Precision Profiling and Components of Variability Analysis for Affymetrix Microarray Assays Run in a Clinical Context

Thomas M. Daly, Carmen M. Dumaul, Crystal A. Dotson, Mark W. Farmen, Sunil K. Kadam, and Richard D. Hockett

From the Division of Experimental Medicine, Genomic Medicine Group, Eli Lilly and Company, Indianapolis, Indiana

Although gene expression profiling using microarray technology is widely used in research environments, adoption of microarray testing in clinical laboratories is currently limited. In an attempt to determine how such assays would perform in a clinical laboratory, we evaluated the analytical variability of Affymetrix microarray probesets using two generations of human Affymetrix chips (U95Av2 and U133A). The study was designed to mimic potential clinical applications by using multiple operators, machines, and reagent lots, and by performing analyses throughout a period of several months. A mixed model analysis was used to evaluate the relative contributions of multiple factors to overall variability, including operator, instrument, run, cRNA/cDNA synthesis, and changes in reagent lots. Under these conditions, the average probeset coefficient of variation (CV) was relatively low for present probesets on both generations of chips (mean coefficient of variation, 21.9% and 27.2% for U95Av2 and U133A chips, respectively). The largest contribution to overall variation was chip-to-chip (residual) variability, which was responsible for between 40 to 60% of the total variability observed. Changes in individual reagent lots and instrumentation contributed very little to the overall variability. We conclude that the approach demonstrated here could be applied to clinical validation of Affymetrix-based assays and that the analytical precision of this technique is sufficient to answer many biological questions. (*J Mol Diagn* 2005, 7:404–412)

Techniques for multiplex gene expression analysis have become increasingly widespread, and provide an opportunity to simultaneously interrogate thousands of genes in a sample to provide a global picture of transcriptional activity. This can be used in experimental systems to gain information about changes in cellular processes in response to experimental manipulations, and to identify potential target genes for future study. In the clinical

realm, adoption of this technology is probably most advanced in the field of oncology, in which molecular profiling of tumors can potentially provide better characterization of tumor types, leading to better prediction of prognosis and more accurate selection of therapy. An example of this is profiling of diffuse large B-cell lymphoma, in which microarray data can be used to differentiate morphologically similar tumors into genetically different groups,^{1,2} which show functional differences in response to standard chemotherapy.³ Such differences in therapeutic response demonstrate why microarray studies can also be useful in the drug development process, in which identification of molecular responsiveness to a compound can help elucidate underlying mechanisms of action for potential therapeutic candidates, allowing subpopulations of patients to be identified where the drug would be most efficacious. However, for microarray data to be clinically applicable in a regulated environment, an understanding of the inherent complexity and variability of the analytical technique must be understood.

The Affymetrix GeneChip platform is the most widely used commercially available microarray for expression analysis. In this technology, pairs of 25-nucleotide oligos are synthesized *in situ* on silica wafers. Each probe pair contains an oligo that exactly matches the target sequence (perfect match, PM) and a second oligo that differs by a single nucleotide in the center of the oligo (mismatch, MM).^{4,5} The presence or absence of a given target sequence in the sample can then be calculated using one of several algorithms based on comparing the PM and MM signals across a probeset of 8 to 16 probe pairs for a given sequence. In addition, the relative expression level of a target sequence can be estimated by the intensity of the signal.

Sample preparation for Affymetrix genechip analysis is a multistep process that includes isolation and cleanup of RNA from target tissue, generation of cDNA by reverse transcription, synthesis of biotinylated cRNA from this cDNA template, hybridization to the Affymetrix chip, and staining using streptavidin-phycoerythrin. Each of these steps has the potential to introduce analytical variability into the final results. For many research applications us-

Accepted for publication April 1, 2005.

Address reprint requests to Thomas M. Daly, Eli Lilly and Company, Lilly Corporate Center, Indianapolis, IN 46240. E-mail: dalytm@lilly.com

ing microarrays, this variability is reduced by batch analyzing samples from a given experiment using a single analyst in a limited number of runs with single lots of reagents. Alternatively, large populations of patients can be analyzed and data analyzed on a group basis, which minimizes the impact of variability in individual analyses by evaluating changes across a population. These approaches contrast sharply with potential clinical applications of microarray technology, in which patient samples will likely be analyzed in real time as they are collected and will need to be analyzed on an individual patient basis, not as part of a combined population. In addition, changes throughout time in reagent lots, analysts, and machine performance could also introduce analytical variation in results obtained on different occasions or at different testing locations. For these reasons, understanding the analytical variability of Affymetrix microarrays in such a context will be critical.

