells seemed to be lost in it. The CEM cells separated from this mixture exhibited the lost characteristics, again very much resembling the original, pure CEM sample.

Figures 4C and D display GEP scatterplots comparing these four samples after both spotted array (Fig. 4C) and short-oligonucleotide array (Fig. 4D) analysis. The generated scatterplots and their R^2-values were found to be very similar between the two methods. As shown earlier, the GEPs of CEM and A2780 cells are indeed different, with R^2-values of 0.808 (Clontech array) and 0.815 (Affymetrix array). The CEM profile is lost in the mix, with R^2-values of 0.802 (Clontech array) and 0.841 (Affymetrix array). After separating these CEM cells from the mixture, the recovered GEP of the sorted CEM cells was found to be virtually identical to the original (pure CEM cells) profile, with R^2-values of 0.990 (Clontech array) and 0.984 (Affymetrix array). For the Clontech array analysis, CEM cells were purified by magnetic bead sorting (Miltenyi Biotec), and for the Affymetrix array analysis, a combination of magnetic bead sorting and flow cytometry/cell sorting was used.

Profiling CD34+ Stem/Progenitor Cells

After validating each step of cell purification for microarray analysis, we tested the method on a "real" biological sample, which was one of our original motivating applications for developing this technology. CD34+ stem/progenitor cells were isolated from cord blood using anti-CD34 antibody–coated Miltenyi magnetic beads. KG-1a cells are a 100% CD34+ cell line, usually used for modeling stem/progenitor cells. In this experiment, they served as a control for monitoring possible GEP distortions caused by cell purification. Figure 5A displays the flow cytometric scattergrams of the four samples after anti-CD4-PE labeling. Sample PS301 contained a mixture of all mononuclear cells isolated from cord blood by the Ficoll-Paque method. Sample PS302 contained CD34+ stem/progenitor cells purified from sample PS301. PS303 contained unpurified KG-1a cells, and PS304 was purified from PS303 using the same method that was used to sort PS302 (an internal control for the effects of cell handling). Flow cytometry data of samples PS303 and PS304 looked very similar, confirming that all KG-1a cells were CD34+ and were sorted. CD34+ stem/progenitor cells constitute less than 1% of all CBMCs (40,41). As expected, in our experiments, CD34+ cells from human cord blood samples showed up only as a very small subset on the PS301 flow cytometry scattergram. Flow cytometry data of PS302 confirmed their successful isolation.

All four samples were analyzed by Affymetrix microarrays. Figure 5B shows the array images of the samples after selecting the same portion of the full image, as in earlier experiments. The images of PS303 and PS304 were virtually identical, suggesting that the sort process did not introduce much distortion into the GEPs (at least not into the genes present in this image segment). Some genes expressed by purified CD34+ cord blood cells were similarly expressed in CBMCs and were not expressed in KG-1a cells (Fig. 5B, arrow 1). Expression levels of other genes were different in CD34+ purified cord blood cells than in any of the other three samples (Fig. 5B, arrow 2). These genes might be characteristic to stem/progenitor blood cells.

The scatterplots shown in Figure 5C also confirm that the sort process did not introduce any distortion into the overall GEPs, since the sorted and unsorted KG-1a cells were truly identical ($R^2 = 0.99$). With an R^2-value of 0.83, the overall GEP of purified CD34+ stem/progenitor cells was very different from the GEP of unsorted CBMCs. Surprisingly, the CD34+ stem/progenitor cells isolated from human cord blood were also very different from the KG-1a cells ($R^2 = 0.81$), despite the fact that these cells are generally used as a model stem/progenitor cell line in experiments. While KG-1a cells, originally derived from a bone marrow tumor, also express CD34 protein, the rest of their gene expression profile is dramatically different from those of a normal CD34+ stem/progenitor cell isolated from normal human cord blood.

