Research article

Computational method for reducing variance with Affymetrix microarrays

Stephen Welle*[1], Andrew I Brooks[2] and Charles A Thornton[3]

Address: [1]Departments of Medicine, Pharmacology & Physiology, University of Rochester, Rochester, NY 14642, USA, [2]Functional Genomics Center, Department of Environmental Medicine, University of Rochester, Rochester, NY 14642, USA and [3]Department of Neurology, University of Rochester, Rochester, NY 14642, USA

E-mail: Stephen Welle* - stephen_welle@urmc.rochester.edu; Andrew I Brooks - andrew_brooks@urmc.rochester.edu; Charles A Thornton - charles_thornton@urmc.rochester.edu

*Corresponding author

Abstract

Background: Affymetrix microarrays are used by many laboratories to generate gene expression profiles. Generally, only large differences (> 1.7-fold) between conditions have been reported. Computational methods to reduce inter-array variability might be of value when attempting to detect smaller differences. We examined whether inter-array variability could be reduced by using data based on the Affymetrix algorithm for pairwise comparisons between arrays (ratio method) rather than data based on the algorithm for analysis of individual arrays (signal method). Six HG-U95A arrays that probed mRNA from young (21–31 yr old) human muscle were compared with six arrays that probed mRNA from older (62–77 yr old) muscle.

Results: Differences in mean expression levels of young and old subjects were small, rarely > 1.5-fold. The mean within-group coefficient of variation for 4629 mRNAs expressed in muscle was 20% according to the ratio method and 25% according to the signal method. The ratio method yielded more differences according to t-tests (124 vs. 98 differences at P < 0.01), rank sum tests (107 vs. 85 differences at P < 0.01), and the Significance Analysis of Microarrays method (124 vs. 56 differences with false detection rate < 20%; 20 vs. 0 differences with false detection rate < 5%). The ratio method also improved consistency between results of the initial scan and results of the antibody-enhanced scan.

Conclusion: The ratio method reduces inter-array variance and thereby enhances statistical power.

Background

Affymetrix microarrays are used by many laboratories to study differences in gene expression associated with experimental treatments, diseases, development, aging, and other conditions. Typically, an arbitrary value for expression ratios (or fold-change values) is chosen to define significant differences in gene expression between conditions. For example, in several studies of aging [1–6], only differences > 1.7-fold in magnitude were considered to be significant. None of the reports indicated whether there were smaller effects that were statistically significant. It has been pointed out that statistically significant differences in gene expression often are of small magnitude (sometimes as low as 1.2-fold), and that larger effects of-

ten are artefacts of high variance [7,8]. For those interested in detecting these smaller effects, it is important to minimize nonspecific sources of inter-array variance.

To understand the approach described in this report, it is necessary to understand the design of Affymetrix microarrays and analysis software (Microarray Suite). There are multiple probe pairs for each mRNA (8–20 for the arrays used in the present study). A probe pair consists of a 25 base oligonucleotide that matches an mRNA sequence (perfect match, or PM probe) and an oligonucleotide with a mismatched base in the center (MM probe). The specific hybridization signal for each probe pair is the difference between the PM intensity and the MM intensity (although the latest version of Affymetrix Microarray Suite, 5.0, has special rules for handling MM probes that have higher signals than their PM partner). No single hybridization condition is optimal for all oligonucleotide probes, so it is inevitable that there is variability among the signals within a probe set. The expression level reported for each probe set (by the Affymetrix "absolute analysis" algorithm) is based on a weighted average of the signals from the individual probe pairs, with signals near the median given more weight than those far from the median. We refer to this as the signal method in this report. The weights assigned to each probe pair can vary from one array to another, but it is unclear whether variable weighting adds significantly to inter-array variance. Microarray Suite also has a procedure ("comparative analysis" algorithm) for comparing two arrays at the level of individual probe pairs. With this algorithm, ratios of signals (PM-MM for each probe pair) from one array to those of the other array are computed first. Weighted averages of these ratios are then computed. We refer to this as the ratio method. This method is supposed to be more precise than the signal method for inter-array comparisons. Thus, many investigators use this algorithm for all possible one-to-one comparisons across groups (e.g., 9 comparisons for 3 arrays per group) and report the average of the ratios as the change in gene expression [1–5,9]. A problem with this approach is that there is no absolute or relative expression level assigned to each mRNA on individual arrays, so that formal statistical approaches (e.g., t-tests) cannot be used to rate the statistical significance of differences. In this report, we describe how we circumvented this problem by using the ratio method to generate a composite gene expression score for each mRNA on each array.

The procedure used to estimate the statistical significance of differences determines which genes, and how many genes, are defined as being differentially expressed. For a comparison between two groups, the t-test is the most commonly used procedure in biological research. However, with 6 arrays per group, even a single outlier can markedly reduce the value of t even when there is no overlap

between groups. Therefore, we also used the nonparametric rank sum test, which is insensitive to a skewed distribution. False detection rates were estimated with the Significance Analysis of Microarrays (SAM) program [8].