For any clinical assay, an understanding of the normal variation of the marker is needed to allow interpretation of patient results. The overall variability of the assay is influenced by preanalytical factors, by the analytical precision of the assay, and by the baseline biological variability of the marker being examined. Biological variability is an inherent characteristic of the analyte being measured, and protocols to evaluate within- and between-patient variability of clinical assays have been described.⁶ However, there is usually little that can be done to influence the amount of biological variability. In contrast, preanalytical variability can often be reduced by identifying sources of variation and limiting their impact. For Affymetrix analyses, the quality of input RNA and the consistency of sample collection and processing methods have been identified as crucial factors. Dumur and colleagues⁷ recently described QC criteria for input sample RNA derived from a variety of sources, and similar criteria have been recommended in a recent best practices document from the Tumor Analysis Best Practices Working Group.⁸ Once preanalytical variation has been reduced as much as possible, the analytical precision of the assay becomes the single biggest factor in determining what level of biological change can be seen.

Because of the complexity of the Affymetrix analytical process, there are many factors that can contribute to analytical variation. Several groups have looked at the overall precision of Affymetrix analyses, and have found average probeset coefficient of variations (CVs) ranging from 8 to 13% on a variety of human and rodent microarrays.^{9–11} However, the majority of these studies were done using a relatively small number of replicates (<15 in most cases) and were analyzed in a limited number of runs. In our study we have tried to more closely mimic potential clinical applications by evaluating precision using a much larger number of technical replicates throughout an extended period of time, using multiple operators, instruments, and reagent lots. In addition, we have specifically looked at the contributions of each of these components to overall variability to identify steps for which rigorous control criteria would provide the most benefit. Using this approach, we have generated a more realistic estimate of the analytical precision of this technology in

clinical use, and provide a model protocol that could be used to validate clinical applications of Affymetrix microarrays.

Materials and Methods

Preparation of RNA Samples

A large pool of cell line RNA was isolated from an antifolate-resistant CCRF-CEM leukemia cell line¹² for use in the precision experiments. Cells (3×10^9) were collected by centrifugation at 3000 rpm and rinsed one time in $1 \times$ phosphate-buffered saline (Invitrogen Life Technologies, Carlsbad, CA). The cell pellet was resuspended in 500 ml of TRIzol reagent (Invitrogen), mixed thoroughly, and then split into 1-ml aliquots that were stored in Trizol at -80°C until use.

Tumor RNA was isolated from a uterine leiomyoma specimen obtained from the Department of Pathology, Indiana University School of Medicine, in accordance with the guidelines of Indiana University and with the approval of the Indiana University-Purdue University Indianapolis institutional review board. The tissue was surgically removed, snap-frozen in liquid nitrogen, and stored at -80°C until ready for use. Two- to three-mm cube sections of tissue were cut off and mechanically homogenized in the presence of TRIzol reagent. Homogenates were pooled together, mixed well, separated into 1-ml aliquots, and stored at -80°C .

General Microarray Analytical Procedure

Total RNA was isolated from the TRIzol aliquots following the manufacturer's instructions, and purified using an RNeasy kit (Qiagen Inc., Valencia, CA). RNA integrity and yield were assessed by determining sample absorbance at 260 and 280 nm and by analysis on the Agilent RNA 6000 Nano LabChip (Agilent Technologies, Inc., Palo Alto, CA). All samples had 260:280 ratios >1.8 and clear 18S and 28S ribosomal RNA bands on the Agilent. Complementary RNA synthesis and gene expression profiling were performed following the protocol described in the Affymetrix GeneChip Expression Analysis Technical Manual,¹³ with only minor changes. Briefly, 5 μg of cleaned total RNA were used to generate double-stranded cDNA by reverse transcription, using a Superscript, double-stranded cDNA synthesis kit (Invitrogen) and an oligo deoxythymidylic acid primer with a T7 RNA polymerase promoter site added to the 3' end. After second-strand synthesis, cDNA was cleaned with a GeneChip Sample Cleanup Module (Affymetrix). Biotin-labeled cRNA was produced by *in vitro* transcription, using the Enzo BioArray high-yield RNA transcript labeling kit (Enzo Diagnostics, Farmingdale, NY). Labeled cRNA was cleaned with a GeneChip Sample Cleanup Module, dried down in a Savant Speed Vac concentrator (Savant Instruments, Inc., Holbrook, NY) and resuspended to a concentration of 1 $\mu\text{g}/\text{ml}$. Twelve μg of the concentrated cRNA product was fragmented by metal-induced hydrolysis at 94°C for 35 minutes. The efficiency

