

Methodology article

Open Access

## Comparing transformation methods for DNA microarray data

Helene H Thygesen\* and Aeilko H Zwinderman

Address: Clinical Epidemiology and Biostatistics, Academisch Medisch Centrum, University of Amsterdam, Meibergdreef 9, 1100 DD Amsterdam, Netherlands

Email: Helene H Thygesen\* - h.h.thygesen@amc.uva.nl; Aeilko H Zwinderman - a.h.zwinderman@amc.uva.nl

\* Corresponding author

Published: 17 June 2004

Received: 02 February 2004

BMC Bioinformatics 2004, 5:77 doi:10.1186/1471-2105-5-77

Accepted: 17 June 2004

This article is available from: <http://www.biomedcentral.com/1471-2105/5/77>

© 2004 Thygesen and Zwinderman; 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:** When DNA microarray data are used for gene clustering, genotype/phenotype correlation studies, or tissue classification the signal intensities are usually transformed and normalized in several steps in order to improve comparability and signal/noise ratio. These steps may include subtraction of an estimated background signal, subtracting the reference signal, smoothing (to account for nonlinear measurement effects), and more. Different authors use different approaches, and it is generally not clear to users which method they should prefer.

**Results:** We used the ratio between biological variance and measurement variance (which is an F-like statistic) as a quality measure for transformation methods, and we demonstrate a method for maximizing that variance ratio on real data. We explore a number of transformations issues, including Box-Cox transformation, baseline shift, partial subtraction of the log-reference signal and smoothing. It appears that the optimal choice of parameters for the transformation methods depends on the data. Further, the behavior of the variance ratio, under the null hypothesis of zero biological variance, appears to depend on the choice of parameters.

**Conclusions:** The use of replicates in microarray experiments is important. Adjustment for the null-hypothesis behavior of the variance ratio is critical to the selection of transformation method.

### Background

When DNA microarray data are used for gene clustering or tissue classification, the signal intensities are usually transformed and normalized in several steps in order to improve comparability and signal/noise ratio. These steps may include subtraction of an estimated background signal, logarithmic transform, subtraction of the reference signal, and more. Different authors use different approaches: the most prominent controversies include Lowess smoothing, heteroscedastic models, subtraction of the reference signal and background correction. The aim of our study was to devise a method to find the optimal transformation method for a given dataset [1,2].

As for Lowess smoothing, the idea is that the non-linear trend, often found in the log (reference)/log (variety)-scatter plot, is an artifact. Pritchard [3], who produced the data we used, expressed concern that this may not always be the case, and called for a method to distinguish artificial non-linearity from biological effects. Also, at the CAMDA '02 Conference [4], concern was expressed that if the non-linear effect is weak, the benefit from smoothing may not be worth its costs in term of possible overfitting.

As for heteroscedastic models, the idea is that the heteroscedasticity often found in the log (reference)/log (variety)-scatter plot (the phenomena that the variance of the difference between the variety signal and the reference signal

depends on the signal intensity), is an artifact. It follows that unbiased normalization would produce homoscedastic data. Some authors, including Rocke et al. [5], Huber et al. [6] and Cui et al. [7], suggest post-normalization homoscedasticity (termed variance stability) as a success criterion for normalization methods. This is also the basis of the heteroscedastic Lowess smoothing, adopted by SNOMAD [8]. However, unbiased normalization may not be synonymous with successful transformation from a practical point of view. To illustrate this, suppose that the true response ( $\log(\text{ratio})$ ) is  $N(0, 1)$ -distributed, whereas the noise is  $N(0, 1)$  for low-intensity genes and zero for high-intensity genes. A variance-stabilizing normalization would, then, shrink the expressions of the low-intensity genes with a factor  $\sqrt{2}$ , but for many applications, a lower shrinking factor (e.g., zero) would work better. (The National Institute of Aging online array analysis tool [9] solves this problem by estimating the technical variance locally). Actually, Brown et al. [10] showed how one can improve the discriminatory power of microarray data by explicitly reducing the weights of measurements suspected of being of low quality (shrinking). Although assigning explicit weights to measurements is uncommon, some methods do so implicitly. For example, using a linear scale instead of the log scale effectively shrinks response levels of low-expressed genes. On the other hand, Huber et al. [6] found consistency between variance stabilization and reproducibility of post-transformation response.

Wernisch et al. [11] used the Pearson correlation coefficient for two replicates of the same experiment. Although that is probably a good success criterion in many cases, it cannot help us decide how to reduce the variety signal and the reference signal to one number, since the reference signal will typically show equally strong correlation as the variety signal. But the reference signal does not in itself tell us anything about the tissue sample.

He et al. [2] aimed at a success criterion more closely related to the application in question. They made the distinction between Type 1 applications (aimed at obtaining p-values for the response of individual genes in parallel) and Type 2 applications (aimed at correlating multi-gene patterns to phenotypes). As for Type 1, they classified each gene as either differentially expressed or non-differentially expressed (based on some threshold for the p-value), and used the accuracy of the classification as a success criterion for the transformation. They proposed several methods for establishing the true set of differentially expressed genes.

Our choice of success criterion is similar to that of He [2], but instead of dichotomous judgements of differentiation, we use quantitative responses. The reasons for this choice are the following:

- One application that we have in mind is the assessment of the normal variability of the expression of either individual genes or genes that are representative for some group of genes, e.g. genes involved in particular pathways (this was the purpose of Project Normal [3], for which the data we used were produced).
- Even if the study is aimed at dichotomous judgements, the information loss from dichotomization may make the assessment of success less sensitive.
- Using the quantitative response levels is simpler, because one does not need to decide on a threshold. Actually, Huber et al. [4], who used a success criterion similar to that of He [2], found that the performance of the Lowess smoother compared to the rank transform in some cases depends on the threshold.

Further, instead of comparing the post-transformation response estimate to true differentiation levels (which are generally unavailable), we measured post-transformation reproducibility in a replicated experiment. When more replicates of the same tissue sample are available, an obvious success criterion is a low within-sample variance. Consider an experiment in which  $S$  tissue samples are each hybridized independently on  $R$  arrays (replicates) with a probeset of  $G$  genes.  $Y_{gsr}$  denotes some indicator (transformed signal) of the expression of gene  $g$  in replicate  $r$  of sample  $s$ . As a measure of reproducibility of the transformed signal, we use the ratio of between-tissue-sample variance to the within-sample variance (the F-statistic), where the latter is averaged over the genes for stability:

$$\frac{\sum_g \text{VAR}_s \left( \frac{1}{n_{repls}} \sum_r Y_{gsr} \right)}{\sum_g \frac{1}{n_{samples}} \sum_s \text{VAR}_r(Y_{gsr})} \tag{1}$$