Results
Normalization method
Comparisons among arrays are meaningful only after accounting for variability in overall target concentration ("target" is the Affymetrix nomenclature for a labeled cRNA that hybridizes with a probe), hybridization efficiency, staining, etc. The normalization procedure recommended by Affymetrix is to multiply raw signals by a scaling factor such that the trimmed mean (excluding 2% highest and 2% lowest) of signals is always the same (500 in this study). This procedure could be problematic if a variable proportion (>2%) of the signals are beyond the linear range of the system. Another concern about the normalization procedure was that the majority of the targets did not produce signals that were significantly greater than those caused by nonspecific hybridization. After the recommended normalization procedure was applied, we confirmed that there was negligible inter-array variability of the mean signal (with 5% of signals trimmed from both the high and low ends) across the 4629 targets that passed the presence / absence filter (described in the next section). The trimmed mean was 649 ± 14 (standard deviation) arbitrary units for the 6 arrays probing RNA from young muscle, and 643 ± 18 for the 6 arrays probing RNA from older muscle. These data were not used to re-scale the arrays because the variance would have been reduced by less than 2%. Plots of signals from individual arrays versus the average of all 12 arrays generally showed the expected scatter around the line of identity (Figure 1A), but a few showed systematic deviations either above or below the line of identity for signals $> \sim 10^4$ arbitrary units (worst-case array shown in Figure 1B). While this problem might be addressed by using a different scaling factor for high-intensity signals [10], few targets produced such high signals, and the magnitude of the effect was small. Thus, the Affymetrix normalization method was employed without modification.

Exclusion of targets based on the absolute detection algorithm
Microarray Suite estimates probabilities that targets are absent ($P_{detection}$) based on ratios of signals from PM probes to those of MM probes, together with the degree of consistency across the multiple probe pairs for each target. As illustrated in Figure 2, the average signal from the multiple probe pairs cannot be used to decide whether a target should be considered present or absent. We restricted the data analyses to targets for which $P_{detection}$ was less than 0.1 for at least 3 of the 6 samples from either the younger or older group. This filter reduced the number of targets
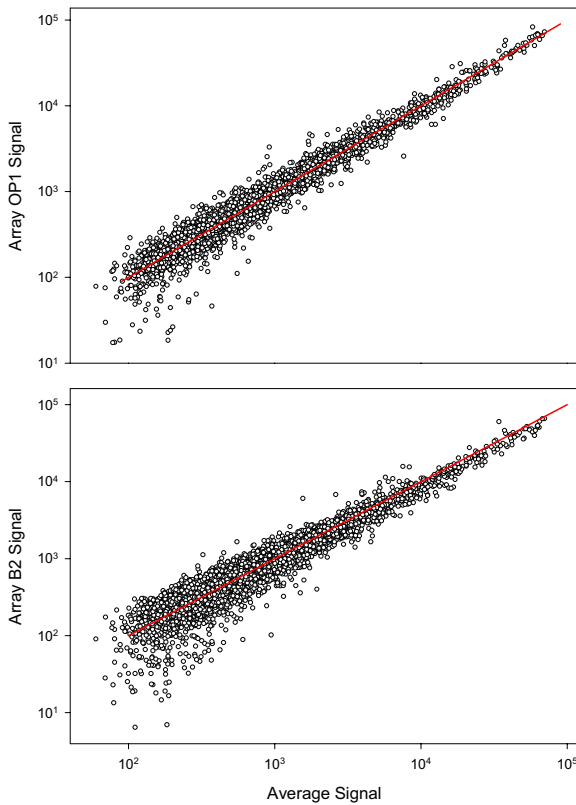
Figure 1
Individual arrays vs. mean of all arrays A. Scatter plot of signals from 4629 probe sets on a typical array vs. mean signals from all 12 arrays. Line of identity is shown. B. Worst-case scatter plot. Same as plot A, except vertical axis represents a different array, the one with the most consistent deviation from the mean of all 12 arrays at high expression levels. Note that almost all signals > 10^4 arbitrary units are below the line of identity.

Figure 2
$P_{detection}$ vs. signal Signal, in arbitrary units, is the average PM-MM intensity difference across all 8–20 probe pairs within a set. $P_{detection}$ is the probability that a target is absent, based on the consistency of PM/MM ratios within a probe set. Values are based on a single array.

included in the statistical analyses to 4629. While excluding data does not affect the nominal value of P for each comparison made with a t-test or rank sum test, it significantly reduces the estimated false detection rate (see **t-Tests** and **SAM** below).

Signal method vs. ratio method

When two arrays are compared, the gene expression ratios obtained by the signal method and those obtained by the ratio method (see **Background** for explanation of terms) are highly correlated. However, the results often differ by more than 1.5-fold (Figure 3). The advantage of the signal method is that Microarray Suite provides, for each target, a number describing the level of gene expression (in arbitrary units) that can be used for t-tests or other statistical procedures. However, according to Affymetrix (Microarray Suite 5.0 User's Guide), comparisons between arrays
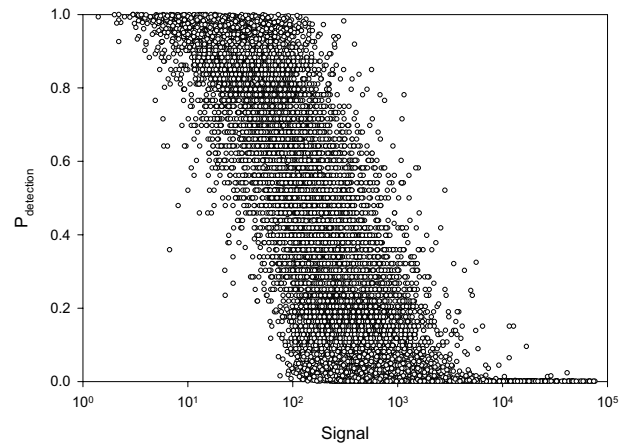
are more precise when the ratio method is used, so the values on the horizontal axis of Figure 3 should be more accurate. The Affymetrix ratio method only provides ratios between two arrays, and does not provide gene expression values for the individual arrays that can be used with standard tests of statistical significance. We therefore extended the ratio method to generate a relative expression score for each target on each array, as follows:

The first step was to name one of the arrays as the baseline in the comparative analysis program (Microarray Suite 5.0). Every other array included in the study was then compared with that baseline array. This procedure yielded, for each target, a set of 11 expression ratios (*r*) representing the relative expression level on each array compared with the baseline array.