Figure 1. Experimental designs for studies performed in this article. **A:** U95Av2 variability assessment study. Diagram represents a single run of eight chips performed by two analysts. *Fluidic stations were varied between analysts on alternate runs. **B:** U133A reagent lot study. Diagram represents a single run of four chips performed by one analyst. *Hybridization mix lots were reversed on alternate runs to incorporate all possible reagent lot combinations. **C:** U133A tumor matrix study. Diagram represents a single run of four chips performed by one analyst. Reagent lots were held constant for all runs in this study.

of the fragmentation procedure was checked by analyzing the size of the fragmented cRNA on the Agilent 2100 bioanalyzer. Each fragmented sample was then used to prepare 200 μ l of hybridization cocktail containing 100 mmol/L MES, 1 mol/L NaCl, 20 mmol/L ethylenediamine tetraacetic acid, 0.01% Tween 20, 0.1 mg/ml herring sperm DNA (Promega Corp., Madison, WI), 0.5 mg/ml acetylated bovine serum albumin (Invitrogen), 50 pmol/L control oligonucleotide B2, 100 pmol/L eukaryotic hybridization controls (Affymetrix), and 6 μ g of fragmented sample. Samples were then hybridized for 16 hours to human genome U95Av2 or U133A arrays (Affymetrix).

After hybridization, GeneChips were washed and stained with streptavidin-phycoerythrin (Molecular Probes, Inc., Eugene, OR), according to the appropriate standard protocol for each chip type. Arrays were scanned using the Affymetrix GeneChip Scanner 3000 and image analysis was performed using Affymetrix Microarray Analysis Suite (MAS), version 5.1¹⁴. Each sample was scaled to a target intensity of 500 for all probe sets. Signal values, detection calls (present, absent, marginal), and *P* values for each detection call were generated using MAS 5.1 Absolute Expression Analysis.¹⁵

Experimental Design for Analytical Precision Studies

A total of 64 HG-U95Av2 chips were used to assess probeset precisions and the contributions of components of variation to the assay, including day/run, analyst, cDNA synthesis reaction, cRNA synthesis reaction, fluidic station, and chip lot. Eight runs of eight chips were performed using aliquots of CEM-MTA cells in 1 ml of TRIzol. Each run consisted of two analysts, each extracting total RNA from one aliquot of cells (Figure 1A). Both analysts worked side-by-side throughout the procedure in this initial experiment, to ensure that all samples were

frozen for the same lengths of time between each step. Each total RNA sample was used as a template for two cDNA synthesis reactions, using the Prologo T7-(dT)₂₄ primer. Each cDNA sample was then *in vitro*-transcribed, fragmented, and hybridized to two separate arrays, yielding a total of four chips per analyst per run. Within each run, the four chips of each analyst were alternated between two fluidics stations. Five runs (40 chips) were completed using one lot of HG-U95Av2 arrays and three runs (24 chips) were performed using a different lot of arrays to assess chip lot-to-lot variation. With the exception of chip lot, the lot number was fixed for all other reagents within the experiment. Chips were always scanned in the same order, 1 to 8, for every run.