**P-value for the variance ratio**

Part of the observed variance ratio can be explained as an artifact of statistical fluctuations. Assume, for simplicity, that  $Y_{gsr}$  were  $N(0, \sigma)$ -distributed (i.i.d.) for all  $g, s, r$ . Then, the expectation of the between-sample variance would become

$$\text{VAR}_{between} \approx n_{genes} \sigma^2 \tag{2}$$

whereas the expectation of the within-sample variance would become

$$\text{VAR}_{within} \approx n_{genes} n_{repls} \sigma^2 \tag{3}$$

Since the number of genes is large,  $VAR_{within}$  becomes reasonably stable, so the expected variance ratio would become

$$\frac{VAR_{between}}{VAR_{within}} \approx \frac{1}{n_{repls}} \quad (4)$$

However, a simple permutation experiment (see results) shows that this estimate is, in many cases, very inaccurate. Presumably, when the post-normalization response is not normal distributed (or if the expressions of individual genes are correlated and/or have heterogenous variance), the above approximation fails. What is worse, it appears that for several transformation methods, the expected variance ratio varies with parameter values, as does its stability. Clearly, one should not choose the parameter that yields the highest variance ratio, if that high variance ratio is solely an artifact of statistical fluctuations. These considerations lead us to two new success criteria:

1. The p-value for the variance ratio under the null hypothesis of zero biological variance.
2. The difference between the obtained variance ratio and the expected variance ratio (under the same null hypothesis).

In this paper, we show how those success criteria depend on parameter values for a number of transformation methods.

## Results

### Data

The study was based on the second data set for the CAMDA 2002 contest ("Project Normal") by Pritchard et al. [3]. In Project Normal, 18 tissue samples from 6 mice (from the testis, kidney and liver from each mouse) were hybridized independently on four identical microarrays. A mixture of all 18 samples was used as a reference. For each spot, the image analysis performed by Pritchard (using GenePix software [12]) provided four values (background and foreground for the variety and the reference). "Flagged" spots (related to absent genes or distorted parts of the image) were omitted. The background signal was ignored. Finally, all signal values were normalized with the median for the channel in question. We did not account for spatial effects, crosstalk or dye effects.

### Box-Cox transformation

In the first experiment, we applied the Box-Cox transform to both signals before subtracting the reference signal from the variety signal. Figure 1 shows the variance ratios for each of the three data sets as a function of  $\lambda$ .

The results suggested an optimal  $\lambda$ -value of -1 for the testis data, 1 for kidney and 10 for liver. These results were surprising, in particular the high variance ratios in the testis data set with negative  $\lambda$ -values, but the results became more plausible when compared to the expected variance ratios under the null hypothesis of zero biological variance.

For a number of lambda values, we computed the variance ratio of 500 transformed and randomly permuted copies of the testis dataset. The 2.5 and 97.5 percentiles of the 500 variance ratios were used as confidence limits under the null hypothesis that gene expressions did not depend on the tissue sample. (See figure 2).

We noticed that the variance of the permutation results became high for high  $\lambda$  values. This is not surprising: for high  $\lambda$  values, the dominance of a few, strongly expressed genes, becomes extreme, leading to less stable averages. We also noticed that the downward slope for negative  $\lambda$  values could be attributed to statistical fluctuations. Measured by p-value,  $\lambda = 0.25$  was the new optimum, but the differences within the range from -1 to 0.5 were hardly significant.

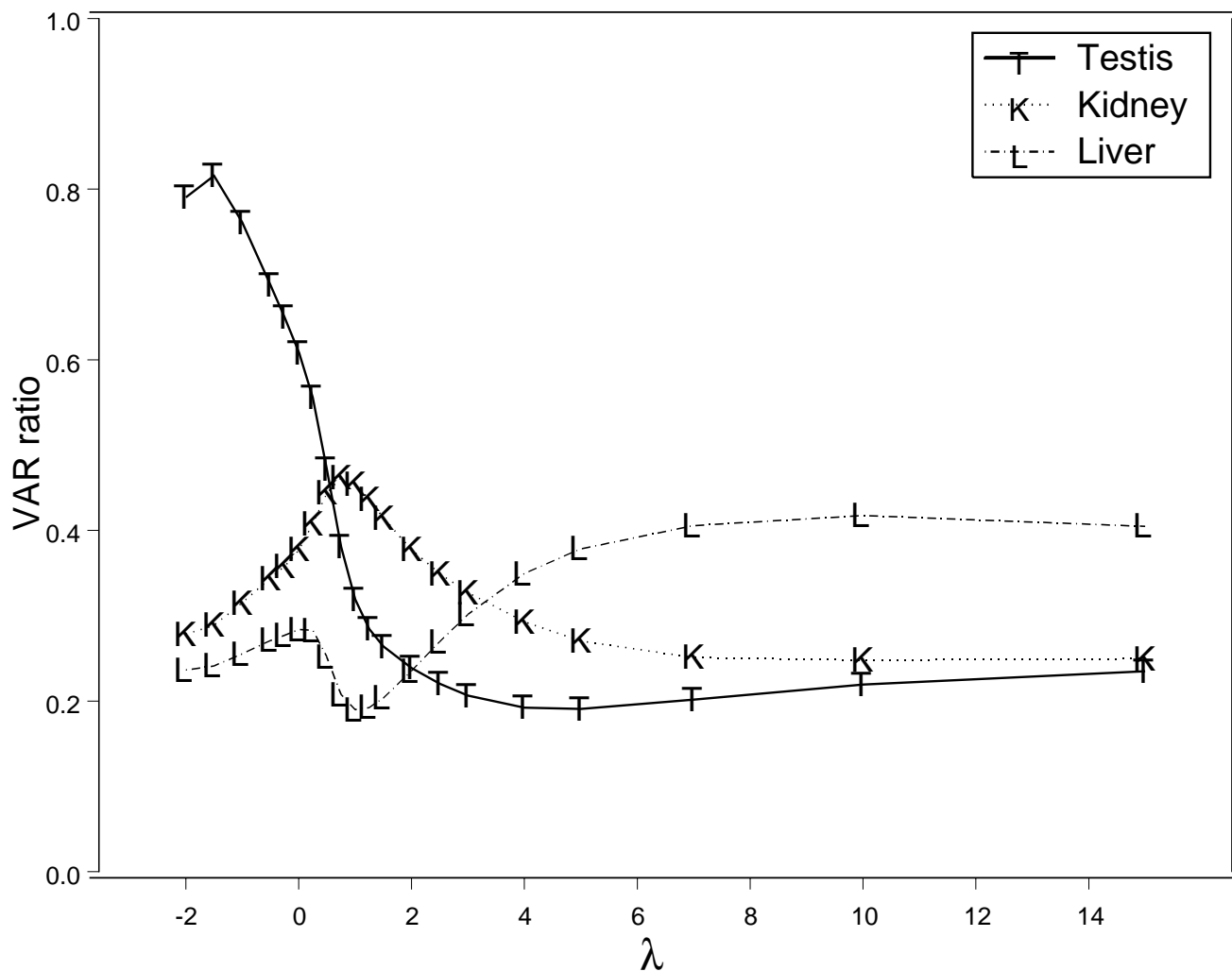
### Partial subtraction of log(reference signal)

As can be seen in figure 3, the kidney data actually improved when the reference signal was ignored. It is not implausible that the effect of subtracting the reference signal could differ between tissue types: the reference could correlate stronger with one tissue than with another. More disturbingly, the variance ratio reaches its maximum for a subtraction factor of 2 for the liver data. A subtraction factor outside the [0, 1]-range does not have any meaningful interpretation. This finding suggests a problem with the reproducibility of the reference signal.