The next step was to compute, for each target, the expression level (*R*) of the baseline array relative to all 12 arrays included in the study. For array #1, *R* was computed with the formula:

$$R_1 = 12/(1 + r_{2 \text{ vs. } 1} + r_{3 \text{ vs. } 1} + ... + r_{12 \text{ vs. } 1})$$

The value of 1 in the denominator of this formula represents the comparison of array #1 with itself. The number of arrays is the numerator rather than the denominator in this formula because the Affymetrix comparative analysis program sets the baseline array as the denominator, so that values of *r* are the inverse of the relevant ratios.
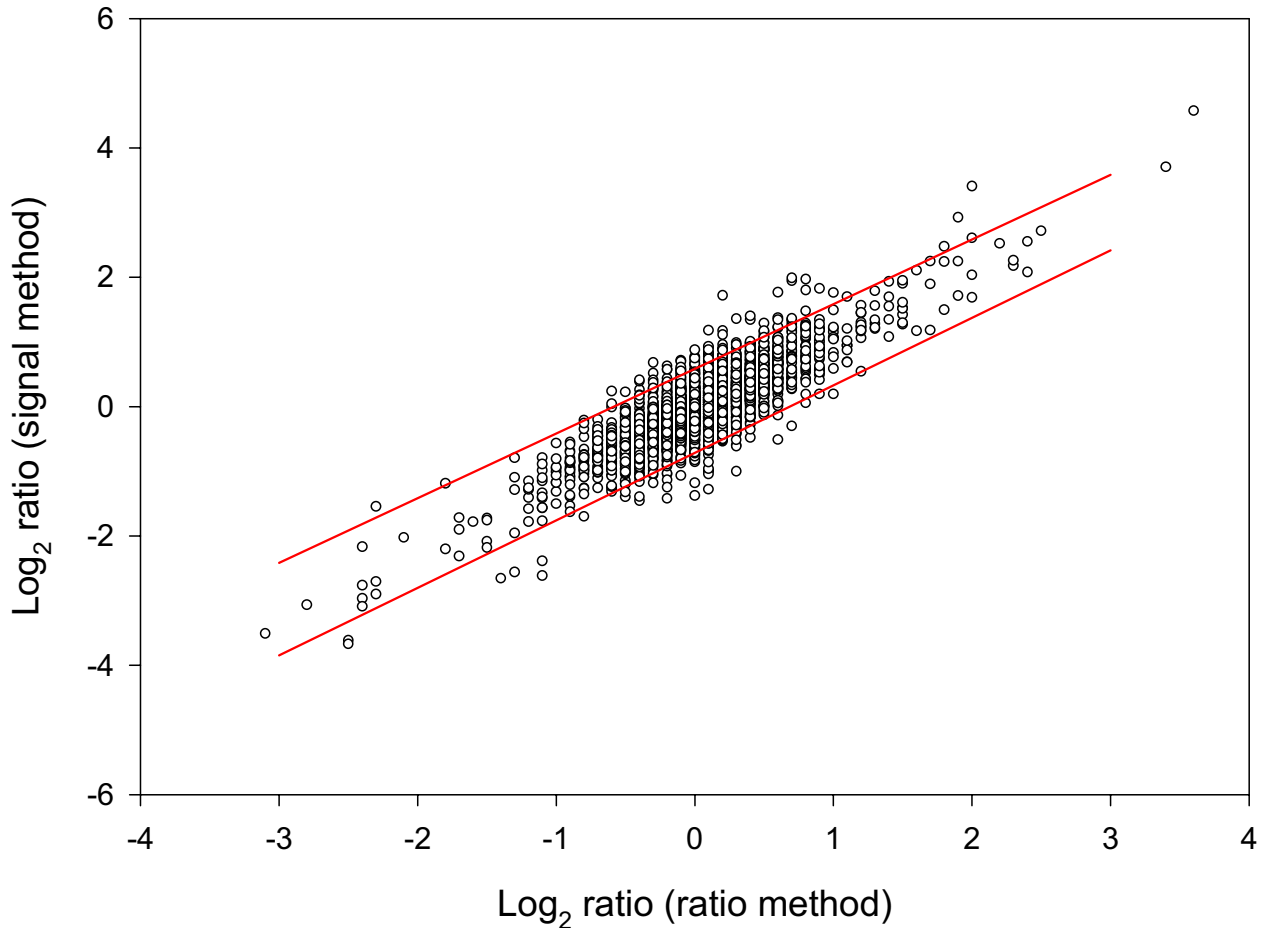
Figure 3
Comparison of two arrays by different methods Horizontal axis shows the ratios between two arrays, for 4629 targets, according to the comparative analysis algorithm, which is the basis of the ratio method. Vertical axis shows ratios between the same arrays according to the absolute analysis algorithm, which is the basis of the signal method. Points outside the red lines have more than 1.5-fold divergence between methods.