To address the effects of reagent lots on overall variability, a total of 48 HG-U133A GeneChips were analyzed in 12 runs of four chips (Figure 1B). To assess the amount of variability introduced to the procedure when operators are not working side-by-side, two analysts each performed 6 of the 12 runs in a randomized order, with all 12 runs beginning on different days, incorporating two lots each of cDNA synthesis kits, cRNA *in vitro* transcription kits, and hybridization mix. Four lots of GeneChip arrays were used. Each run was performed by a single analyst, and consisted of four separate extractions of total RNA from stored aliquots of CEM-MTA cells in TRIzol. Each total RNA sample was then used for a single cDNA synthesis reaction, using the Affymetrix T7-Oligo (dT) promoter primer, which was subsequently transcribed, labeled, fragmented, and hybridized to one U133A array. Two reagent lots of cDNA synthesis, *in vitro* transcription (IVT), and hybridization reagents were varied for each reaction so that each of the eight possible combinations was tested three separate times by each analyst. In addition, a second lot of cRNA cleanup kits was used for the final four runs. All samples in a given run were hybridized to a single lot of chips.

Tumor tissue matrix studies were performed using a total of 16 U133A GeneChips analyzed by two analysts in four runs of four chips each (Figure 1C). For each run, analysts each extracted total RNA from two aliquots of homogenized tissue, stored in TRIzol. Each total RNA sample was used as a template for two cDNA synthesis reactions that were then transcribed, fragmented, and hybridized to a single U133A array per sample. A single reagent lot was used for all reagents within this experiment.

Statistical Analysis of Data

A variance components analysis was done by gene and then summarized across genes using SAS (SAS Institute Inc., Cary, NC). Genes were categorized by the percentage of the chips on which they were called present or marginal or by median Affymetrix detection *P* value. The CV for each probeset was calculated as $100 \times \sqrt{\text{variance}} / \text{mean}$, where the mean is the overall average of the chip values within each experiment. Variance was calculated using a mixed model approach¹⁶ with run and cDNA/cRNA production treated as nested random effects and analyst, machine, and chip lot treated as fixed effects in the initial study. For the reagent lot experiment, chip lot and run were treated as nested random effects and analyst, cDNA kit, IVT kit, hyb mix, and RNA clean-up kit were treated as fixed. The residual variance is composed of pure chip-to-chip variability, which represents the underlying variation in the assay contributed by factors other than those specifically tested above. The fixed effect means were converted to variance components under strict assumptions,¹⁷ which assumes the same two analysts, machines, and lots are to be continuously used with equal probability for any sample. The square root of the sum of the variance components (total variance) is used to estimate the SD of a gene measurement on one chip.

Results

Variability Components of Analytical Process

The initial set of experiments was designed to assess the overall analytical variability of microarray results and to determine which components of the analytical process contributed the most to the overall variability of the system. RNA isolated from a leukemic cell line was analyzed by two analysts using 64 Affymetrix U95Av2 chips in eight runs throughout a period of 3 months (Figure 1A). Eight chips were lost due to fluidic station failure in a single run, and one chip was excluded due to an inability of the software to align the grid and generate (a .chp) file. All data from the remaining 55 chips were included in the final analysis, regardless of the background and signal values or any other quality control values obtained.

The majority of genes were consistently classified in the same category (P/M versus A) throughout the experiment. Of the 12,625 probesets 7087 had identical calls on all 55 chips analyzed, while 83.5% of probesets were

concordant on 50 or more chips (Figure 2A). CVs for individual probesets ranged from 6 to 353% and was related to signal intensity, with the highest CVs occurring in probesets with mean signal intensities of lower than 100 (Figure 2B). However, the majority of the highly variable probesets were for absent calls. Present probesets (defined as probesets with median *P* values of <0.06 in the 55 chips) had much more reproducible signals, with an average CV of 21.9% (SD 9.6%) and a 95th percentile of 40% CV. It is also interesting to note this relatively low CV continues to hold for probesets with *P* values >0.06 , and only starts to rise at a faster rate for *P* values >0.1 (Figure 2D). This suggests that from an analytical standpoint, a cutoff of 0.06 is relatively conservative.

The contribution of various factors to the overall variance was calculated using a mixed model containing both controllable and uncontrollable factors. The six sources of variance analyzed were analyst, fluidic station, chip lot, cDNA synthesis/IVT reaction, run-to-run, and chip-to-chip (residual). Overall, the largest component to variability was found to be chip-to-chip variation, which accounted for ~40% of the total variability (Figure 3A). Run-to-run, cDNA/IVT synthesis, and chip lot variability made up approximately equal parts of the remainder. Analyst and fluidic station variation contributed very little ($<5\%$) to the overall variability under these experimental conditions.