Because the quality of the reference signal is probably a property of the array rather than a biological effect, it may be too rigid to choose an over-all subtraction factor for all 24 arrays for one tissue. In fact, for the second replicate of the testis of mouse 1, which we omitted because of bad image quality, ignoring the reference signal turned out to work better. As for partial subtraction of log(reference signal), our findings did not change when we corrected for the behavior of the variance ratio under the null hypothesis of zero biological variance.

### Baseline shift

We computed variance ratios for the three data sets after shifting all measurements upwards by a number of medians (or averages) for the channel, and subsequently taking the logarithm. For the testis data, a baseline relative to the mean worked better than one relative to the median. For the kidney, the reverse was true. We also tried a baseline



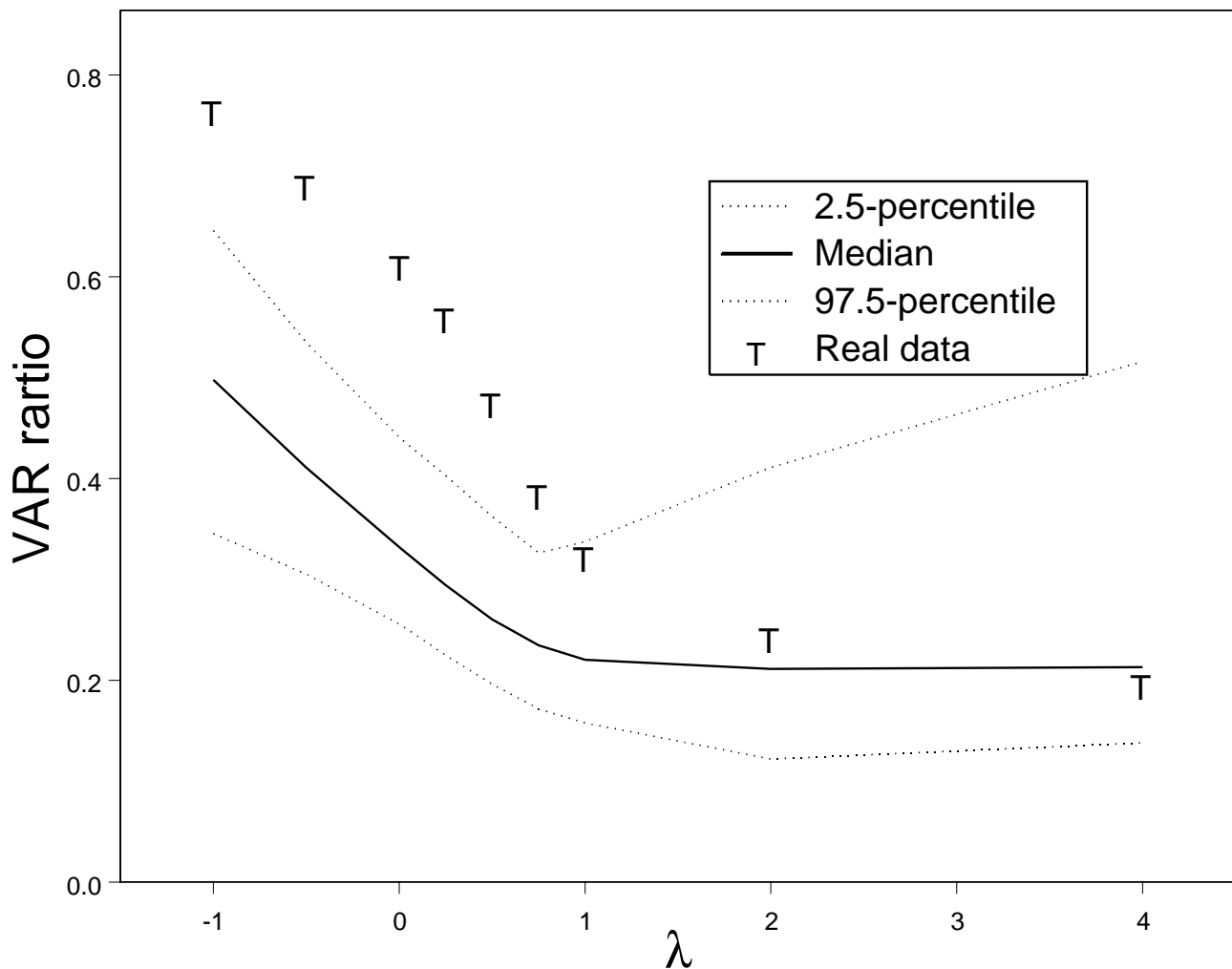
**Figure 1**  
**Box-Cox transform.** Variance ratios for each of the three datasets, under Box-Cox transformation, for different values of  $\lambda$ .

relative to the median (or average) background. The idea was that the baseline shift may work by reducing the influence from spots with a foreground signal not much stronger than the background. But a baseline relative to the foreground appeared to work much better.

However, the permutation test suggested that the good performance of the baseline shift could be an artifact of statistical fluctuations. Figure 4 compares the variance ratio under different baseline shift levels to the median and the 97,5-percentile from the distribution under the null hypothesis of zero biological variance.

**Smoothing**

The kidney data appeared to improve the most with a smooth band of 0.75, which is also the default in S-Plus. The testis data did not improve through smoothing. We also tried to smooth log(variety) as a function of log (reference) (that is, without rotating the scatter plots), but that resulted in slightly lower variance ratios. First-order local regression (Lowess) generally worked slightly better than second-order local regression (Loess). Also experiments that accounted for heteroscedasticity, by normalizing the residuals relative to the local standard deviation of the residuals, yielded disappointing results. That being



**Figure 2**

**Box-Cox relative to null hypothesis.** Variance ratios for the testis dataset, under Box-Cox transformation, for different values of  $\lambda$ . Compared to the expected variance ratio under the null hypothesis of zero biological variance, and its 2.5% and 97.5% confidence limits (percentiles of the variance ratio with permutated data).