Table 1: Number of differences detected: comparison of signal and ratio methods

Criterion for difference	Signal Method	Ratio Method
t-test, P < 0.01	98	124
rank sum test, P < 0.01	85	107
SAM, false detection < 20%	56	124
SAM, false detection < 5%	0	20

A different array was then named as the baseline. E.g., for array #2 as the baseline:

$$R_2 = 12/(r_{1 \text{ vs. } 2} + 1 + r_{3 \text{ vs. } 2} + ... + r_{12 \text{ vs. } 2})$$

These steps were repeated until all 12 arrays had been named as the baseline. The values R_1 through R_{12} were then used for comparisons between age groups with t-tests, rank sum tests, and SAM as described below.

For the 4629 probe sets that passed the presence / absence filter, the expression ratios (mean value in old group / mean value in young group) generated by the signal method and those generated by the ratio method were highly correlated (r = 0.89). There also was a fairly good correlation between the signal and ratio methods with respect to the level of statistical significance (log P) of the age-related differences (r = 0.74). The advantage of the ratio method was that it usually reduced the within-group variance so
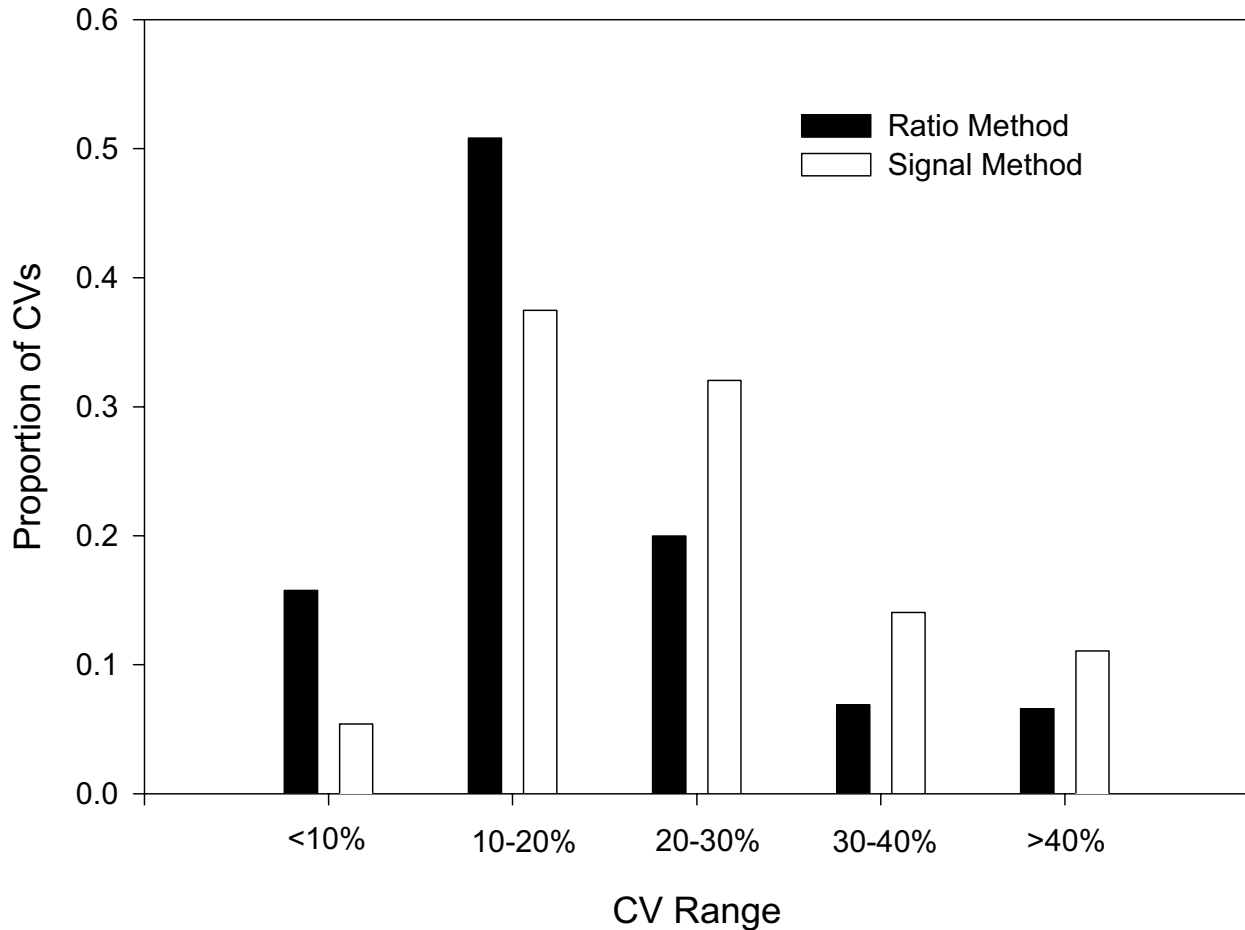
Figure 4
Frequency distribution of coefficients of variation (CVs) Distribution of 4629 CVs obtained by the ratio method (solid bars) and the signal method (open bars). CVs are average of within-group CVs in young and old groups.