Effects of Reagent Lots on Variability

To accurately compare results generated at different times, lot-to-lot variability should be minimal for the reagents used in an analytical system. We next evaluated the effects of differing lots of reagents on analytical variability. Before the start of these experiments, Affymetrix released the U133A chip format. An initial experiment replicating a single run of the study above using U133A chips showed substantially equivalent performance to the U95Av2 in terms of overall analytical variability and components of variation (data not shown). Based on this, U133A chips were used for all remaining experiments.

Twelve runs of four chips were performed using the same cell line RNA preparation as before. Analyses were run by two different analysts on 12 different occasions throughout a 7-month period. Four different chip lots were included in the study, and two different lots of cDNA synthesis kits, IVT kits, cRNA cleanup kits, and hybridization cocktail were used (Figure 1B). All other assay components were fixed to a single lot. No run failures occurred, and data from all 48 chips were included in the final analysis.

Concordance of P+M calls was again good between chips, with 17,130 out of 22,283 probesets (76.9%) giving identical calls on $>90\%$ of the chips analyzed. As before, the highest CVs occurred at signal intensities <100 , mainly in absent calls (Figure 4A). Mean CV for present probesets was 27.2% (SD, 9.7%), with 90.8% of present probesets having a CV of $<40\%$. As a population, the variability of U133A probesets under the tested conditions was slightly higher than that seen for the U95Av2

Figure 2. **A:** Concordance of P+M calls across the 55 chips analyzed. The majority of probesets were consistently called either P+M or A on all chips. **B:** Coefficient of variation of signal intensity was calculated for each probeset, and plotted against the mean intensity for all probesets. **C:** Expanded view of CV versus intensity for present probesets only. **D:** Probeset CV as a function of median *P* value. **Dotted line** indicates *P* = 0.06.

Figure 3. Components of variability analysis. Probesets are divided into present (median *P* value < 0.06) or absent (median *P* value > 0.06) for each study. Mean with SD error bars are plotted. **A:** U95Av2 variability assessment study. **B:** U133A reagent lot study.

Figure 4. Probeset variability of U133A Affymetrix chips. **A:** Coefficient of variation versus signal intensity for both present and absent probesets. **B:** Comparison of CVs between U95Av2 and U133A chips for probesets called present.

chips (Figure 4B). Once again, chip-to-chip variation was the largest contributor to overall variability, responsible for ~55% of the overall variability seen in the present probesets. The only reagents that substantially contributed to variation were different RNA cleanup kits and chip lots (Figure 3B). Changes in reagent lots of the cDNA synthesis kits, IVT kits, and hybridization mixes contributed minimal amounts to the overall variation. Interestingly, run-to-run and analyst variability were higher in this study than in the U95Av2 study, with each contributing ~10% of the overall variability. This may be due to the longer time period throughout which these studies were run (7 months as opposed to 3 months) as well as the fact that the analysts were no longer working side-by-side, and may more accurately represent the effects of different analysts in a clinical laboratory context.

Matrix Effects on Analytical Variability

The previous experiments were performed using RNA prepared from a cultured cell line. To determine whether these results were applicable to RNA isolated from a more clinically relevant source, we next looked at RNA obtained from flash-frozen tumor resection tissue. RNA isolated from a large uterine leiomyoma was obtained and analyzed on 16 U133A chips in four runs by two analysts (Figure 1C). Probeset variability for present probes in this matrix was similar to that seen using cell line RNA, with a mean CV of 29.1% and 86.9% of the present probesets having a CV of <40%. Chip-to-chip variability contributed 64% of the overall variation, while run-to-run, analyst, and cRNA/IVT reactions all contributed an approximately equal part of the remainder (data not shown).

To determine whether probeset performance was consistent between samples, we compared probeset CVs for the 7434 probesets that were present in both the cell line and tumor RNA samples. As shown in Figure 5A, the correlation between probeset CVs in the two experiments was low. This is not entirely surprising because CV is related to signal intensity and the relative signal intensities were different for most probesets between samples. If the results are divided into quadrants using the 40% CV cutoff described above, one can see that the majority of probesets showed less than a 40% CV in both the cell line and tumor homogenate (Figure 5A). Table 1 lists the probesets with the highest variability in the other three quadrants. In general, probesets that were highly variable in one sample but not the other tended to have lower median signal intensities in the more variable sample (Figure 5B). It is interesting to note that many of the probesets that showed high variability in both samples were AFFX ribosomal RNA probesets, five of which have been removed from the U133A vs 2.0 GeneChips (Affymetrix website: http://www.affymetrix.com/support/help/faqs/hgu133_2/faq_14.jsp).