Figure 5D is a software-generated (Spotfire, version 7.2) heat map comparing, by hierarchical clustering, the ex-
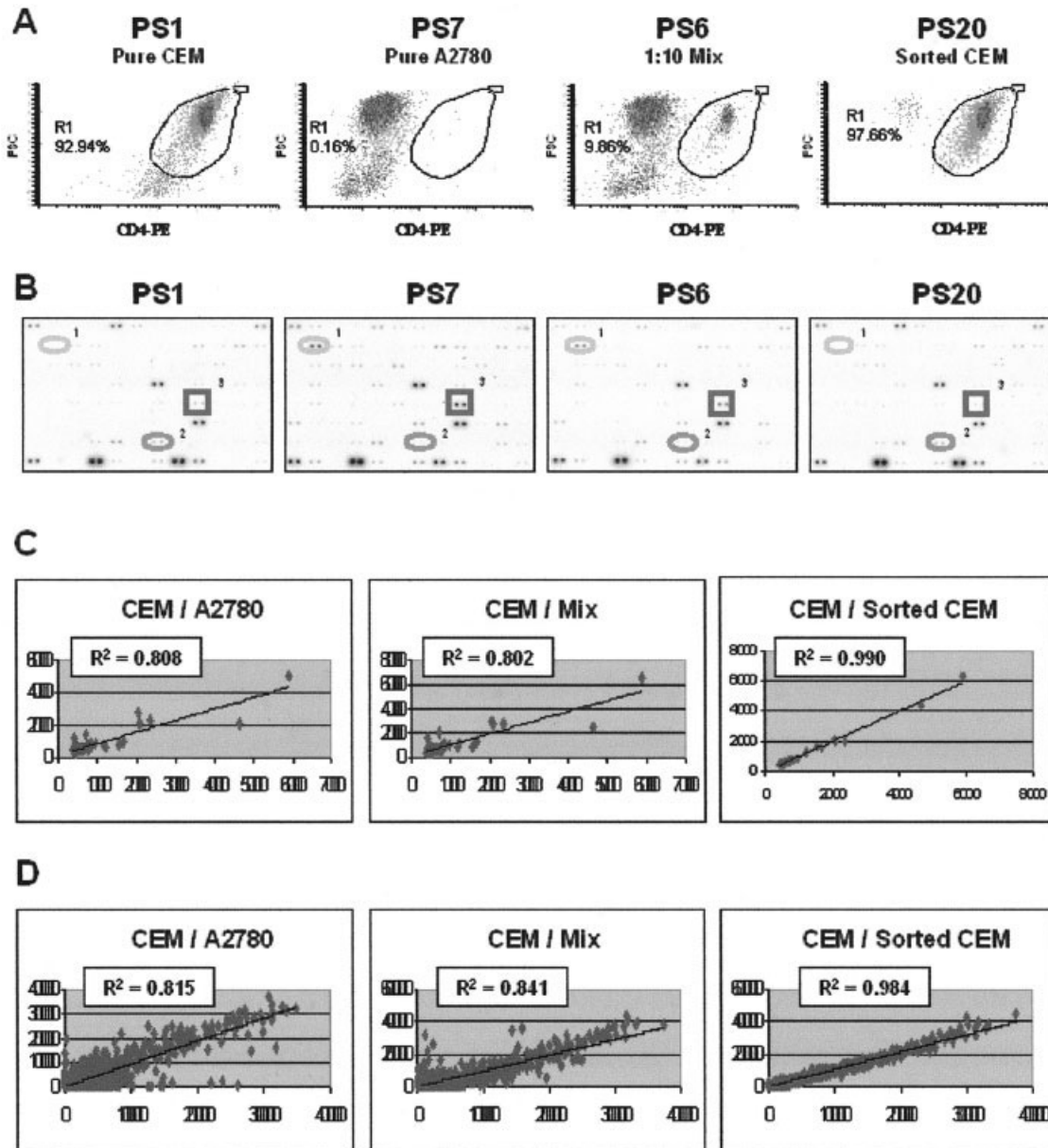
Fig. 4. Recovered gene expression profiles after cell sorting. **A:** Flow cytometry scattergrams of pure CEM and A2780 cells, a 10% CEM/A2780 cell mixture, and purified CEM cells recovered from the 10% mixture. **B:** Spotted microarray images of the same four samples. Genes expressed in only one of the cell types are in circles 1 and 2. Differentially expressed genes are in square 3. **C–D:** Scatterplots comparing CEM cells to A2780 cells, the 10% CEM/A2780 cell mixture, and to the recovered CEM cells based on spotted microarray results (C) and Affymetrix results (D).