Table 2: Correlation coefficients of results of initial scan and antibody-enhanced scan for 4629 probe sets with respect to expression ratios (mean old / mean young) and statistical significance (from t-tests) of the differences between young and old

Variable	Signal Method	Ratio Method
expression ratio	0.57	0.75
P, from t-test	0.39	0.56
log P, from t-test	0.55	0.68

that the same mean difference between young and old was associated with a higher level of statistical significance. The average within-group coefficient of variation (CV, standard deviation / mean) was 20% with the ratio meth-od and 25% with the signal method (average CVs were the same for young and old groups). The distribution of CVs improved significantly with the ratio method (Figure 4). Table 1 shows that more differences were detected by the ratio method whether we used t-tests, rank sum tests, or SAM to define significant differences. Moreover, consistency between the initial scan and the antibody-enhanced scan was significantly improved by the ratio method, for both expression ratios and for the statistical significance of differences between young and old (Table 2). With the signal method, 38% of the differences significant at $P < 0.01$ (by t-test) on the initial scan were also significant at $P < 0.01$ on the antibody-enhanced scan, and 65% were significant at $P < 0.05$ on the antibody enhanced scan. With the ratio method, 51% of the differences significant at $P < 0.01$ on the initial scan were also significant at $P < 0.01$ on the antibody-enhanced scan, and 77% were significant at $P < 0.05$ on the antibody enhanced scan.
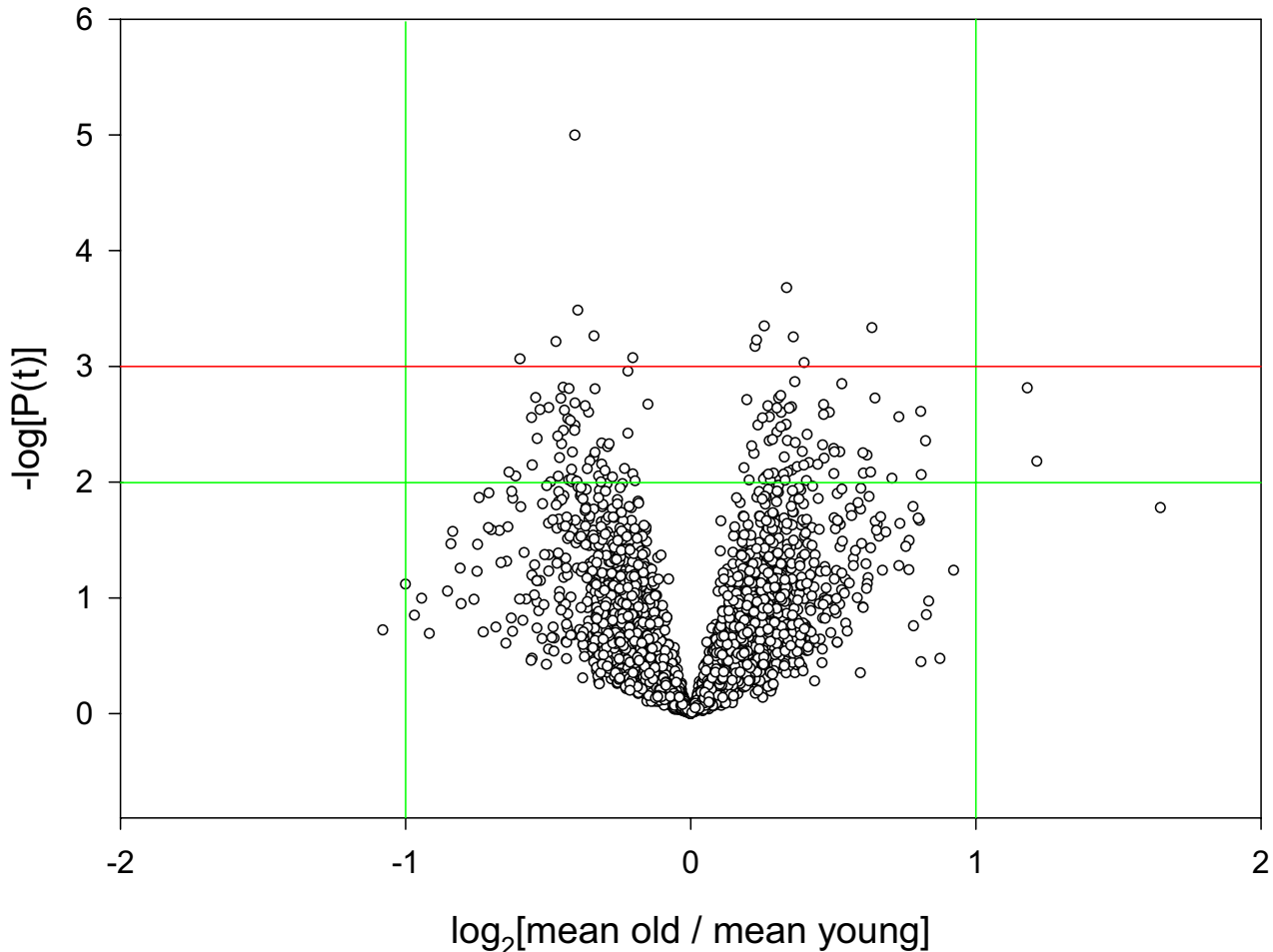
Figure 5
Volcano plot Statistical significance by t-tests [-log(P)] vs. expression ratio (mean old / mean young) for 4629 targets that passed the presence / absence filter. Note $\log_2$ scale on horizontal axis. Vertical lines represent 2-fold difference between young and old. Upper horizontal line represents P = 0.001. Lower horizontal line represents P = 0.01.

t-Tests

A plot of expression ratio vs. statistical significance (Figure 5) shows that differences with high statistical significance (by t-test) usually were less than 1.7-fold in magnitude. The validity of at least some of the small differences is demonstrated by the fact that ~1.3-fold differences were detected (P < 0.01) for 17 genes encoding proteins involved in mitochondrial electron transport or ATP synthesis (Table 3). This finding replicates our previous research, in which SAGE and quantitative RT-PCR assays demonstrated ~1.3-fold declines in older muscle of several mRNAs encoding components of NADH dehydrogenase, cytochrome c oxidase, and ATP synthase complexes [11]. For all of these mRNAs, both the signal and ratio methods detected the differences at P < 0.03, whereas the ratio method was a bit more likely to detect them at P < 0.01

(14/17 genes) than was the signal method (12/17 genes, Table 3).