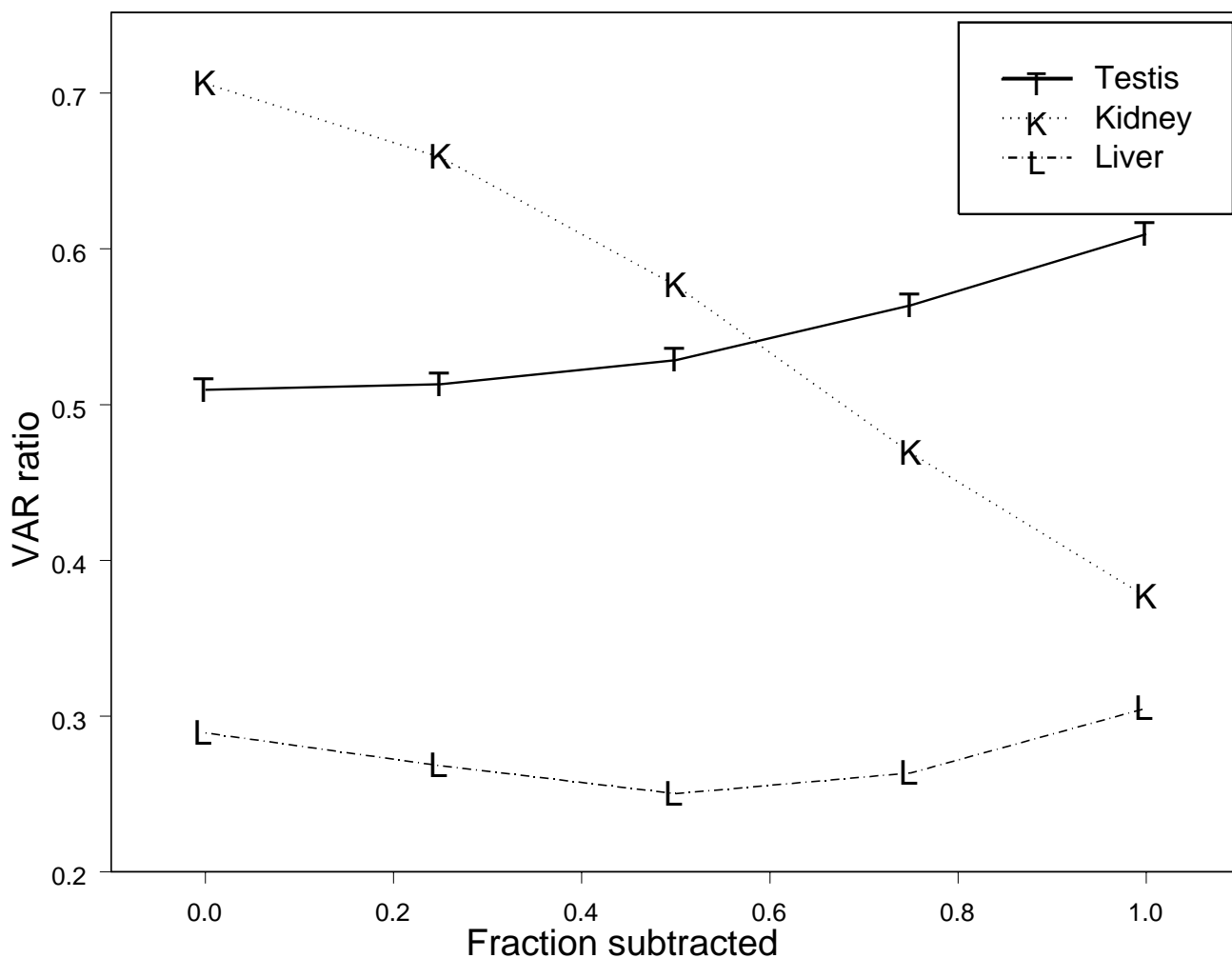
said, the difference in variance ratio between the various smoothing methods was very small.

It is not clear why this difference between the testis data and the kidney data is observed. In general, the scatter plots for the testis data look nicer than those for the kidney data, suggesting that smoothing is not necessary for the testis data and only introduces an overfit which reduces the variance ratio. On the other hand, two of the mice had highly elevated expression levels of nearly all genes. Although the expression levels were divided by the global median, some tissue-sample-specific nonlinearity in those two samples may exist, e.g. because of a saturation effect. This would justify smoothing. However,

smoothing would lead to a lower variance ratio because a tissue-sample specific artifact can not be distinguished from a biological effect. This problem illustrates that independent replicates, where four different tissue samples from each mouse were used, would have been better. For the Lowess smoothing, adjusting for the behavior of the variance ratio under the null hypothesis did not change our findings.

#### **Background correction**

The background signal of a spot, estimated by the GenePix software, uses the median signal of all pixels in the neighbourhood of the spot. We investigated the effect of subtracting this background from the signals before



**Figure 3**  
**Partial subtraction of log(reference signal).** Variance ratios for each of the three datasets, under partial subtraction of log(reference signal), for different values of  $\alpha$ .

transformation, which was either the rank transform, rank(variety)-rank(reference), or log (variety)/log (reference). In the latter case, signals below some threshold were set equal to that threshold, and the procedure was repeated with different thresholds. We also tried using a kernel smoother, in which the local background estimates from Genepix were smoothed with an exponential kernel, restricted to the 12 nearest neighbour spots. In all cases, background correction slightly reduced the variance ratio.

