

Methodology article

Open Access

# Multifactorial experimental design and the transitivity of ratios with spotted DNA microarrays

Jeffrey P Townsend\*

Address: Department of Plant and Microbial Biology, 321 Koshland Hall, University of California, Berkeley, CA 94720, USA

Email: Jeffrey P Townsend\* - townsend@nature.berkeley.edu

\* Corresponding author

Published: 02 October 2003

BMC Genomics 2003, 4:41

Received: 21 June 2003

Accepted: 02 October 2003

This article is available from: <http://www.biomedcentral.com/1471-2164/4/41>

© 2003 Townsend; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

## Abstract

**Background:** Multifactorial experimental designs using DNA microarrays are becoming increasingly common, but the extent of the transitivity of cDNA microarray expression measurements across multiple samples has yet to be explored.

**Results:** A strong correlation between direct and transitive inference for significantly differentially expressed genes is demonstrated, using subsets of a dye-swap loop design.

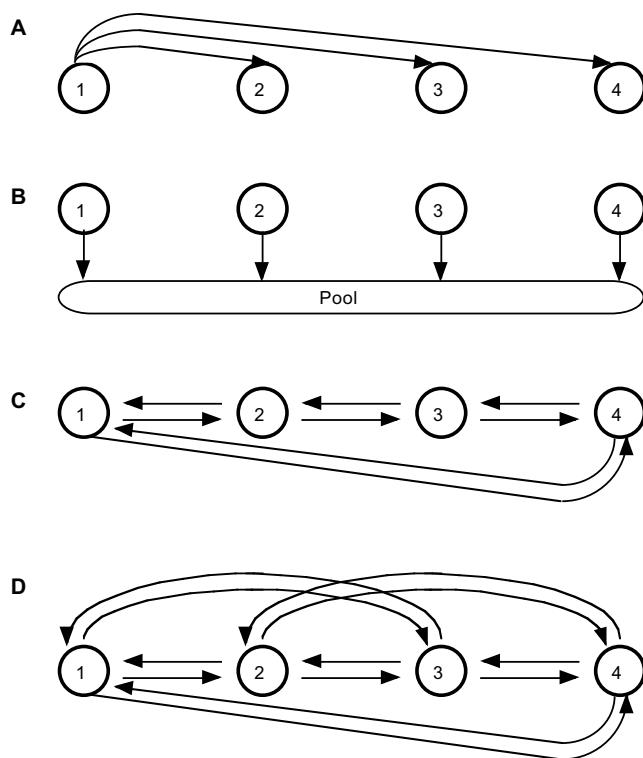
**Conclusions:** In experimental design, opportunities for transitive inference should be exploited, while always ensuring that comparisons of greatest interest comprise direct hybridizations.

## Background

As the use of microarrays to measure gene expression on a genomic scale becomes more common, biologists are designing increasingly complex experiments. Spotted DNA microarrays are used almost exclusively for relative measurements between two genotypes, environments, or developmental stages. Consequently, experimental design when only two states are of interest is well developed. A typical experiment might involve a comparison between a strain grown under an experimental treatment and the same strain grown under control conditions, or, alternatively, a comparison between a mutant and wildtype organism grown under identical conditions. Messenger RNA isolated from the two samples can be labeled with two different fluorophores and directly competitively hybridized against spotted DNA on a microarray. Replicating the comparison with fluorophores swapped controls for the effects of the dye [1-3]. Numerous methods have been proposed for the statistical analysis of these two-sample experiments (e.g. [4-7]). More complex experimental designs, in contrast, may comprise more

than two samples as characterized by their genotype, environment, or developmental stage.

In complex designs, the pairwise nature of spotted DNA microarray hybridizations permits multiple comparison structures. Most multiple-comparison spotted DNA microarray studies to date can be characterized as one of three types of experimental design. In the "reference sample" strategy depicted in Figure 1A, each sample is tested against a single, common standard (e.g. [8-10]). The advantage of this approach is the simplicity with which it is implemented. However, it has three disadvantages. First, because it pairs the reference strain with every other strain, it provides more information on the expression levels of the reference strain than on any other. Unless the reference strain is of primary interest, the focus in the experimental design upon its expression level is wasteful. Second, all comparisons of other strains in the design to each other are made in a purely transitive fashion, which is less precise and subject to greater experimental noise than direct comparison [11,12]. Third, use of a single

**Figure 1**

A depiction of common strategies for microarray experimental design. A) Each sample is tested against a single, common standard. B) A pooled sample, made up of equal parts of a number of samples, is used for comparison to the samples of interest. C) Each strain is compared head-to-head with other strains in a circular fashion, or, D) with added cross-comparisons, a multiple-pairwise fashion. This last example is a classical balanced block design, with dye swaps.

reference may lead to difficulty in estimating gene expression levels for genes that are at low expression level in the reference strain compared to other strains. The expression levels estimated for these genes will be particularly inaccurate because experimental error in the measurement of the sample with lower gene expression will lead to enormous variance in the ratios observed.