The P values generated by the t-tests were not adjusted for multiple comparisons. However, a Bonferroni correction would be too stringent for exploratory research [12]. A useful alternative to P in studies involving thousands of comparisons is the estimated false detection rate, which is the ratio of the expected number of chance differences (P × number of comparisons) to the number of differences observed. If we use P < 0.01 to define a significant difference, we should expect ~46 chance differences (0.01 × 4629 comparisons) if there is no effect of aging on gene expression. Because 124 differences were significant at P < 0.01 (by the ratio method), the estimated false detection rate was 46/124, or 37%. When no presence / absence filter was applied (12533 probe sets included in the analy-

Table 3: Reduced expression in older muscle of genes involved in energy metabolism (ER = expression ratio = mean value in old / mean value in young)

Gene	ER signal method	ER ratio method	P signal method	P ratio method
ubiquinol cytochrome c reductase binding protein	.72	.73	.0009	.0015
cytochrome c oxidase Vb	.65	.76	.0214	.0021
cytochrome bc-1 complex core protein II	.82	.77	.0137	.0022
mitochondrial ADP/ATP translocator	.74	.74	.0036	.0016
cytochrome c oxidase VIIc	.61	.75	.0026	.0032
ATP synthase subunit F6	.74	.72	.0078	.0006
cytochrome c oxidase VIc	.78	.79	.0252	.0016
cytochrome c oxidase VIIb	.69	.73	.0021	.0019
adenylate kinase I	.76	.76	.0155	.0003
ubiquinol cytochrome c reductase hinge protein	.71	.72	.0082	.0041
holocytochrome c-type synthase	.77	.81	.0016	.0046
cytochrome c oxidase VIIa2	.73	.73	.0099	.0062
NADH dehydrogenase KFYI	.83	.87	.0199	.0098
mitochondrial aspartate aminotransferase	.72	.76	.0079	.0100
cytochrome c oxidase 4	.81	.82	.0057	.0104
cytochrome c	.69	.70	.0016	.0108
ATP synthase c (P1 gene)	.75	.77	.0052	.0112

sis), the estimated false detection rate (ratio method) increased from 37% to 73% because there were fewer differences (at P < 0.01) among the "absent" mRNAs than expected by chance (48 observed vs. 79 expected by chance).

Mann-Whitney rank sum tests
The Mann-Whitney rank sum test was used to detect transcripts for which there was little or no overlap of values between groups. This test revealed 107 differences at P < 0.01 (exact P = 0.00866 at rank sum cutoff values) when the ratio method was used. We would expect to find 40 differences by chance alone (0.00866×4629 genes), so the false detection rate (40/107 = 37%) is the same as that estimated from t-tests. There were 21 differences significant at P < 0.01 by rank sum tests but not by t-tests according to the ratio method. Thus, for exploratory research being done to generate lists of mRNAs that warrant further study, use of both parametric and nonparametric tests is one way to significantly expand the list.

SAM
SAM computes a value, termed d [$d = (\text{mean}_1 - \text{mean}_2)/(s_d + s_0)$], that is similar to t [$t = (\text{mean}_1 - \text{mean}_2)/s_d$]. The "exchangeability factor", s_0, minimizes the number of extreme d values among targets with small variances in signal intensity. When absolute signals are analyzed, these small variances are associated with targets that are more difficult to quantify accurately because of low concentra-

tions. We already have filtered most of these targets from the analysis. When data based on the signal method were analyzed, s_0 was very small (lowest percentile of the s_d values). When relative expression levels were computed by the ratio method, all means were close to 1 regardless of the absolute signal intensity. In this case, s_0 was large (53rd percentile of the s_d values), and lower s_d values were associated with stronger rather than weaker signals. This problem precluded the use of SAM for data normalized in this fashion. However, by multiplying each value of *R* by the gene-specific mean signal (mean of all 12 arrays), we generated a data set compatible with SAM.

SAM lists genes for which d exceeds (by an adjustable threshold termed Δ) the value that would be expected by chance (d_e). Values of d_e are generated by computing the d distribution numerous times with random permutations of the group assignments (we instructed SAM to perform 100 permutations). The average distribution from these permutations defines the values of d_e. Reducing Δ expands the list of "significant" genes, but also increases the false detection rate. When we chose a value of Δ corresponding to a false detection rate < 20%, there were 124 differences according to the ratio method but only 56 differences according to the signal method. There were 20 differences for which the false detection rate was < 5% when the ratio method was used (including 9 for genes involved in energy metabolism that are listed in Table 3),

but none when the signal method was used. When no presence / absence filter was applied, the highest-ranked differences had false detection rates of 30% even with the ratio method, and only 34 genes achieved this level. Thus, it is very important to remove noisy data before using SAM.

Discussion

Careful subject selection and consistency in experimental conditions, sample collection procedures, and sample processing obviously are needed if small differences in gene expression are to be detected. A further reduction in total within-group variance can be achieved by using the ratio method described in this report. This method is based on the Affymetrix comparative analysis algorithm, which was designed for comparisons between two arrays. To use the procedure for groups rather than individual arrays, we assigned each target on each array a score that was the average ratio from all one-to-one comparisons of that array with every array included in the study. The best illustration of the increase in power gained by the ratio method was the fact that that 20 differences were below the 5% false detection rate (by SAM) when this method was used, whereas no differences below the 5% false detection rate were found when the signal method was used. The major drawback of the ratio method is increased computational time.