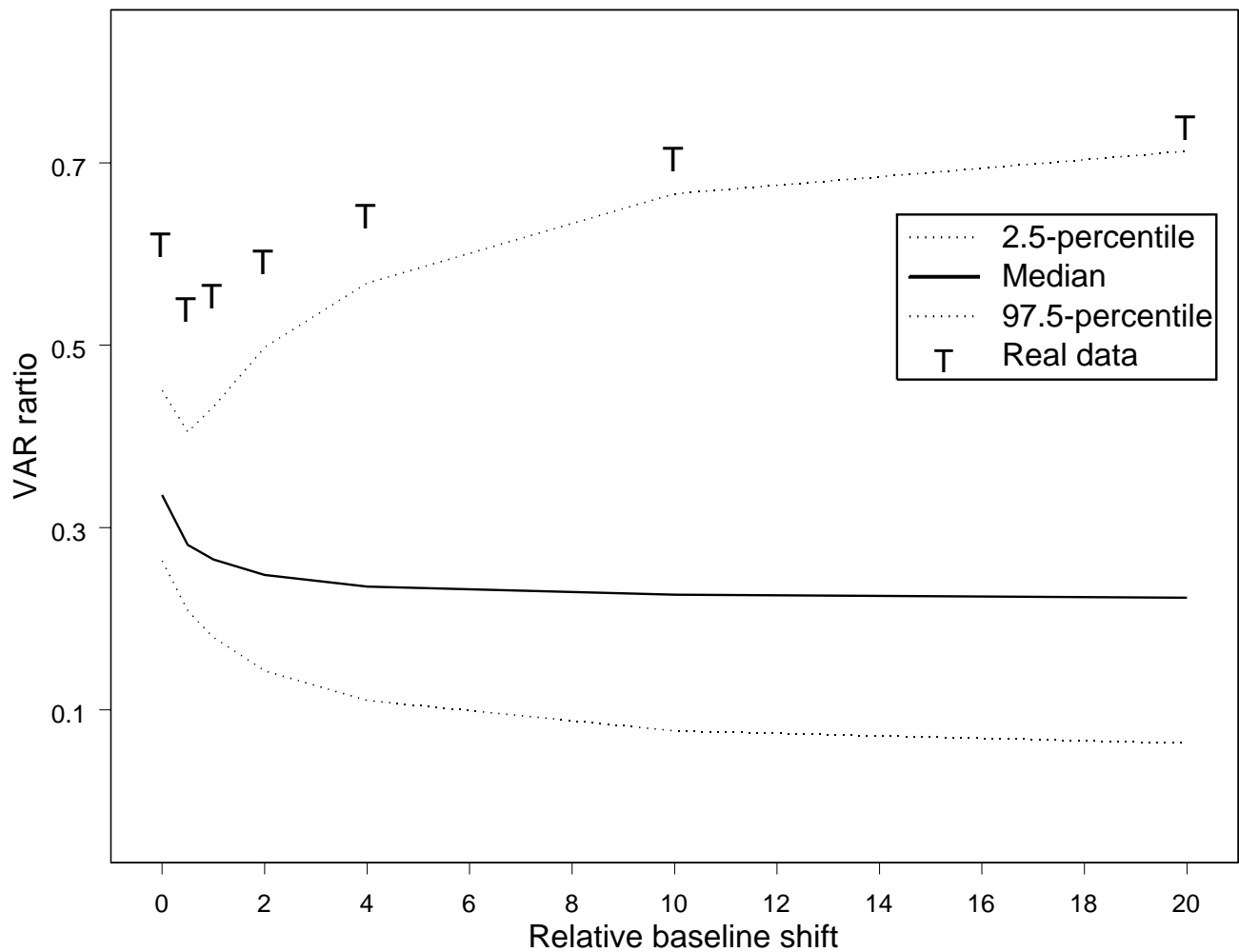
**Reproducibility of our findings**

In a study such as this, attention should be paid to possible overfitting, in particular in the presence of outliers. The variance ratio is not a robust method, in principle, but since the variances are averages over 5776 genes, an out-

lier must be rather extreme if it is to skew the results significantly. Replacing the variance ratio with the slightly more robust measure

$$\frac{\text{Median}_g(\text{VAR}_s(\frac{1}{n_{\text{repls}_{gs}}} \sum_r Y_{gsr}))}{\sum_g \frac{1}{n_{\text{samples}}} \sum_s \text{VAR}_r(Y_{gsr})} \tag{5}$$

yielded almost identical results. For the Box-Cox transformation (figure 5) and the baseline shift (figure 6), we nevertheless divided the genes randomly into ten subsets and repeated the calculations for the testis data set for each subset. In general, the trend is the same for the subsets as



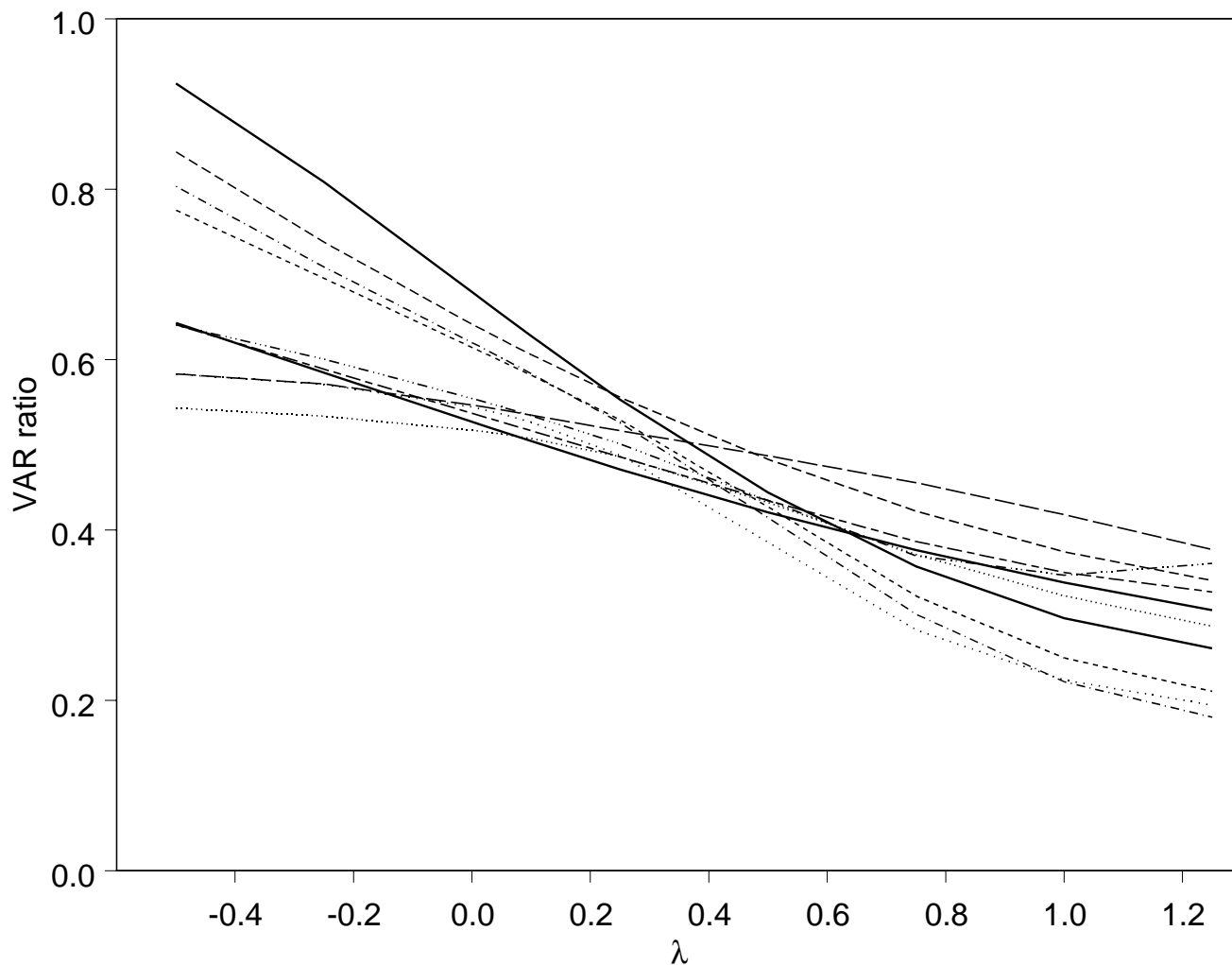
**Figure 4**  
**Baseline shift.** Variance ratios for the testis dataset, under logarithmic transform with a baseline shift relative to the mean, for different values of  $a$ . Compared to the expected variance ratio under the null hypothesis of zero biological variance, and its 2.5% and 97.5% confidence limits (percentiles of the variance ratio with permuted data).

for the whole dataset, although the overall level of the variance ratio, as well as the optimal parameter value, varied somewhat between the subsets.