pression levels of all of the genes judged to be "valid" (approximately 4,800 genes) for the four samples. Each gene is represented by a colored stripe, in which light green represents very low expression levels, and light red represents very high expression levels. The genes are arranged into clusters (both by gene expression levels in the vertical axis) and by cell sample (in the horizontal axis), by hierarchical clustering, to visualize characteristic groups of genes as patterns. Again, CD34+ stem/progenitor cells are truly different from both KG-1a cells and from CBMCs, although some groups of genes were ex-

pressed similarly. Hierarchical clustering analysis found that the GEP of CD34+ stem/progenitor cells isolated from cord blood was still slightly closer to the GEP of the KG-1a cells than to the mature blood cells in their original cord blood mixture. Prior purification of these CD34+ cells from cord blood was necessary to uncover their characteristic profile.

DISCUSSION

In this project, we set out to examine both the capabilities and limitations of the microarray approach in study-
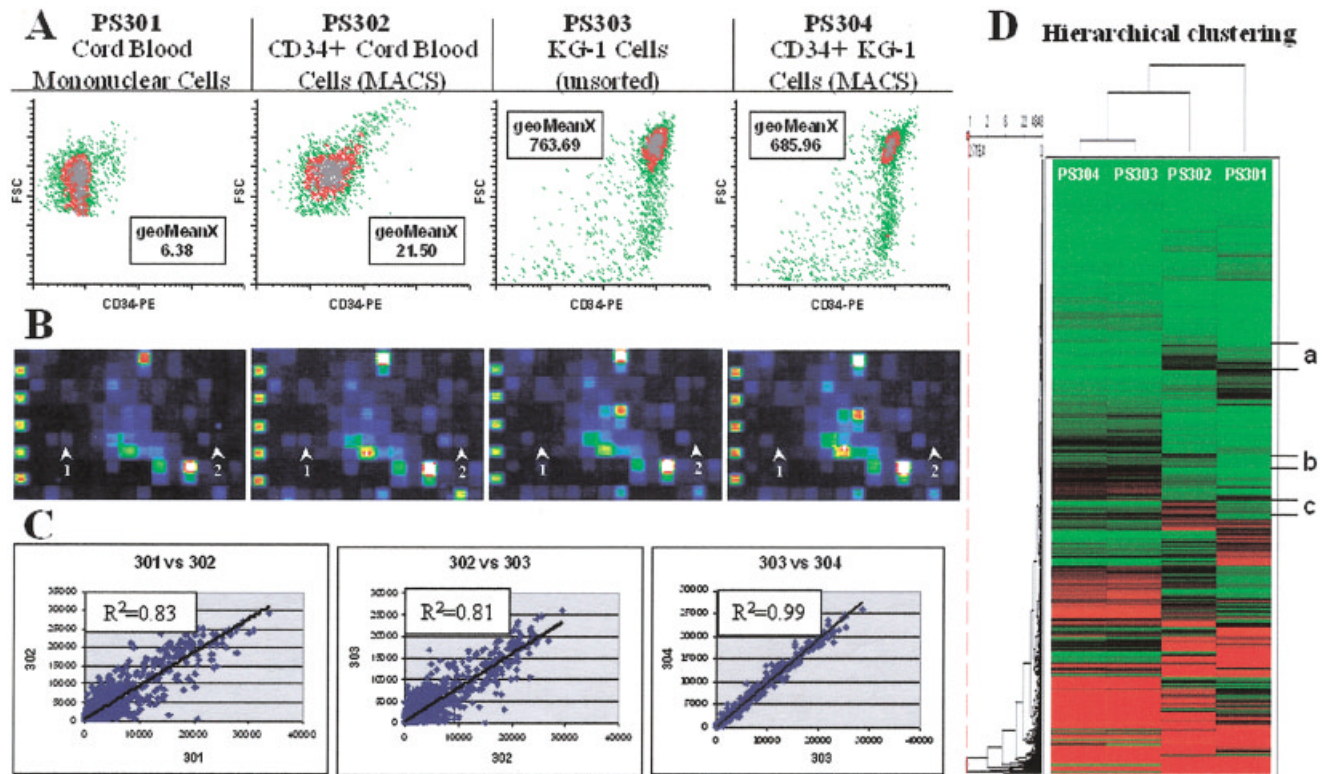
FIG. 5. Microarray analysis of purified CD34+ cord blood stem/progenitor cells. **A:** Flow cytometry scattergrams of CBMCs, MACS-purified CD34+ cord blood stem/progenitor cells, unsorted KG-1a cells, and MACS-sorted KG-1a cells. **B:** Affymetrix microarray images of the same four samples. Arrow 1 is pointing at a sequence expressed only in the first two samples. Arrow 2 is pointing at a sequence expressed only in CD34+ cord blood stem/progenitor cells. **D:** Heat map of the same four samples. Samples and genes are ordered by Spotfire hierarchical clustering analysis based on normalized expression levels. Some groups of genes (a, b, and c) are differentially expressed in CD34+ cord blood stem/progenitor cells.