It has been suggested that inter-array variance can be reduced by ignoring data from MM probes, or by using an alternative computation to take advantage of the MM data [13–16]. In previous versions of Microarray Suite, MM signals were always subtracted from PM signals, which often led to negative expression levels. The newer version (5.0), used in this study, has a different procedure for dealing with high signals from MM probes. It is not clear whether alternative algorithms for using the MM signals, or ignoring MM signals, would improve the accuracy of the data. We therefore used the Affymetrix algorithm, which is likely to be used by most investigators until there is definitive evidence that alternative methods are superior.

There is no consensus about the optimal statistical approach for finding differences in expression among thousands of genes. When a specific hypothesis is being tested, "shopping" for the best statistical test to support the hypothesis should be avoided. In contrast, when the goal is to generate descriptive information to guide decisions about future research directions, there is no reason not to use multiple approaches to obtain as much information as possible. For a comparison of two groups, the t-test is simple and robust, and does not require special software. Some investigators may be wary about using t-tests rather than mean differences to rank genes because one or two extreme values can reduce t even when there is no overlap

between groups. The rank sum test can be used to detect such effects. Log transformation of the data also can minimize the impact of outliers with high signals. However, log transformation reduces t when the outliers are close to zero. It has been suggested that this feature of the log transformation is advantageous because it can exclude effects that are artefacts of noisy data at low expression levels [17]. We prefer to filter noisy data by using the $P_{detection}$ algorithm.

SAM [8] is an alternative to t-tests or rank sum tests. The false detection rates computed by SAM were markedly increased when we did not apply our presence / absence filter. When the filter was used, SAM generated a lower estimate of the number of false differences than estimates based on multiplying P(t) or P(rank sum) by the number of comparisons. Perhaps this observation can be explained by the fact that expression levels of many genes are correlated with expression levels of many others. The number of random differences to be expected among thousands of comparisons with t-tests or rank sum tests is based on the assumption that comparisons are independent of one another. SAM computes the false detection rate with a method that does not rely on this assumption.

Some of the arrays included in this study (n = 4) probed RNA pooled from multiple subjects, whereas others (n = 8) probed RNA from individual subjects. The heterogeneous nature of the samples conceivably could influence the statistical properties of this data set. However, there was good uniformity among arrays in terms of scaling factors and percentages of probe sets with $P_{detection} < 0.1$, and both age groups were comprised of 2 pools and 4 individual samples, which should minimize the influence of using both pooled and individual samples (see Additional file 1). After the data analyses described in this report were completed, we reanalyzed the individual samples along with 8 new individual samples (total of 8 individual samples per age group) using U133A and U133B arrays, which have ~44000 probe sets with 11 probe pairs per set. The reduction of within-group variance by the ratio method was replicated (Table 4), demonstrating that it is not unique to U95A arrays, and is not an artefact of including both pooled and individual samples. We cannot guarantee that the same reduction in variance will occur in all cases. If variance caused by biological heterogeneity or by inconsistent laboratory procedures is very high, then the difference between the signal and ratio methods might be trivial in relation to overall variance. The proposed computational method appears to reduce the inflation of variance caused by variable weighting of individual probes within a set, but cannot compensate for variance from other sources.

Table 4: Mean CVs for all targets passing presence / absence filter in analysis of U133A and U133B arrays

	Mean CV with signal method (%)	Mean CV with ratio method (%)	Reduction in mean CV with ratio method (%)
U133A, initial scan, 9276 targets			
8 samples from young men	26.5	21.2	20
8 samples from old men	26.5	19.2	28
U133A, antibody-enhanced scan, 10665 targets			
8 samples from young men	25.8	20.2	22
8 samples from old men	23.5	18.4	22
U133B, initial scan, 5766 targets			
8 samples from young men	29.8	23.9	20
8 samples from old men	27.5	22.1	20
U133B, antibody-enhanced scan, 7571 targets			
8 samples from young men	28.1	22.2	21
8 samples from old men	26.1	20.4	22

The data generated in this study have been deposited in the NCBI Gene Expression Omnibus (GEO, accession numbers GSM2390 through GSM2401, Series accession number GSE80) [http://www.ncbi.nlm.nih.gov/geo/] and the AMDeC Microarray Resource Center Gene Expression Database [www.amdec.org]. Both the signal data and the ratio data have been deposited. The Affymetrix files (*.exp, *.dat, *.cel, *.chp) can be obtained from the corresponding author.

Conclusions

The ratio method reduces inter-array variance and thereby enhances statistical power. SAM is very sensitive to noisy data, which should be removed a priori.

Methods

Subjects

The subjects were 16 young (21–31 yr) and 19 older men (62–77 yr). All had normal neuromuscular function and were healthy according to history, physical examination, and laboratory tests. None of them was involved in any type of regular exercise program involving strenuous activity. All subjects gave written consent after the procedures and risks were explained. The research was approved by the University of Rochester Research Subjects Review Board.

Procedures

Subjects were asked to refrain from any activity more strenuous than walking for 3 days before the muscle biopsy. Each subject stayed at the University of Rochester General Clinical Research Center the night before the biopsy

to minimize variability between subjects in the amount of activity performed on the day of the biopsy. The needle biopsy was obtained from the vastus lateralis. The skin and muscle were anesthetized with lidocaine a few minutes before tissue removal. The muscle sample was frozen in liquid nitrogen within 30 seconds of removal, then stored at -70°C until analysis.