**Discussion**

The optimal transformation method will differ from application to application. Therefore, experiments should be designed in a way that makes it possible to identify the optimal transformation method. Replicating each array, as in the Project Normal datasets, is an excellent approach. Replicating each gene within each array would enable the same analysis.

The variance ratio is, as any performance measure would be, a stochastic entity. Therefore, its absolute value may be less meaningful than its p-value under the hypothesis of zero biological variance. Under certain assumptions (normal distribution, homogene variance, zero correlation between genes), the p-value of the variance ratio can be computed from the F distribution, but those assumptions were, in general, not met in this study. One can resort to a random permutation experiment in order to obtain that p-value. This step is important, because optimizing the p-value may lead to different parameter values than optimizing the crude variance ratio.



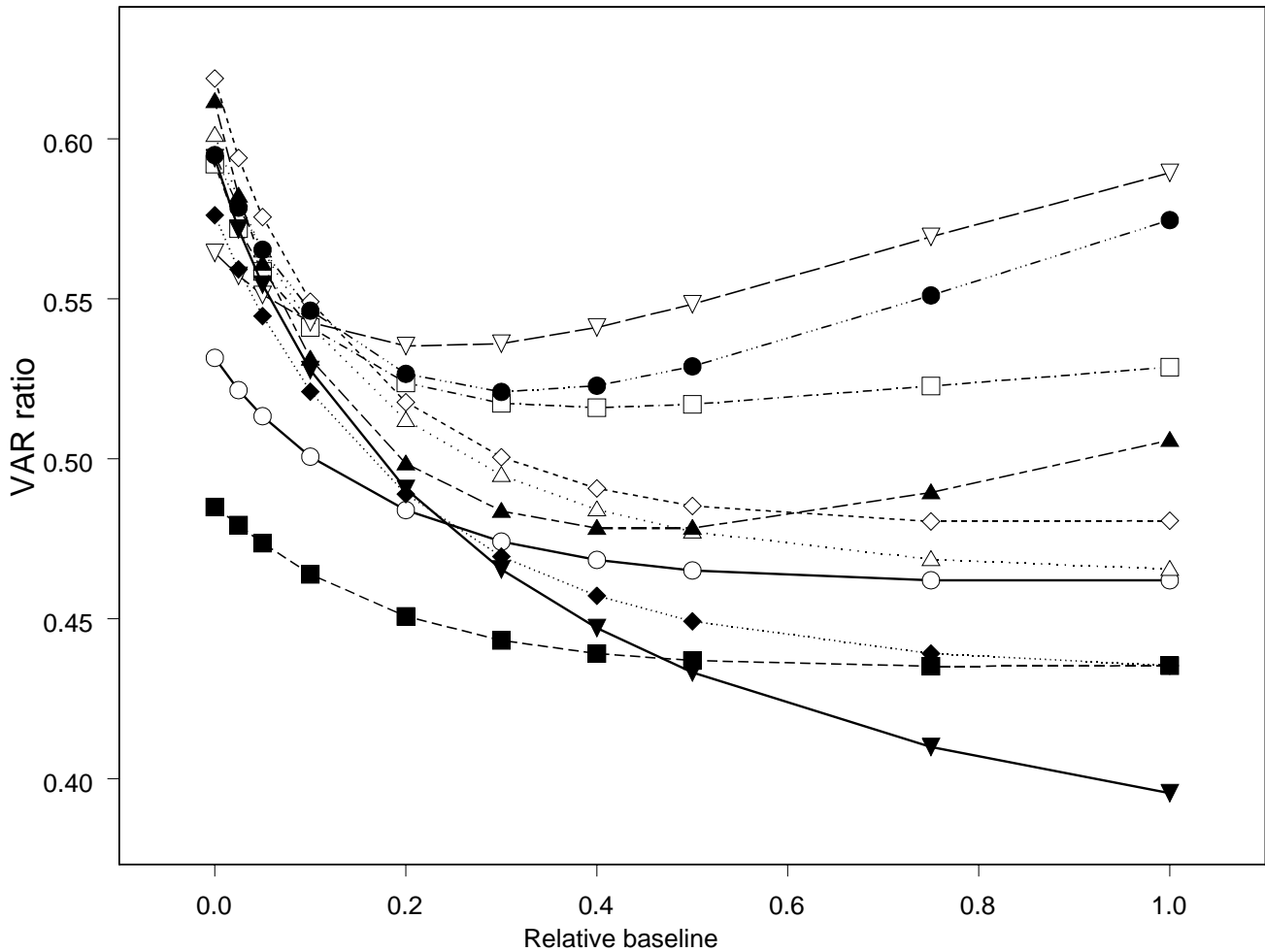
**Figure 5**  
**Box-Cox for subsets of the genes.** Variance ratio for the Box-Cox transform as a function of  $\lambda$ , for 10 complementary subsets of the genes. Testis data.

**Conclusions**

Figure 7 compares the highest possible variance ratio, relative to the distribution under the null hypothesis of zero biological variance, for each of the three data sets, for three transformation methods. The variance ratios were standardized so that the median variance ratio from the permutation experiment corresponded to a value of zero, and the upper quartile to a value of one. The difference between the variance ratio found and the median variance ratio from the permutation experiments is divided by the difference between the median and the upper quartile of the latter. We would have preferred to compute the p-value of the variance ratio, but this was not feasible

because in many cases none of the permutation experiments resulted in variance ratios above the variance ratio for the unpermuted data. For the testis and liver data, a parameter of 0.25 in the Box-Cox transform gave the best results, but none of the methods performed substantially better than the simple log-ratio. For the kidney data, smoothing, ignoring the reference signal, or a Box-Cox parameter of 0.75 all gave better results. It is possible that a combination of those three methods would be better still. This being said, we limited our study to optimizing the parameters globally for all arrays within a dataset. Simultaneous optimization of the parameters for individual arrays, if feasible, is likely to yield better results.





**Figure 6**  
**Baseline shift for subsets of the genes.** Variance ratio for the baseline-shift transform as a function of the baseline relative to the mean, for 10 complementary subsets of the genes. Testis data.