Discussion

In this study, we have performed extended precision profiling of the Affymetrix human microarray platform, using two generations of chips and greater than 100 analyses. The design was structured to mimic potential clinical laboratory performance of this testing, by varying reagent lots, machines, and operators, and by performing analyses throughout an extended period of time. No postanalytical controls were applied, and all arrays were

Figure 5. Comparison of U133A probe CVs from two different materials. **A:** The CVs for the 7434 probesets present in both the cell line and tumor RNA samples are compared. Quadrants are divided at 40% CV for each sample, percentages indicate the number of probesets in that quadrant. **B:** Relative signal intensities of probesets that performed differently in the two samples (**A: top left and bottom right**). **Diagonal line** represents equal intensity. Probesets tended to have lower signal intensities in the sample with the higher CV.

included in the final precision analysis regardless of chip hybridization parameters such as %P, background, or 3':5' ratios. Even under these extremely inclusive conditions, overall precision was surprisingly good for an assay of this complexity, with mean probeset CVs of <30% for present probesets in all experiments. Although the precisions of individual probesets varied widely, the majority of present probesets had CVs of <40% under the conditions tested. Not surprisingly, variation increased as signal intensities decreased, with the largest CVs seen in probesets with signal intensities of <100.

Chip-to-chip (residual) variation was the largest component of overall variation in all experiments, suggesting that controllable factors such as reagent lots, analyst, and fluidic station play a limited role in the precision of this technique. This is promising from the standpoint of clinical application of Affymetrix genechip arrays, because it suggests that assays could be performed in a clinical laboratory context without seriously compromising precision. Although the precise cause of the underlying chip-to-chip variation is unknown, possibilities may include subtle differences in within-lot chip manufacturing, chip-to-chip surface variability in hybridization conditions, or likely a composite of multiple factors. Among the controllable factors, chip lot and cRNA cleanup kits were the most notable, suggesting that efforts to better standardize these steps would yield the most return in terms of improved precision.

By the nature of this experimental design, the CVs generated here represent an upper estimate of overall variability because no acceptance/rejection criteria were applied to chip output. The elimination of chips by using hybridization QC metrics⁸ would likely reduce the CVs for many probesets. We intentionally did not include such

manipulations in our study because the precise criteria used can vary from institution to institution, and because we wanted to capture an estimate of the maximal variability of the assay throughout time. However, adoption of standardized criteria could potentially be useful in this regard.

Data generated using the paradigm demonstrated in this study can be applied in a number of ways. One use would be to establish QC material for potential clinical applications. Probeset CVs could be used to derive control ranges for a standard RNA sample, which could then be run as a QC material in production runs. These control ranges could be used to set acceptance/rejection criteria for each run either based on a whole-chip analyses, or on a limited subset of probesets. This would be most applicable for uses such as tumor profiling, in which the QC criteria could be limited to those probesets that make up a tumor signature. Because CVs can vary substantially between probesets and within probesets depending on signal intensity, it will be important to empirically determine appropriate CV ranges in an assay-specific manner using a control material that represents the appropriate probesets for the target application.

Another potential application of the approach demonstrated here is to identify probesets with consistently poor precision so that inclusion of these probesets in diagnostic signatures or in QC metrics can be avoided. One limitation of this study is that not all probesets represented on the chips could be examined because many RNAs were not expressed in the tissues tested. Fifty-three percent of the U133A probesets were present in one of the two samples tested. In addition, because CV is related to signal intensity, targets that happened to have low expression levels in the two RNAs tested could ap-