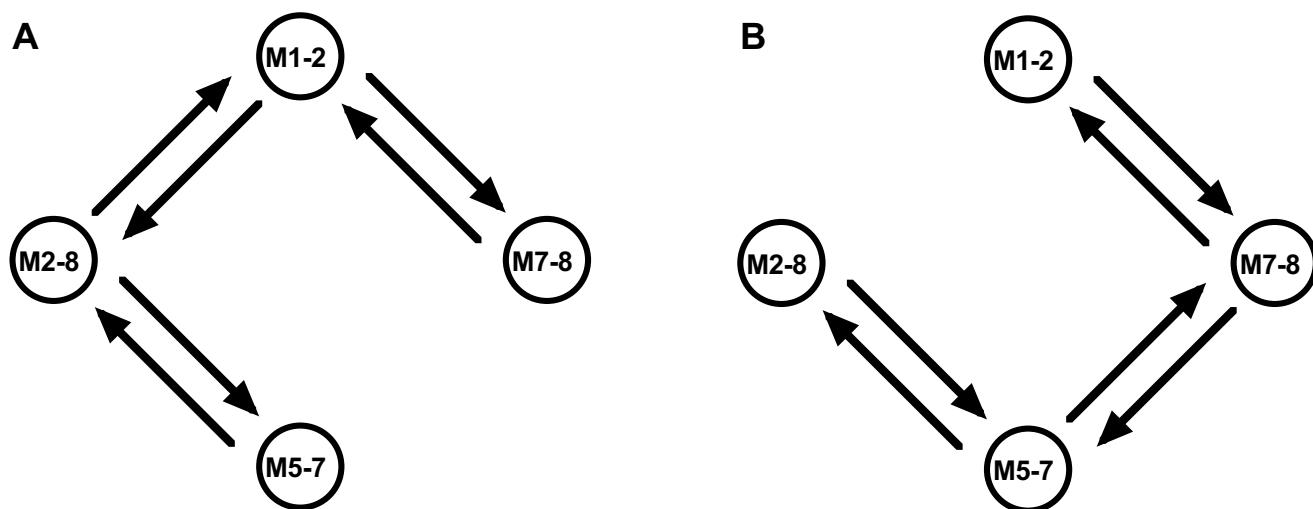
A second kind of design, the "pooled reference sample" strategy depicted in Figure 1B (e.g. [13]), is appealing in part because it solves this latter problem by making a cocktail of small amounts of mRNA from each of the samples to be compared. Thus, every gene expressed at a high level in any sample will be present at a reasonable concentration in the pooled sample. While pooling RNA samples from within populations of interest to form a set of pooled samples is statistically valid and efficient when comparing among populations [14], pooling a small

amount of each sample of interest to form a reference sample is still subject to the first and second disadvantages of the strategy depicted in Figure 1A. These disadvantages are that all comparisons are made directly to a reference that in this design is not of biological interest, and that all biologically interesting comparisons are composed of noisier transitive inference. Furthermore, employing a pooled sample as a reference may lead to difficulties in expanding or repeating an experiment if the pooling procedure itself cannot be easily or precisely replicated.

In the "loop" and all-pairwise strategies depicted in Figures 1C and 1D (e.g. [15–17]), each strain is compared directly with other strains, in a circular or multiple-pairwise fashion. The ability to detect differences is maximized, since the comparisons are between individual strains or conditions. The problems of using a common standard are avoided. A disadvantage of this method is that the way to interpret the raw data is not initially intuitive. Ratios observed across various pairwise comparisons are not immediately comparable. Moreover, the extent to which transitive data (whereby expression levels are inferred through a chain of comparisons to other samples) should be trusted, is as yet unexplored.

The challenge of converting ratio data across various pairwise comparisons into comprehensible expression levels has been solved by increasingly accessible classical ANOVA and Bayesian methods of analysis (e.g. [11,18,19]), and the complexity of the raw data should not discourage intrepid experimentalists from using these broadly powerful experimental designs. An intuitive and appealing outcome of all these methods is a relative measure of expression level for all samples. This relative measure is of an arbitrary unit. If desired, it can easily be scaled across genes so that one particular sample is the reference (*i.e.* has relative expression level of one), in which case the relative expression level estimates from replicated data may be interpreted exactly as would the raw data from a reference-sample experiment. Alternatively, expression levels for a gene may be scaled across samples examined so that the lowest of the expression levels is one. In some cases, expression levels may be scaled to counts of mRNA per cell by comparison to data sets from methods such as Serial Analysis of Gene Expression (SAGE) [19,20]. For most purposes, this scaling is of little relevance to the experimental question asked. Of much greater importance is the power to detect differences among the samples, and this depends on the power of the experimental design.