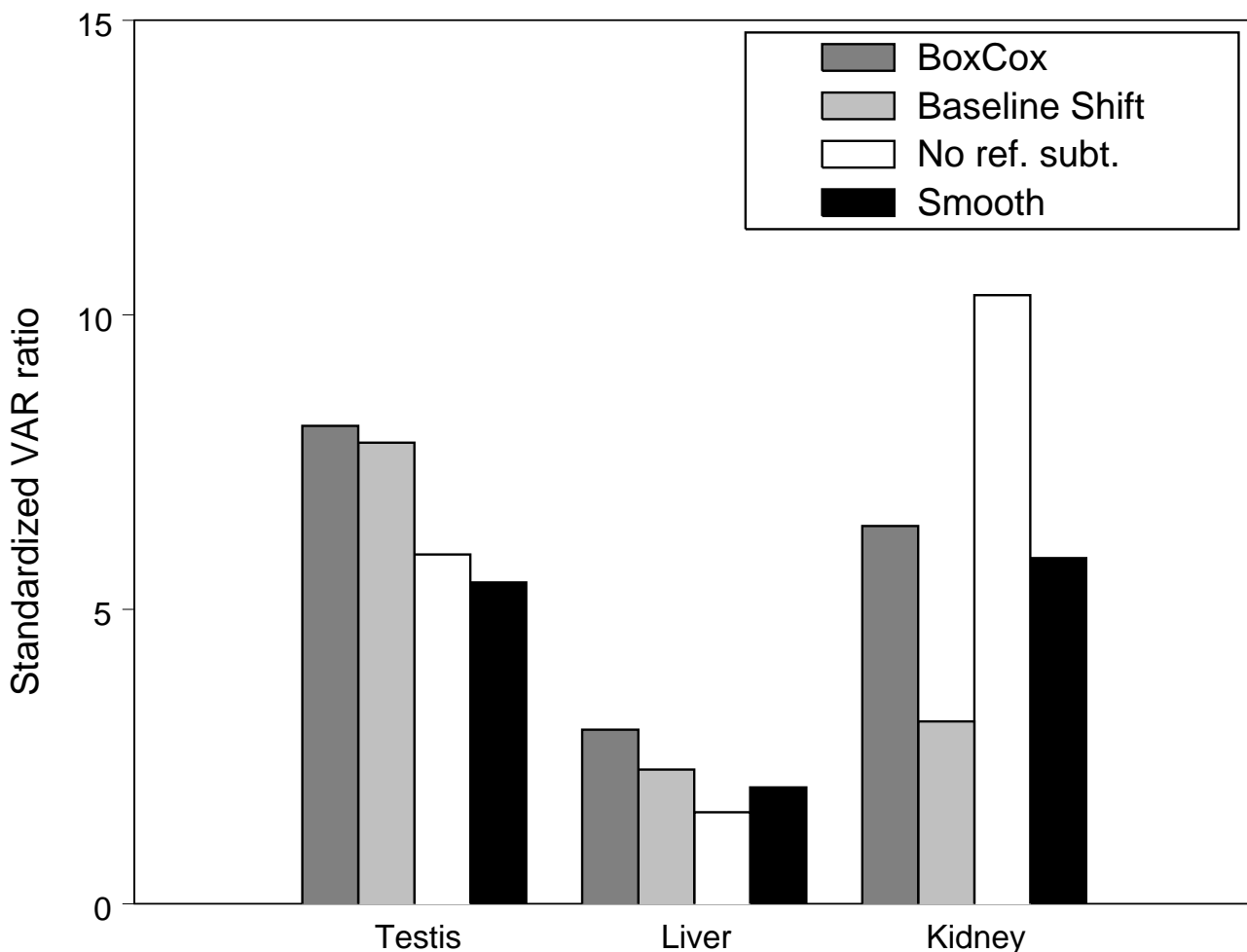
Background correction, based on the median of neighborhood pixel intensities, did not prove to be beneficial. It is possible that other image processing techniques could produce more useful background estimates.

As compared to the Standard Figures of Merit proposed by He et al. [2], our method is simpler, but He's method is more closely related to applications that rely on dichotomous judgements of differentiation of gene expressions. It will be interesting to see how the two approaches relate to each other if applied to the same data sets. Such a comparison may also resolve the question of whether the per-gene response should be dichotomized or not.

**Methods**

**General approach**

For a number of different transformation methods, we computed the variance ratio for a number of parameter values, leading to a choice of parameter value that maximizes the variance ratio. This was done separately for each of three data sets (kidney, liver, testis), leading to three different optimal parameter values. Further, for a number of transformation methods, we computed the distribution of the variance ratio under the assumption of zero between-sample variance. The optimal parameter choice was then the one leading to the highest variance ratio, or the lowest p-value, relative to the distribution of the variance ratio under the null hypothesis.



**Figure 7**

**Summary of results.** Maximal variance ratios for each of the three data sets, under four different transformation methods. The median variance ratio under the null hypothesis of zero biological variance was subtracted, and the residue was divided by the difference between the median and the upper quartile of the latter.

**Box-Cox transform**

The common practice of taking the logarithm of the measured intensities can be justified by the assumption that most variance sources are multiplicative rather than absolute. Under that assumption, the logarithm of the intensity has desirable properties such as an approximately Gaussian distribution and homoscedastic residuals, i.e. all genes have equal influence on the model fitting, whether they give rise to high or low intensities. The question of whether to use a log scale or a linear one was raised by, among others, Efron et al. [13], who came to the conclusion that the log scale is better than the linear scale. If some variance sources (such as background noise) are

absolute, one could use e.g., the square root as a compromise between a linear and a logarithmic scale, as in e.g., Amaratunga and Cabrera [14]. The generalization of this is the Box-Cox transform

$$B_{\lambda}(x) = \begin{cases} \log(x) & \text{if } \lambda = 0 \\ \frac{x^{\lambda} - 1}{\lambda} & \text{if } \lambda \neq 0 \end{cases} \quad (6)$$

We used the following transform:

$$Y = B_{\lambda}(x_{variety}) - B_{\lambda}(x_{reference}) \quad (7)$$

for different values of  $\lambda$ , thereby finding the  $\lambda$ -value that led to the highest variance ratio. A consequence of this is that for a  $\lambda$ -value of 1 (linear scale), the reference signal is subtracted from the variety signal rather than divided by. We have chosen this deliberately. Alternatively, one could use

$$Y = B_{\lambda}(x_{\text{variety}}/x_{\text{reference}}).$$

#### **Partial subtraction of log(reference signal)**

It is sometimes argued, for example by Wolfinger [15], that one should ignore the reference signal. Although this partly depends on what biological question is asked (whether absolute gene expressions, or gene expressions relative to a given reference, are more meaningful), we will here find the degree to which the reference signal should be subtracted in order to maximize the variance ratio. The transformed signal becomes  $Y = \log(x_{\text{variety}}) - a \log(x_{\text{reference}})$  where  $a$  is a parameter. We have limited the analysis to the log transform, but the same analysis could easily be carried out with alternative methods, e.g., Box-Cox.