Table 1. U133A Probesets Showing the Highest Variability

Probeset	Gene	Description	% CV		Median signal	
			Tumor	Cell line	Tumor	Cell line
Group A: high CV in tumor and cell line						
AFFX-HUMRGE/M10098.5.at			273.9	111.9	647	3395
AFFX-r2-Hs18SrRNA-5.at			224.8	107.2	3205	1041
AFFX-r2-Hs28SrRNA-5.at			142.2	143.1	4521	561
AFFX-HUMRGE/M10098.M.at			72.5	109.7	205	729
AFFX-r2-Hs18SrRNA-M.x.at			73.6	101.8	1058	648
211696.x.at	HBB	Hemoglobin, beta	314.5	54.5	1479	769
204118.at	CD48	CD48 antigen (B-cell membrane protein)	257.1	59.2	474	3524
AFFX-HUMRGE/M10098.3.at			215.1	85.0	1112	1402
AFFX-r2-Hs18SrRNA-3.s.at			202.6	81.4	100	1662
217466.x.at	RPS2	Ribosomal protein S2	101.6	73.9	1363	2333
AFFX-r2-Hs28SrRNA-3.at			99.8	85.4	15686	1984
AFFX-BioDn-5.at			78.0	87.1	189	4056
Group B: high CV in cell line						
213757.at	EIF5A	Eukaryotic translation initiation factor 5A	30.5	182.1	297	788
214001.x.at	RPS10	Ribosomal protein S10	36.5	174.5	536	318
212952.at	CALR	Calreticulin	41.6	168.4	3160	644
213350.at	RPS11	Ribosomal protein S11	45.5	153.0	253	3548
219138.at	RPL14	Ribosomal protein L14	25.9	143.1	399	289
212044.s.at	RPL27A	Ribosomal protein L27a	27.6	128.4	2236	1466
202028.s.at	CITED1	Cbp/p300-interacting transactivator	16.0	127.2	3516	2093
208937.s.at	ID1	Inhibitor of DNA binding 1	33.9	126.6	138	3200
213826.s.at	NULL	ESTs, highly similar to H33.HUMAN HISTONE H3.3	23.3	118.8	466	325
216246.at	RPS20	Ribosomal protein S20	35.4	116.4	1095	657
221943.x.at	RPL38	Ribosomal protein L38	10.5	108.5	4014	2752
Group C: high CV in tumor						
217028.at	CXCR4	Chemokine (C-X-C motif), receptor 4 (fusin)	335.1	33.1	185	195
214617.at	PRF1	Perforin 1 (pore forming protein)	333.9	33.6	171	185
210915.x.at	TRB@	T cell receptor beta locus	319.9	21.1	317	83
213193.x.at	TRB@	T cell receptor beta locus	317.8	24.3	384	1247
219014.at	LOC51316	Hypothetical protein	303.5	26.2	187	646
39248.at	AQP3	Aquaporin 3	302.0	26.7	106	404
212588.at	PTPRC	Protein tyrosine phosphatase, receptor type, C	291.4	41.8	263	2264
211339.s.at	ITK	IL2-inducible T-cell kinase	285.9	34.6	147	833
211742.s.at	EVI2B	Ecotropic viral integration site 2B	279.1	40.8	193	329
201422.at	IFI30	Interferon, gamma-inducible protein 30	272.3	26.9	547	158

Only probesets that were present in both experiments were included.

pear to have an unrealistically high CV in our study that might be improved at higher target levels. This is a natural limitation of using cellular RNA as a test material because no sample will express all genes represented on the chip. Once again, this demonstrates the need to tailor potential QC materials to the assay in question so that all of the relevant probesets are adequately represented in the material.

In conclusion, we have described a paradigm for precision profiling of Affymetrix microarrays that could be used as part of a validation protocol for clinical applications of this technology. The ultimate goal with any test is to eliminate as much analytical variability as possible, thereby reducing the contribution of nonbiological effects on the reported result. The results reported here give an estimate of the magnitude of these analytical effects, and serve as a marker against which process modifications can be compared. The variability measured in these experiments is probably an upper limit, as no acceptance/rejection criteria were applied to these samples and all data were included in the final analysis. Analytical precision could potentially be improved by such an approach,

either by using standard Affymetrix chip-based parameters such as background and 5':3' ratios or by incorporating control materials in the analytical process, using limits derived from experiments such as these. Defining appropriate control strategies will be important for the widespread dissemination of this platform to support clinical applications of this technology.