Powerful experimental design depends upon a thoughtful consideration of the statistical consequences of comparison structure [12]. Variation may be introduced into experimental procedure within the protocols of

**Figure 2**

"Dropout" design used to examine the transitivity of microarray results. A) Subset of comparisons from a full dye swap loop used to acquire direct results. B) Subset of comparisons from the full dye swap loop used to acquire indirect (transitive) results.

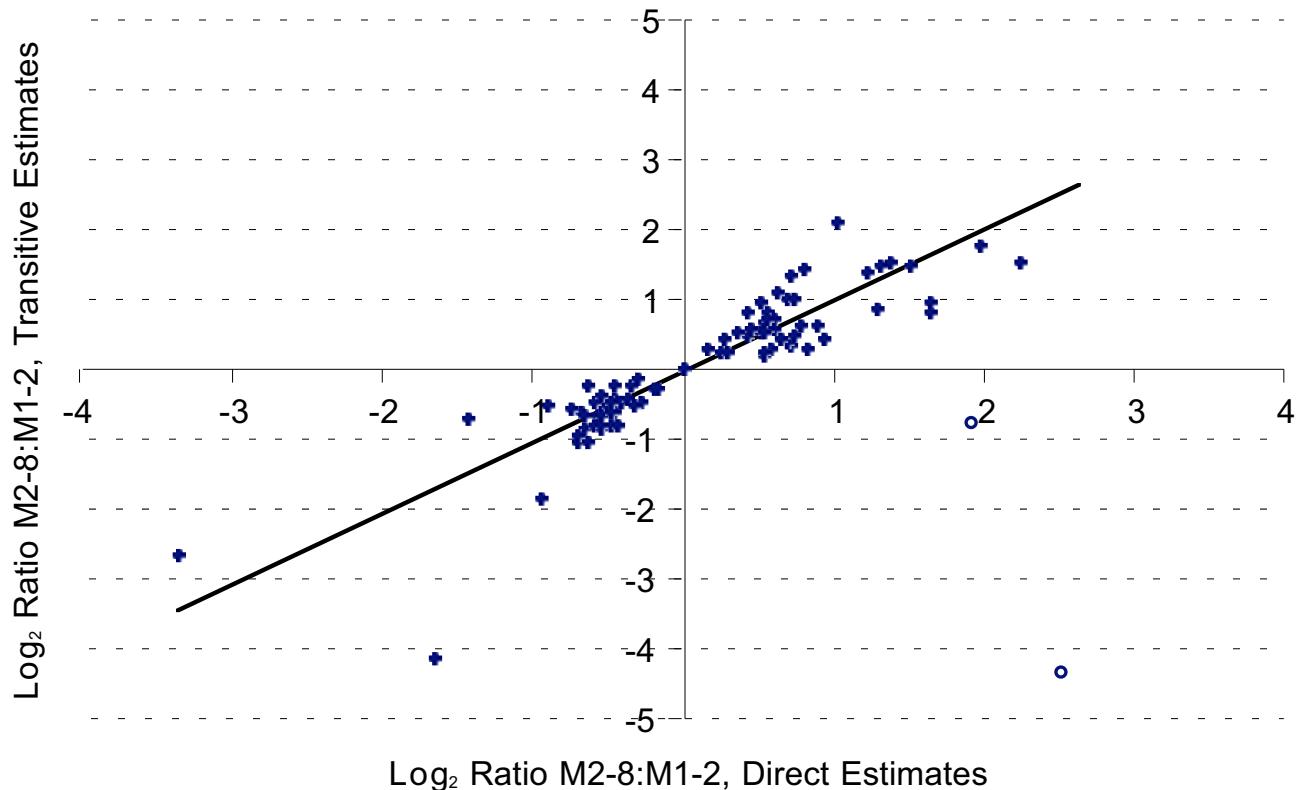
microarray production and implementation at multiple stages. Here the extent to which variation in parameters of the production and measurement process contribute to error variance of spotted DNA microarray measurements is analyzed. It is demonstrated that measured ratios, properly estimated, are transitive from sample to sample. Furthermore, it is demonstrated that transitive inference, a key component of all three types of complex experimental designs, is unbiased with regard to direct comparisons. Transitive inference is shown to be accurate provided replication is sufficient. Inference from direct comparisons, however, is more accurate, and should be maximized in any experimental design.

## Results and Discussion

All of the experimental designs in Figure 1 rely on transitive inference, in different ways, and each implicitly assumes that the ratio measurements from microarray experiments are not nonlinearly compressed or stretched compared to the true values. If compression of true ratios is nonexistent or linear, then the results of direct comparisons should be highly and linearly correlated with results arrived at indirectly from another (independent) subset of the same design (Figure 2). In the set containing the direct comparison (Figure 2A), ratios of gene expression of two samples are estimated from two direct hybridizations. In the set comparing the focal samples indirectly (Figure 2B), ratios of gene expression of the same two samples are esti-

mated by transitive inference. With the indirectly estimated ratio between two samples on the y-axis, and the directly estimated ratio on the x-axis, the slope of gene expression level estimates should have a highly significant linear regression. Its slope should be one, though increased variance may arise along the y-axis associated with transitive inference. This expectation is observed in Figure 3, which plots, for all statistically significantly differentially expressed genes, the Log<sub>2</sub> ratio of estimates of expression levels from the direct competitive hybridizations of strains M2-8 and M1-2 on the DNA microarray, against the Log<sub>2</sub> ratio of the purely transitive estimates of expression levels of M2-8 compared to M1-2. Results obtained from an indirect series of comparisons corroborate results obtained from direct comparisons.