ing gene expression in real biological samples and defined, model cell mixtures. To determine if meaningful data could be obtained by this method, even in the unfortunate scenario when the investigated cell type was a small minority in the sample, we tested the effects of cell-subset ratios and sample processing methods on the overall GEP, as well as on individual gene expression levels. We modeled biological samples by cell mixtures of two cell types, with different ratios of each type, and analyzed their GEP by both spotted (Clontech) and short-oligonucleotide microarrays (Affymetrix).

We found that without applying any cell separation, the cell type in the majority dominated the overall GEP, while the GEPs of minor cell subsets were washed out. Looking at the overall GEP when investigating a minor cell subset of a cell mixture is like "only seeing the tip of the iceberg." The differences between the GEPs of the gradually changing cell mixtures convincingly mirrored the changes in cell ratios. Summarizing our model cell mixture experiments, we concluded that the gene expression profiles for mixed cell populations are, as expected, the combined expression profiles for each cell subpopulation, weighted according to its relative frequency in the cell mixture.

When trying to determine the sensitivity of the MA approach in analyzing cell mixtures, we found that, in our model, the overall GEP of a more than 75% pure sample (PS2 and PS3) was indistinguishable from a 100% pure sample (PS1). The number of outlier genes within these samples as seen in the Trellis plots was very small (fewer than 10 out of more than 12,000 genes), considering that the raw data was not preprocessed prior to scatterplot analysis. However, these outlier genes might indicate that purity requirements can be very different for monitoring individual genes, depending on whether those same genes are expressed at high or low levels in the contaminating cell types.

Our results indicated that, in the case of minor cell subsets, to be able to see more than just the tip of the iceberg, cell purification is necessary. Any purification method takes time, and mRNA is a "moving target," with possible degradation during the experiment. RNAs can be produced very rapidly and some mRNAs may be degraded in a matter of minutes in live cells, while others may be stable over several hours (32,42). Is it possible to "freeze the GEP in time" by cell fixation until the cells get purified and delivered to the microarray? Most purification methods require labeling of the target cells. How much will the labeling process alter the GEP of the labeled cells? To address these questions, along with the reasonable concern that the more a sample is processed the more dis-

torted its GEP might get, we also tested handling effects on overall sample GEP. We showed that after antibody labeling and methanol fixation the overall GEP remained unaltered, and even omitting steps traditionally used to improve RNA quality did not have a significant effect on the overall GEP. Again, the presence of a few outlier genes indicated that individual genes might be much more affected by certain processing steps; obviously, antibody labeling of a surface receptor on a live cell might trigger certain pathways altering the expression levels of the genes involved. Nevertheless, we were able to conclude that the overall GEP of a sample (representing the vast majority of all genes) is more robust and resistant to sample processing than has been generally appreciated.