Analysis of gene expression by high density oligonucleotide arrays

RNA was extracted from the muscle samples as described previously [11]. All RNA samples were of high quality as indicated by the pattern of staining with ethidium bromide in an agarose gel after electrophoretic separation. RNA was probed with Affymetrix HG-U95A high density oligonucleotide arrays, which have ~12500 probe sets. Twelve arrays were examined: four (two for each age group) that probed RNA pooled from 4–8 subjects and eight (four for each age group) that probed RNA from a single subject. Pooling of RNA was necessary in some cases because most of the RNA from some samples had been used for other purposes. Additional file 1 shows the source of RNA, scaling factors, and percentage of transcripts present ($P_{detection} < 0.1$) for each array.

Analytical procedures were carried out using standard operating procedures developed and validated by the University of Rochester Microarray Core Facility. After hybridization and washing, arrays were stained with phycoerythrin-streptavidin, which binds to the biotinylated cRNAs hybridized with the probes. The initial scan detected the fluorescence of the phycoerythrin-streptavidin. The

analyses described in this report are based on data from this initial scan. After the initial scan, signals were amplified by staining the array with biotin-labeled anti-streptavidin antibody followed by phycoerythrin-streptavidin. Analyses of the antibody-enhanced scans are not presented, except for correlations with the initial scans, since the same statistical issues are relevant to both scans. These scans supported the results discussed above. Data from both scans were deposited in the GEO and AMDeC databases.

Software

The statistical algorithms of Microarray Suite 5.0 were used with the default parameters to generate signals, ratios of signals between two arrays, and $P_{detection}$ values. Data generated by Microarray Suite were exported to Microsoft Excel for further analysis. SAM runs within Excel.

Authors' contributions

SW was responsible for data analyses and was the principal author. AIB was responsible for generating microarray data, and contributed to writing the manuscript. CAT was responsible for obtaining the muscle samples, and contributed to writing the manuscript. All authors read and approved the final manuscript.

Additional material

Additional Files

Source of RNA, scaling factors, and percentage of probe sets with Pdetection < 0.1 for 12 arrays included in this study

Click here for file

[http://www.biomedcentral.com/content/supplementary/1471-2105-3-23-S1.xls]

References

1. Lee C-K, Klopp RG, Weindruch R, Prolla TA: **Gene expression profile of aging and its retardation by caloric restriction.** *Science* 1999, **285:**1390-1393
2. Lee C-K, Weindruch R, Prolla TA: **Gene-expression profile of the ageing brain in mice.** *Nature Genetics* 2000, **25:**294-297
3. Ly DH, Lockhart DJ, Lerner RA, Schultz PG: **Mitotic misregulation and human aging.** *Science* 2000, **287:**2486-2492
4. Cao SX, Dhahbi JM, Mote PL, Spindler SR: **Genomic profiling of short- and long-term caloric restriction effects in the liver of aging mice.** *Proc Natl Acad Sci USA* 2001, **98:**10630-10635
5. Kayo T, Allison DB, Weindruch R, Prolla TA: **Influences of aging and caloric restriction on the transcriptional profile of skeletal muscle from rhesus monkeys.** *Proc Natl Acad Sci USA* 2001, **98:**5093-5098
6. Jiang CH, Tsien JZ, Schultz PG, Hu Y: **The effects of aging on gene expression in the hypothalamus and cortex of mice.** *Proc Natl Acad Sci USA* 2001, **98:**1930-1934
7. Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G: **The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster.*** *Nature Genetics* 2001, **29:**389-395
8. Tusher VG, Tibshirani R, Chu G: **Significance analysis of microarrays applied to the ionizing radiation response.** *Proc Natl Acad Sci USA* 2001, **98:**5116-5121
9. Chen Y-W, Zhao P, Borup R, Hoffman EP: **Expression profiling in the muscular dystrophies: identification of novel aspects of molecular pathophysiology.** *J Cell Biol* 2000, **151:**1321-1336
10. Schadt EE, Li C, Su C, Wong WH: **Analyzing high-density oligonucleotide gene expression array data.** *J Cell Biochem* 2000, **80:**192-202
11. Welle S, Bhatt K, Thornton CA: **High-abundance mRNAs in human muscle: comparison between young and old.** *J Appl Physiol* 2000, **89:**297-304
12. Claverie J-M: **Computational methods for the identification of differential and coordinated gene expression.** *Hum Mol Genet* 1999, **8:**1821-1832
13. Naef F, Hacker CR, Patil N, Magnasco M: **Empirical characterization of the expression ratio noise structure in high-density oligonucleotide arrays.** *Genome Biol* 2002, **3:**research0018
14. Zhou Y, Abagyan R: **Match-Only Integral Distribution (MOID) Algorithm for high-density oligonucleotide array analysis.** *BMC Bioinformatics* 2002, **3:**3
15. Chu T-M, Weir B, Wolfinger R: **A systematic statistical linear modeling approach to oligonucleotide array experiments.** *Mathematical Biosciences* 2002, **176:**35-51
16. Chudin E, Walker R, Kosaka A, Wu SX, Rabert D, Chang T, Kreder DE: **Assessment of the relationship between signal intensities and transcript concentrations for Affymetrix GeneChip arrays.** *Genome Biol* 2001, **3:**research0005
17. Long AD, Mangalam HJ, Chan BYP, Tolleri L, Hatfield GW, Balsi P: **Improved statistical inference from DNA microarray data using analysis of variance and a Bayesian statistical framework.** *J Biol Chem* 2001, **276:**19937-19944