Two outliers have not been included in the regression in Figure 3. These outliers serve to demonstrate the importance of replication and of the use of direct comparisons and closed loops in microarray experimental design. The slight outlier is the gene *OPT2*. As an outlier, it particularly demonstrates the importance of replication in preventing singular erroneous measurements from steering studies toward erroneous conclusions. Examination of the raw microarray data on the expression level of this gene reveals that the *OPT2* spot on one of the two microarray comparisons between M1-2 and M7-8 (Figure 2) had been excluded from analysis due to poor spot morphol-

**Figure 3**

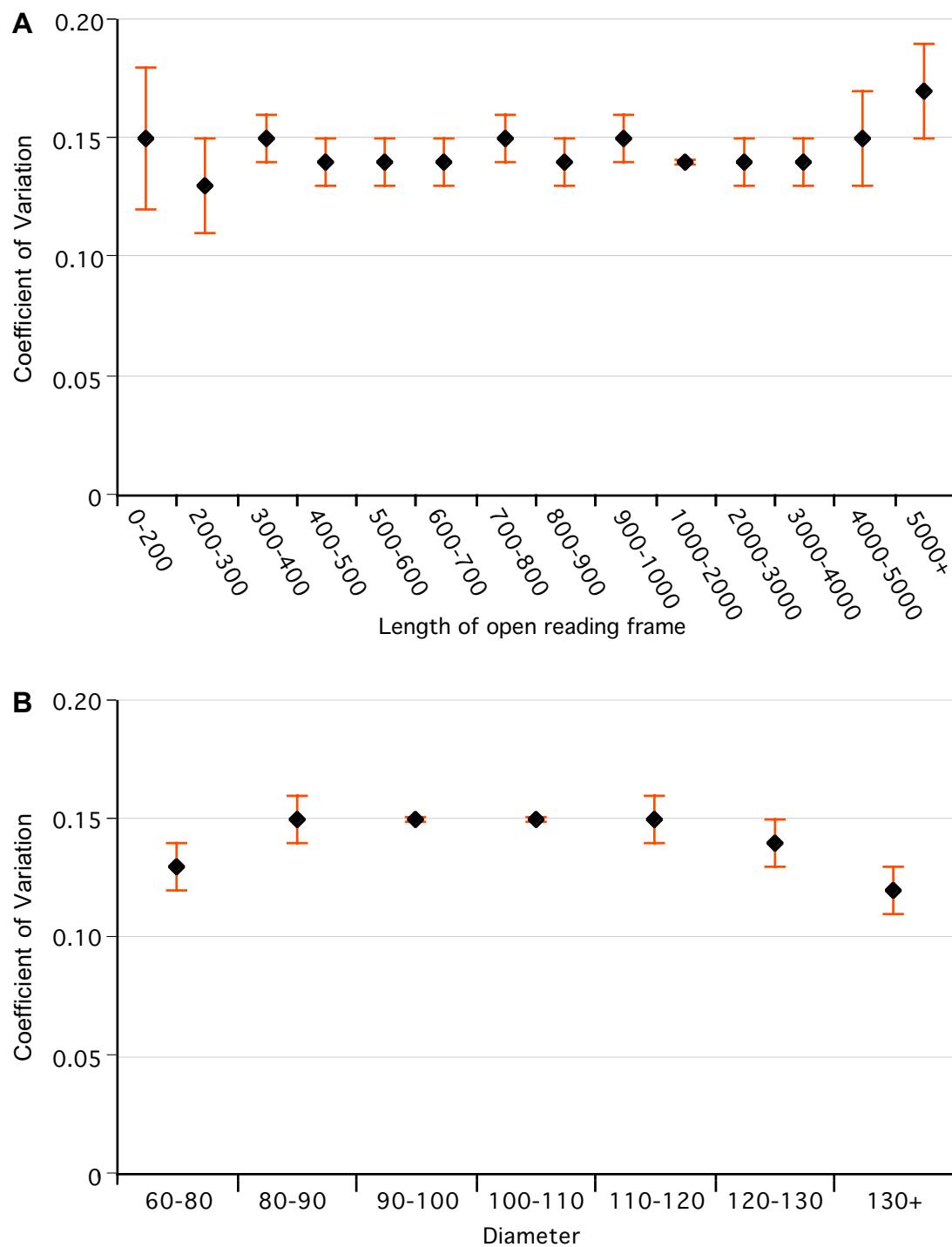
Correlation of direct and indirect estimates of gene expression level. Log-log scatter plot of estimates of gene expression level for the 87 genes with non-overlapping credible intervals, comparing estimates from direct comparisons (X-axis) to estimates from independent, transitive comparisons (Y-axis). Excluding the two outliers (open circles) in Cartesian quadrant IV, which represent estimates different not just in magnitude but also direction, the regression is  $y = 1.001 \times - 0.06$ , with a correlation  $r^2 = 0.8$  ( $P << 0.001$ ).