Methanol fixation of the antibody-labeled cells prior to cell purification and MA analysis turned out to be a rather fortunate choice. It did not unfavorably alter the detection/selection process by either immunomagnetic cell separation or flow cytometric cell sorting. Both PE- and FITC-labeled antibodies used in these experiments maintained good separation characteristics after methanol postfixation. In summary, this fixation method not only preserved the GEP of labeled cells, but also allowed fluorescence-based labeling for cell sorting.

To address the question of how much purity we need in a sample, we showed that generally it is not necessary to achieve 100% purity. In our model for the overall GEP, anything above 75% pure was found to be indistinguishable from the pure sample. This level of purity can be achieved by two rounds of magnetic bead sorting. One round typically results in about 70% purity; two rounds raises the purity to approximately 90%; while after three rounds, it is generally above 95%. One round of magnetic bead sorting followed by one round of flow cytometry/cell sorting results in about 90–98% purity as well. Studying individual genes, however, might require much higher or lower sample purity, depending on the gene's relative expression levels in the cell subsets. With good cell biomarkers and techniques, multiparameter flow cytometry/cell sorting can be used to obtain purities of more than 99%. This degree of purity may be needed for correct GEP analysis of low-expressing genes in which even 90% purity may be insufficient to obtain accurate GEP results for those specific genes.

To test just how much of the "iceberg" can be revealed, we purified the minority cell subset of a 10% cell mix in which the GEP of the minor cell subset was shown to be covered by the background cells. Using both magnetic bead cell purification and flow cytometry/cell sorting, we managed to recover the "hidden" GEP virtually perfectly, also proving that the sort process itself did not distort the profile. The almost perfectly recovered profiles suggest that both magnetic bead and conventional flow cytometry/cell sorting purification methods are capable of purifying cells without significantly distorting their GEP. Results from the control KG-1a cells from the cord blood experiment confirmed this finding, since the cells that went through the sort process matched the unsorted cells, with an R^2-value of 0.99. We concluded that for meaning-ful gene expression microarray profiling a minor cell subset of a cell mixture, purification of these cells is not only necessary, but also very much achievable, recovering the "pure profile" without any significant distortion, despite the concerns expressed previously in the literature (18,33,39). In our hands, following the procedures described in this work, the effects of sample handling on the GEP were minimal and not significant.

As a proof-of-principle experiment, we measured the GEP of purified, CD34+ cord blood stem/progenitor cells. Since these cells are present in cord blood at a less than 1% minority of all mononuclear cells (40,41), their GEP had been heavily masked by the overwhelming presence of mature, contaminating, cell types. The true stem/progenitor cell-GEP was "invisible" without purification. We showed that the recovered GEP of these cells was characteristically different from both CBMCs and KG-1a cells. This result seriously questions the use of KG-1a cells as a model cell line for stem/progenitor cells in gene expression studies, even though that cell line was originally established from a bone marrow tumor.

For the cord blood experiment, we needed to pool several samples. Individual cord blood samples could not be directly analyzed, simply because they did not provide enough purified CD34+ cells necessary for one microarray analysis. This problem would be even more serious if we wanted to further purify this cell subset based on the cells' other surface antigen properties. Unfortunately, many biological samples do not provide enough purified cells of a certain cell type for direct gene expression profiling. For these samples, nondistorting RNA-amplification is necessary prior to microarray analysis.

In summary, we found the results presented here very promising. Both Clontech and Affymetrix arrays performed at a very high level of reproducibility, the generated profiles proved to be surprisingly robust, and hidden GEPs could be accurately recovered from cell mixtures by cell separation techniques. MAT, based on specific cell subpopulations, could truly become a driving technology not only in genomics, but also in the emerging field of cytomics, which aims at the understanding of the molecular architecture and functionality of cell systems (cytomes) by single-cell analysis in combination with exhaustive bioinformatic knowledge extraction (37).