References

1. Rosenwald A, Wright G, Leroy K, Yu X, Gaulard P, Gascoyne RD, Chan WC, Zhao T, Haioun C, Greiner TC, Weisenburger DD, Lynch JC, Vose J, Armitage JO, Smeland EB, Kvaloy S, Holte H, Delabie J, Campo E, Montserrat E, Lopez-Guillermo A, Ott G, Muller-Hermelink HK, Connors JM, Braziel R, Grogan TM, Fisher RI, Miller TP, LeBlanc M, Chiorazzi M, Zhao H, Yang L, Powell J, Wilson WH, Jaffe ES, Simon R, Klausner RD, Staudt LM, Alizadeh AA, Widhopf G, Davis RE, Pickeral OK, Rassenti LZ, Botstein D, Byrd JC, Grever MR, Cheson BD, Chiorazzi N, Kipps TJ, Brown PO: Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma: relation of gene expression phenotype to immunoglobulin mutation

- genotype in B cell chronic lymphocytic leukemia. *J Exp Med* 2003, 198:851–862
2. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson Jr J, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM: Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000, 403:503–511
3. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, Gascoyne RD, Muller-Hermelink HK, Smeland EB, Giltner JM, Hurt EM, Zhao H, Averett L, Yang L, Wilson WH, Jaffe ES, Simon R, Klausner RD, Powell J, Duffey PL, Longo DL, Greiner TC, Weisenburger DD, Sanger WG, Dave BJ, Lynch JC, Vose J, Armitage JO, Montserrat E, Lopez-Guillermo A, Grogan TM, Miller TP, LeBlanc M, Ott G, Kvaloy S, Delabie J, Holte H, Krajci P, Stokke T, Staudt LM: The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med* 2002, 346:1937–1947
4. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ: High density synthetic oligonucleotide arrays. *Nat Genet* 1999, 21(Suppl 1):20–24
5. Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H, Brown EL: Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 1996, 14:1675–1680
6. Fraser CG: *Biological Variation: From Principles to Practice*. Washington DC, AACCC Press, 2001
7. Dumur CI, Nasim S, Best AM, Archer KJ, Ladd AC, Mas VR, Wilkinson DS, Garrett CT, Ferreira-Gonzalez A: Evaluation of quality-control criteria for microarray gene expression analysis. *Clin Chem* 2004, 50:1994–2002
8. The Tumor Analysis Best Practices Working Group: (Hoffman EP, Awad T, Palma J, Webster T, Hubbell E, Warrington JA, Spira A, Wright G, Buckley J, Triche T, Davis R, Tibshirani R, Xiao W, Jones W, Tompkins R, West M): Expression profiling—best practices for data generation and interpretation in clinical trials. *Nat Rev Genet* 2004, 5:229–237
9. Novak JP, Sladek R, Hudson TJ: Characterization of variability in large-scale gene expression data: implications for study design. *Genomics* 2002, 79:104–113
10. Bakay M, Chen YW, Borup R, Zhao P, Nagaraju K, Hoffman EP: Sources of variability and effect of experimental approach on expression profiling data interpretation. *BMC Bioinformatics* 2002, 3:4
11. Shippey R, Sendera TJ, Lockner R, Palaniappan C, Kaysser-Kranich T, Watts G, Alsobrook J: Performance evaluation of commercial short-oligonucleotide microarrays and the impact of noise in making cross-platform correlations. *BMC Genomics* 2004, 5:61
12. Schultz RM, Chen VJ, Bewley JR, Roberts EF, Shih C, Dempsey JA: Biological activity of the multitargeted antifolate, MTA (LY231514), in human cell lines with different resistance mechanisms to antifolate drugs. *Semin Oncol* 1999, 26(Suppl 6):68–73
13. Affymetrix GeneChip Expression Analysis Technical Manual. Santa Clara, Affymetrix, 2002, pp 1.1.3–2.4.16
14. Affymetrix Microarray Suite 5.1 User's Guide. Santa Clara, Affymetrix, 2002
15. Affymetrix GeneChip expression analysis: Data analysis fundamentals. Santa Clara, Affymetrix, 2002, pp 13–15
16. Littell RC, Milleken GA, Stroup WW, Wolfinger RD: *SAS System for Mixed Models*. Cary, SAS Institute Inc., 1996
17. Fleiss JL (Ed): *The Design and Analysis of Clinical Experiments*. Philadelphia, John Wiley and Sons, 1985, p. 21 (equation 1.38)