ogy and low signal to background. The level of *OPT2* mRNA in M1-2 in the "indirect" comparison was therefore based on a single microarray comparison. Whereas the ratio of M1-2:M7-8 from analysis of the full data set was 0.88, the single comparison yielded a ratio of 2.5. The result from the full data set [16] recapitulates the result of the direct comparison subset, calling attention to the value of replication in microarray experimental design.

The egregious outlier is *COS12*. By comparison with results obtained from the full dataset [16], *COS12* is incorrectly estimated by the transitive comparisons. This is perhaps unsurprising, as the estimates for *COS12* are that expression levels in both M2-8 and M1-2 (7.8 and 3.9, respectively) are higher than in the other two (1.0 and 2.4, respectively). Thus, in the direct comparison, the actual quantities of mRNA are relatively close. The comparison should be relatively accurate. In contrast, two out of three links of the transitive inference rely on ratios

observed with big differences in actual abundance between the two samples. In fact, further examination reveals that none of the raw measurements are incorrect in assessing in which sample has higher expression of *COS12*. The error derives entirely from the accumulated noise of transitivity, and it vanishes when the circle of comparisons (Figure 2B) is closed.

In order to identify the parameters of DNA microarray experiments that contribute to the experimental error variance, Figure 4 relates experimental error variance to characteristics of the reporter spot. One concern is that polymerase chain reaction (PCR) products spotted on slides may have been better or worse for subsequent hybridization depending on open reading frame (ORF) length. Spots might have better signal when the initial *Taq* polymerase amplification protocol used for all ORFs was optimal for their length, or when GIBCO Elongase was used on longer, reamplified genes. Figure 4A shows that

**Figure 4**

Length of ORF, average spot diameter, and the coefficient of variation of gene expression level. The coefficient of variation estimated for a given gene does not depend upon A) the ORF length, or B) the average spot diameter. Error bars are +/- 2 standard errors of the mean. Larger error bars at either extreme are due to smaller sample size in those size classes and do not reflect a greater standard deviation of the measurements.

there is no particular effect of length of the ORF on the error variance of the reporter spot. The large standard error bars at the extremes are a consequence of the few ORFs annotated below 200 base pairs or above 4000 base pairs. This result confirms that the PCR and hybridizations performed were equally successful across many different ORF lengths.

When microarrays are printed, variation in the size of reporter spots can arise from variation in tip deposition, DNA and salt concentrations, and rehydration time. A concern is that small or large spot size, even when passing visual inspection and intensity thresholds, may be anti-correlated with consistency of microarray ratio results. Figure 4B shows the effect of average spot diameter (across microarrays) on experimental error variance. Across the considerable range of average spot diameters, average error variance remains virtually constant. There is little difference in consistency of observed ratios across a wide range of spot sizes.

Figure 5 relates absolute expression level (from SAGE data [20]) to the coefficient of variation, the average absolute log ratio, and the intensity of microarray spots. One might be concerned that genes with low gene expression level yield especially noisy data. However, the coefficient of variation for measurements of relative gene expression level does not vary across absolute gene expression levels (Figure 5A).

One might also be concerned that DNA microarray measurements of meagerly expressed genes would have a different degree of "compression" than measurements of genes expressed abundantly, so that, on average, the log ratios of one or the other would be inflated. Figure 5B demonstrates that there is no such inflation; the average absolute log ratio is consistent across absolute gene expression levels. Very abundantly expressed genes had higher intensity on DNA microarray chips (Figure 5C). However, it should be noted that nearly 50% of the genes lie in the SAGE bin labeled "zero" in Figure 5. Most of the rest lie in the bin labeled "one". Thus, the variation in intensity with absolute expression level is only present at a significant scale for a minor fraction of genes in the genome, and genes within even the "zero" class are frequently well-measured (Figure 5A).