LITERATURE CITED

1. Chaussabel D, Sher A. Mining microarray expression data by literature profiling. Genome Biol 2002;3:R55.
2. Shoemaker DD, Linsley PS. Recent developments in DNA microarrays. Curr Opin Microbiol 2002;5:334–337.
3. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995;270:467–470.
4. Lander ES. Array of hope. Nat Genet 1999;21(1 suppl):3–4.
5. Brown PO, Botstein D. Exploring the new world of the genome with DNA microarrays. Nat Genet 1999;21(1 suppl):33–37.
6. Bowtell DD. Options available—from start to finish—for obtaining expression data by microarray. Nat Genet 1999;21(1 suppl):25–32 [erratum, Nat Genet 1999 Feb;21:241].
7. Holloway AJ, van Laar RK, Tothill RW, Bowtell DD. Options available—from start to finish—for obtaining data from DNA microarrays II. Nat Genet 2002;32(suppl):481–489.
8. Leung YF, Cavalieri D. Fundamentals of cDNA microarray data analysis. Trends Genet 2003;19:649–659.
9. Copland JA, Davies PJ, Shipley GL, Wood CG, Luxon BA, Urban RJ. The use of DNA microarrays to assess clinical samples: the transition from bedside to bench to bedside. Recent Prog Horm Res 2003;58:25–53.
10. Petricoin EF, 3rd, Hackett JL, Lesko LJ, Puri RK, Gutman SI, Chumakov K, Woodcock J, Feigal DW, Jr, Zoon KC, Sistare FD. Medical applications of microarray technologies: a regulatory science perspective. Nat Genet 2002;32(suppl):474–479.
11. Shaughnessy J, Jr. Primer on medical genomics. Part IX: scientific and clinical applications of DNA microarrays—multiple myeloma as a disease model. Mayo Clin Proc 2003;78:1098–1109.
12. Nakeff A, Sahay N, Pisano M, Subramanian B. Painting with a molecular brush: genomic/proteomic interfacing to define the drug action profile of novel solid-tumor selective anticancer agents. Cytometry 2002;47:72–79.
13. Los G, Yang F, Samimi G, Manorek G, Guerorguieva IM, Howell S, van Erp N, Breaux JK. Using mRNA expression profiling to determine anticancer drug efficacy. Cytometry 2002;47:66–71.
14. Staunton JE, Slonim DK, Coller HA, Tamayo P, Angelo MJ, Park J, Scherf U, Lee JK, Reinhold WO, Weinstein JN, et al. Chemosensitivity prediction by transcriptional profiling. Proc Natl Acad Sci USA 2001;98:10787–10792.
15. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, et al. Molecular portraits of human breast tumours. Nature 2000;406:747–752.
16. Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, et al. Molecular classification of cutaneous malignant melanoma by gene expression profiling. Nature 2000;406:536–540.
17. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530–536.
18. Spruill SE, Lu J, Hardy S, Weir B. Assessing sources of variability in microarray gene expression data. Biotechniques 2002;33:916–920, 922–923.
19. Mirnics K. Microarrays in brain research: the good, the bad and the ugly. Nat Rev Neurosci 2001;2:444–447.
20. Packeisen J, Korsching E, Herbst H, Boecker W, Buerger H. Demystified. . .tissue microarray technology. Mol Path 2003;56:198–204.
21. Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JC, Trent JM, Staudt LM, Hudson J, Jr, Boguski MS, et al. The transcriptional program in the response of human fibroblasts to serum. Science 1999;283:83–87.
22. Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. [see Comment]. Lancet 2003;362:1439–1444.
23. Russo G, Zegar C, Giordano A. Advantages and limitations of microarray technology in human cancer. Oncogene 2003;22:6497–6507.