#### **Adding a baseline signal**

When we restricted the analysis to genes in certain intensity ranges, it appeared that the log transform worked better for high-intensity genes, and the linear scale better for low-intensity genes. This makes sense because the background signal (which is an additive variance source) is relatively more influential for low-intensity genes. So instead of the "static" BoxCox, which assumes the optimal  $\lambda$  to be the same for all genes, a "dynamic" BoxCox, with a  $\lambda$ -value approaching 1 for low-intensity genes and 0 for high-intensity genes, may work better. A simple way to mimic this is to add some constant to all gene expressions before taking the logarithm. This additive constant, or baseline shift, could be set proportional to either the median or the average of the gene expressions for one channel – that is, we choose a different baseline shift for each channel. (Actually, because we normalized the data by setting the median of each channel to one, the baseline shift becomes the same for all channels if we choose a shift proportional to the median of the channel). The transformation becomes

$$Y = \log(a x_{0,\text{variety}} + x_{\text{variety}}) - \log(a x_{0,\text{reference}} + x_{\text{reference}}) \quad (8)$$

where  $x_0$  is either the average or the median expression over the genes, and  $a$  is a parameter. Other possible compromises between the linear scale and the logarithmic scale are the so-called Generalized Logarithm [5], and the arsinh transform [6].

#### **Smoothing**

The well-known banana shape, sometimes seen in  $\log(\text{variety})/\log(\text{reference})$ -scatter plots, is generally attributed to non-linear measurement effects. We try to neutralize such effects by applying the following trans-

form:  $Y = \log\left(\frac{x_{\text{variety}}}{x_{\text{reference}}}\right) - f(\log(x_{\text{reference}} * x_{\text{variety}}))$  where

$f$  is some local regression fit to the variety signal as a function of the reference signal. We used first-order Lowess regression and tried different values for the bandwidth parameter [16].

#### **Background correction**

It can be debated whether background correction actually improves data quality. The reason why we have chosen not to correct for background is that many of the transformation methods that we studied cannot deal with negative numbers, which raises the question what to do with those spots that have higher background than foreground. A simple transformation, which does not have problems with negative values, is the rank. Thus, we applied the following transformations:

$$Y = \text{rank}(x_{\text{variety}}) - a \text{rank}(x_{\text{reference}})$$

where  $a$  is a parameter. The variance ratios for these transformations were computed both with and without background correction. We also tried with the log transform, where signal values below a fixed threshold were set equal to that threshold. The arsinh transform suggested by Huber [6] is another solution to the problem with the negative signal intensities.

#### **P-value for the variance ratio**

As explained in section 1 (Background), the raw variance ratio may not be comparable for different transformation methods and different parameter values. Therefore, we computed the p-value of that variance ratio relative to the null hypothesis that the expression of a gene depends only on the gene, and not on the tissue sample. The distribution of the variance ratio under that null hypothesis is computed with a simple permutation method, similar to the method described by Efron et al. [13]. The same permutation test is used by the National Institute of Aging [9], but they apply it locally to account for heteroscedasticity.

Consider the following design matrix for the unpermuted variance ratio (6 tissue samples (=mice), 4 replicates of each):

1 2 3 4 5 6

1 2 3 4 5 6

1 2 3 4 5 6

1 2 3 4 5 6

This means that the first replicate of the first tissue sample is substituted for the first replicate of the first sample, and so on.

In each permuted calculation, the rows were randomly permuted, e.g.,

1 4 3 6 2 5

4 1 2 6 3 5

2 5 6 3 1 4

4 5 1 3 2 6

Here, replicate 2 of the virtual tissue sample 1 refers to replicate 2 of physical tissue sample 4, and so on. Because we always substituted with a replicate with the same dye, dye effects are retained. On the basis of such a permutation, the variance ratio is computed. Repeating this with 500 different permutations yields 500 sample variance ratios. Finally, we compared the observed (unpermuted) variance ratio to the quantiles from the 500 samples.

### Authors' contributions

Both authors were involved in the definition of the variance ratio estimator. HT defined the transformation methods under study. HT did the literature review. Both authors contributed to the manuscript.

### References

1. Tsodikov A, Szabo A, Jones D: **Adjustments and measures of differential expression for microarray data.** *Bioinformatics* 2002, **18**:251-260.
2. He YD, Dai H, Schadt EE, Cavet G, Edwards SW, Stepaniants SB, Duenwald S, Kleinhanz R, Jones AR, Shoemaker DD, Stoughton RB: **Microarray Standard Data Set and Figures of Merit for Comparing Data Processing Methods and Experiment Designs.** *Bioinformatics* 2003, **19**(8):956-965.
3. Pritchard CC, Hsu L, Delrow J, Nelson PS: **Project normal: Defining normal variance in mouse gene expression.** *PNAS* 2001, **98**(23):13266-71.
4. **CAMDA '02 conference** [<http://www.camda.duke.edu>]
5. Rocke DM, Durbin B: **Approximate variance-stabilizing transformations for gene-expression microarray data.** *Bioinformatics* 2003, **22**(19(8)):966-72.
6. Huber W, Von Heydebreck A, Sultmann H, Poustka A, Vingron M: **Variance stabilization applied to microarray data calibration and to the quantification of differential expression.** *Bioinformatics* 2002, **18**(Suppl 1):S96-S104.
7. Cui X, Kerr MK, Churchill GA: **Transformations for cDNA Microarray Data.** *Statistical Applications in Genetics and Molecular Biology* 2003, **1**(Article 4):2.
8. Colantuoni C, Henry G, Zeger S, Pevsner J: **SNOMAD (Standardization and Normalization of MicroArray Data): Web-accessible Gene Expression Data Analysis.** *Bioinformatics* 2002, **18**(11):1540-1 [<http://pevsnerlab.kennedykrieger.org>].
9. National Institute of Aging: **Online Microarray Data Analysis Tool.** [<http://lgsun.grc.nia.nih.gov/ANOVA>].
10. Brown CS, Goodwin PC, Sorger PK: **Image metrics in the statistical analysis of DNA microarray data.** *PNAS* 2001, **98**(14):8944-49.
11. Wernisch L, Kendall SL, Soneji S, Wietzorrek A, Parish T, Hinds J, Butcher PD, Stoker NG: **Analysis of whole-genome microarray replicates using mixed models.** *Bioinformatics* 2003, **19**(1):53-61.
12. Axon: **GenePix Pro microarray scanning and analysis software.** [[http://www.axon.com/GN\\_GenePixSoftware.html](http://www.axon.com/GN_GenePixSoftware.html)].
13. Efron B, Tibshirani R, Storey JD, Tusher V: **Empirical Bayes Analysis of a Microarray Experiment.** *JASA* 2001, **96**(456):1151-1160.
14. Amaratunga D, Cabrera J: **Analysis of Data From Viral DNA Microchips.** *JASA* 2001, **96**(456):1161-1170.
15. Wolfinger R, Chu T: **Simultaneous Assessment of Transcriptional Variability and Tissue Effects in Normal Mouse.** *Proceedings of CAMDA '02 2003* [<http://www.camda.duke.edu>].
16. Venables WN, Ripley BD: **Modern Applied Statistics with S-Plus.** Springer-Verlag; 1994:250. ISBN 3-540-94350

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)