## Conclusions

Microarrays can yield insight into both qualitative differential expression and quantitative ratios of gene expression in respective samples, provided that measurements are replicated and appropriately analyzed. These results are of importance to users of any method for gene expression analysis that yields quantitative estimates of gene expression with a powerful and replicated interconnected

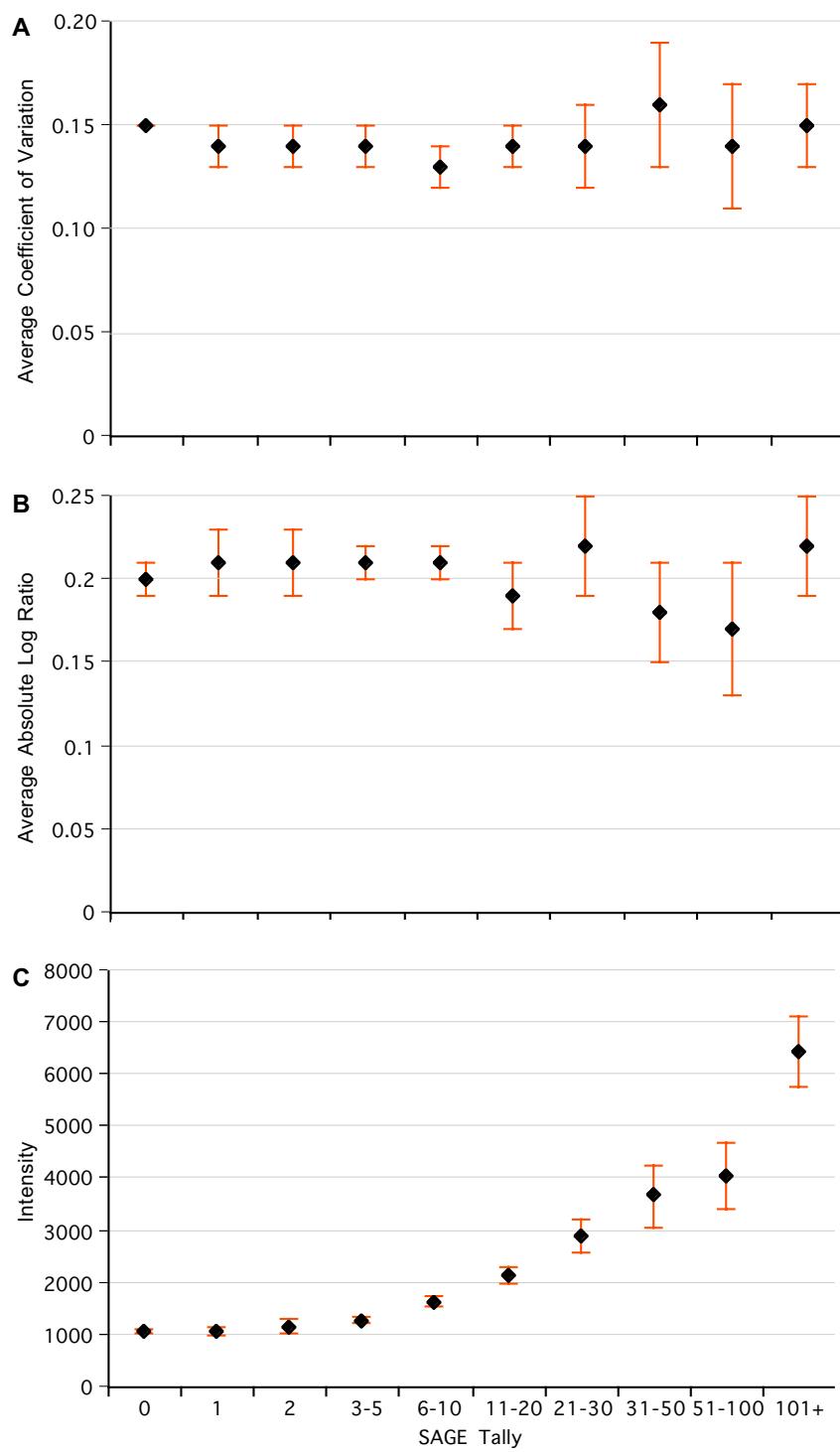
experimental design. Studies on cDNA microarray experimental design have consistently pointed out that reference-sample design (Figure 1A, e.g. [8,9,21,22]) yields weak statistical information [12,18,19,23] about the non-reference samples. Many of these studies appropriately advocate classical balanced block or hierarchical designs, which are statistically powerful. It is, however, possible to construct multiple balanced block or hierarchical designs for any experiment. For all designs, it remains advisable to maximize direct comparisons relative to transitive inference. As complex interconnected experiments comprise more samples, most conceivable designs will not put every possible pair of samples on an array. In these cases, the power to detect differences is maximized by including direct comparisons that ensure a minimal transitive path between any two samples. For example, these direct comparisons might cross a circle of samples or cross between treatments at the same time point in a time course. Not only do these cross-comparisons contribute to the power of an experiment by substituting direct for indirect inference, but also they increase the number of paths of transitive inference, improving the quality of transitive inferences. Although a straight reference-sample design (Figure 1A) never involves more than two arrays of distance, it also never gains from multiple paths of transitive inference.

These microarray results affirm the success of PCR products used as reporter spots on microarrays in producing equally valid data across all reasonably common ORF lengths. They also indicate that average spot diameter has little effect on the consistency of reporters. Most importantly, it is demonstrated that a current protocol, normalization, and statistical method yield transitive results that are correlated in direction and magnitude with direct results.