24. Boess F, Kamber M, Romer S, Gasser R, Muller D, Albertini S, Suter L. Gene expression in two hepatic cell lines, cultured primary hepatocytes, and liver slices compared to the in vivo liver gene expression in rats: possible implications for toxicogenomics use of in vitro systems. Toxicol Sci 2003;73:386–402.
25. Chen Y, Knosel T, Kristiansen G, Pietas A, Garber ME, Matsuhashi S, Ozaki I, Petersen I. Loss of PDCD4 expression in human lung cancer correlates with tumour progression and prognosis. J Path 2003;200:640–646.
26. Kaminski N, Allard JD, Pittet JF, Zuo F, Griffiths MJ, Morris D, Huang X, Sheppard D, Heller RA. Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. Proc Natl Acad Sci USA 2000;97:1778–1783.
27. Lock C, Hermans G, Pedotti R, Brendolan A, Schadt E, Garren H, Langer-Gould A, Strober S, Cannella B, Allard J, et al. Gene-microarray analysis of multiple sclerosis lesions yields new targets validated in autoimmune encephalomyelitis. Nat Med 2002;8:500–508.
28. Whitney AR, Diehn M, Popper SJ, Alizadeh AA, Boldrick JC, Relman DA, Brown PO. Individuality and variation in gene expression patterns in human blood. Proc Natl Acad Sci USA 2003;100:1896–1901.
29. Lee S, Zhou G, Clark T, Chen J, Rowley JD, Wang SM. The pattern of gene expression in human CD15+ myeloid progenitor cells. Proc Natl Acad Sci USA 2001;98:3340–3345.
30. Ivanova NB, Dimos JT, Schaniel C, Hackney JA, Moore KA, Lemischka IR. A stem cell molecular signature. Science 2002;298:601–604.
31. Ramalho-Santos M, Yoon S, Matsuzaki Y, Mulligan RC, Melton DA. "Stemness": transcriptional profiling of embryonic and adult stem cells. Science 2002;298:597–600.
32. Auer H, Lyianarachchi S, Newsom D, Klisovic MI, Marcucci, Kornacker K. Chipping away at the chip bias: RNA degradation in microarray analysis. Nat Genet 2003;35:292–293.
33. Schuchhardt J, Beule D, Malik A, Wolski E, Eickhoff H, Lehrach H, Herzel H. Normalization strategies for cDNA microarrays. Nucleic Acids Res 2000;28:E47.
34. Knight J. When the chips are down. [see Comment]. Nature 2001;410:860–861.
35. Rogojina AT, Orr WE, Song BK, Geisert EE, Jr. Comparing the use of Affymetrix to spotted oligonucleotide microarrays using two retinal pigment epithelium cell lines. Mol Vis 2003;9:482–496.
36. Sasik R, Calvo E, Corbeil J. Statistical analysis of high-density oligonucleotide arrays: a multiplicative noise model. Bioinformatics 2002;18:1633–1640.
37. Valet G, Leary JF, Tárnok A. Cytomics - new technologies: towards a human cytome project. Cytometry 2004;59A:167–171.
38. Leary JF, Reece LN, Szaniszlo P, Prow T, Wang N. 2001. High-throughput cell analysis and sorting technologies for clinical diagnostics and therapeutics. In: Farkas DL, Leif RC, editors. Optical diagnostics of living cells IV. Proceedings of SPIE, Vol. 4255. p 16–27.
39. Goldsworthy SM, Stockton PS, Trempus CS, Foley JF, Maronpot RR. Effects of fixation on RNA extraction and amplification from laser capture microdissected tissue. Mol Carcinog 1999;25:86–91.
40. Hao QL, Shah AJ, Thiemann FT, Smogorzewska EM, Crooks GM. A functional comparison of CD34 + CD38- cells in cord blood and bone marrow. Blood 1995;86:3745–3753.
41. Thoma SJ, Lamping CP, Ziegler BL. Phenotype analysis of hematopoietic CD34+ cell populations derived from human umbilical cord blood using flow cytometry and cDNA-polymerase chain reaction. Blood 1994;83:2103–2114.
42. Lam LT, Pickeral OK, Peng AC, Rosenwald A, Hurt EM, Giltnane JM, Averett LM, Zhao H, Davis RE, Sathyamoorthy M, et al. Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. Genome Biol 2001;2:R41.