## Methods

Microarrays containing 6218 open reading frames from the Saccharomyces Genome Project were prepared and used as in [24] and [16], resulting in high experimental reliability, strong signal, and low background. A few spots from these arrays were excluded from analysis when both fluorescence signals were within two standard deviations of the distribution of the intensities of the background pixels of that spot, or if printed spot morphology was poor. Samples were natural isolates of wine yeast (*S. cerevisiae*) from Montalcino, Italy, so that quantitative variation in gene expression among isolates may exist at many loci and at a variety of levels [16].

Subsets of the data (Figure 2, Additional file 1, Additional file 2) from a loop design (Figure 1C) were comparatively analyzed using a Bayesian analysis of gene expression levels [19] with a common coefficient of variation among

**Figure 5**

Effect of SAGE mRNA count on coefficient of variation, average absolute log ratio, and signal intensity. SAGE yields data on the absolute expression level, in counts of mRNA per cell, of genes. A) The coefficient of variation estimated for a given gene does not generally appear to depend upon the absolute expression level. B) The average absolute log ratio does not generally appear to depend upon the absolute expression level. C) Intensity of a microarray spot does depend upon absolute expression level, at least for highly abundant genes. Error bars are  $\pm 2$  standard errors of the mean. Larger error bars at higher SAGE levels are due to smaller numbers of genes at those levels and do not reflect a greater standard deviation of the measurements.

samples (Additional file 3, Additional file 4). The Bayesian analysis of gene expression levels uses all transitive and direct comparisons to yield estimates and 95% credible intervals for the level of gene expression in each sample examined, relative to the sample with lowest expression. The 95% credible intervals are the Bayesian equivalent of 95% confidence intervals, representing the central 95% of the range of values that are credible for the parameter estimated. Gene expression levels were called significantly different when 95% credible intervals did not overlap between a pair of samples, indicating that no single level of gene expression was credible for both of a pair of samples. This procedure is conservative compared to using a P value for differential expression of 0.05.

The 87 genes that were significant by this criterion were then used in further analysis. Genes that were not statistically significantly different were excluded because these genes have minimal or no differences in gene expression level among samples, and therefore the estimates of their expression level are primarily influenced by experimental noise. Estimates of gene expression level that are not statistically significant should not be considered credibly different, and are generally discarded in any functional study. Thus, only the estimates of expression level for genes that show statistically significant differences in expression were informative with regard to the transitivity of microarray ratio data. If all genes are included in the regression in Figure 3, the estimated expression levels via direct or indirect comparisons are only barely correlated ( $y = 0.16 \times - 0.08$ ,  $r^2 = 0.07$ ). This level of correlation occurs because the vast majority of the points in the regression constitute mostly noise, and only 87 constitute genuine signal.

In the Bayesian method, the results are reported relative to gene expression in the sample with lowest expression. That sample is assigned an expression level of one, and other samples are scaled appropriately. This has the intuitive appeal that all gene expression level measurements are positive. The inferred expression levels are of arbitrary scale, rather than in absolute counts of mRNA abundance per cell; this is a consequence the inherently comparative nature of the two-color spotted microarray technology and is unavoidable in any analysis of an experiment conducted with any of the experimental designs in Figure 1.

Potential sources of error variance were ruled out by examination of the degree of error variance across categories of ORFs using the full dataset [16] rather than the subsets in Figure 2. For instance, ideal data would show no change in error variance across deposited ORF length, average spot diameter, or ORF absolute expression level as measured by SAGE [25].

## Additional material

### Additional file 1

[Click here for file](#)

[<http://www.biomedcentral.com/content/supplementary/1471-2164-4-41-S1.txt>]

### Additional file 2

[Click here for file](#)

[<http://www.biomedcentral.com/content/supplementary/1471-2164-4-41-S2.txt>]

### Additional file 3

[Click here for file](#)

[<http://www.biomedcentral.com/content/supplementary/1471-2164-4-41-S3.txt>]

### Additional file 4

[Click here for file](#)

[<http://www.biomedcentral.com/content/supplementary/1471-2164-4-41-S4.txt>]

## Acknowledgements

Thanks to Rachel Whitaker, Alison Galvani, and three anonymous reviewers for comments on the manuscript.

## References

- Kerr M, Kathleen and Churchill Gary A.: **Statistical design and the analysis of gene expression microarray data.** *Genetical Research* 2001, **77**:123-128.
- Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G and Gibson G: **The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*.** *Nature Genetics* 2001, **29**:389-395.
- Liang Mingyu, Briggs Amy G., Rute Elizabeth, Greene Andrew S. and Cowley Allen W.: **Quantitative assessment of the importance of dye switching and biological replication in cDNA microarray studies.** *Physiological Genomics* 2003, **14**:199-207.
- Ideker T, Thorsson V, Siegel AF and Hood LE: **Testing for differentially-expressed genes by maximum-likelihood analysis of microarray data.** *JOURNAL OF COMPUTATIONAL BIOLOGY* 2000, **7**:805-817.
- Baldi Pierre and Long Anthony: **A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes.** *Bioinformatics* 2001, **17**:509-519.
- Theilhaber Joachim, Bushnell Steven, Jackson Amanda and Fuchs Rainer: **Bayesian estimation of fold-changes in the analysis of gene expression: the PFOLD algorithm.** *Journal of Computational Biology* 2001, **8**:585-614.
- Tseng George C., Oh Min-Kyu, Rohlin Lars, Liao James C. and Wong Wing Hung: **Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects.** *Nucleic Acids Research* 2001, **29**:2549-2557.

8. DeRisi JL, Iyer VR and Brown PO: **Exploring the metabolic and genetic control of gene expression on a genomic scale.** *Science* 1997, **278**:680-686.
9. Gasch Audry P., Spellman Paul T., Kao Camillo M., Carmel-Harel Orna,, Eisen Michael B., Storz Gisela, Botstein David and Brown Patrick O.: **Genomic expression programs in the response of yeast cells to environmental changes.** *Molecular Biology of the Cell* 2000, **11**:4241-4257.
10. White Kevin P., Rifkin Scott A., Hurban Patrick and Hogness David: **Microarray analysis of Drosophila development during metamorphosis.** *Science* 1999, **286**:2179-2184.
11. Kerr M. Kathleen, Martin Mitchell and Churchill Gary A.: **Analysis of variance for gene expression microarray data.** *Journal of Computational Biology* 2000, **7**:819-837.
12. Yang Yee Hwa and Speed Terry: **Design issues for cDNA microarray experiments.** *Nature Reviews* 2002, **3**:579-588.
13. Arbeitman Michelle N., Furlong Eileen E. M., Imam Farhad, Johnson Eric, Null Brian H., Baker Bruce S., Krasnow Mark A., Scott Matthew P., Davis Ronald W. and White Kevin P.: **Gene expression during the life cycle of Drosophila melanogaster.** *Science* 2002, **297**:2270-2275.
14. Peng Xuejun, Wood Constance, Blalock Eric M., Chen Kuey Chu, Landfield Philip W. and Stromberg Arnold J.: **Statistical implications of pooling RNA samples for microarray experiments.** *BMC Bioinformatics* 2003, **4**:26.
15. Oleksiak MF, Churchill Gary A. and Crawford DL: **Variation in gene expression within and among natural populations.** *Nature Genetics* 2002, **32**:261-266.
16. Townsend Jeffrey P., Cavalieri Duccio and Hartl Daniel L.: **Population genetic variation in genome-wide gene expression.** *Molecular Biology and Evolution* 2003, **20**:955-963.
17. Ranz JM, Castillo-Davis CI, Meiklejohn Colin D. and Hartl Daniel L.: **Sex-dependent gene expression and evolution of the Drosophila transcriptome.** *Science* 2003, **300**:1742-1745.
18. Wolfinger Russell D., Gibson Greg, Wolfinger Elizabeth D., Bennett Lee, Hamadeh Hisham, Bushel Pierre, Afshari Cynthia and Paules Richard S.: **Assessing gene significance from cDNA microarray expression data via mixed models.** *Journal of Computational Biology* 2001, **8**:625-637.
19. Townsend Jeffrey P. and Hartl Daniel L.: **Bayesian analysis of gene expression levels: statistical quantification of relative mRNA level across multiple treatments or samples.** *Genome Biology* 2002, **3**:research0071.1-71.16.
20. Velculescu Victor E., Zhang Lin, Zhou Wei, Vogelstein Jacob, Basrai Munira, Bassett Douglas E., Hieter Phil, Vogelstein Bert and Kinzler Kenneth W.: **Characterization of the yeast transcriptome.** *Cell* 1997, **88**:243-251.
21. Shamji Alykhan, Kuruvilla Finny G. and Schreiber Stuart: **Partitioning the transcriptional program induced by rapamycin among the effectors of the Tor proteins.** *Current Biology* 2000, **10**:1574-1580.
22. Furlong Eileen E. M., Andersen Erik C., Null Brian, White Kevin P. and Scott Matthew P.: **Patterns of gene expression during drosophila mesoderm development.** *Science* 2001, **293**.
23. Kerr M. Kathleen and Churchill Gary A.: **Experimental design for gene expression microarrays.** *Biostatistics* 2001, **2**:183-201.
24. Eisen MB and Brown PO: **DNA arrays for analysis of gene expression.** *Methods Enzymol.* 1999, **303**:179-205.
25. Velculescu Victor E., Zhang Lin, Vogelstein Bert and Kinzler Kenneth: **Serial analysis of gene expression.** *Science* 1995, **270**:484-487.

Publish with **BioMed Central** and every scientist can read your work free of charge

*"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."*

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

