Tutorial section

There is no silver bullet – a guide to low-level data transforms and normalisation methods for microarray data

Abstract

To overcome random experimental variation, even for simple screens, data from multiple microarrays have to be combined. There are, however, systematic differences between arrays, and any bias remaining after experimental measures to ensure consistency needs to be controlled for. It is often difficult to make the right choice of data transformation and normalisation methods to achieve this end. In this tutorial paper we review the problem and a selection of solutions, explaining the basic principles behind normalisation procedures and providing guidance for their application.

Keywords: *microarrays, experimental bias, data normalisation, low-level data-transforms, microarray data analysis*

INTRODUCTION

Microarray technology provides a powerful tool for large-scale gene expression measurements. Typical application areas include genomics research, quantitative systems biology and the more traditional screening for candidate genes. Microarray data are the end-product of a fairly intricate experimental process,[1–3] and adequate analysis must meet the challenge of separating true biological signals from experimental bias and random fluctuations. A wide range of both commercial and freely available software has been developed to address this need. The choice between the large number of different analysis methods available, however, can be daunting. Unfortunately, there is no 'silver bullet': no single 'catch-all' method solves the problem of appropriate microarray data analysis. The right path will depend on the biological problem examined, the implemented experimental design and even the microarray type.

Box 1: Experimental design

Efficiency of data analysis begins with good experimental design. It must be emphasised that it is always advisable to consult both a statistician and someone with experience conducting similar microarray experiments *before* embarking on sample/data collection. Much effort and resources have been wasted by inappropriately designed experiments, for which an analyst can often only provide a 'post mortem' of what has gone wrong in experimental design. Essentially, good design considers all of the following issues:

- A clearly defined biological question/hypothesis much facilitates analysis. For exploratory studies: how will biological patterns be distinguished from technical

ones? What independent means of validation are there?

- A sensible experimental arrangement must be chosen. For example, a dye swap design addresses dye bias effects, while the use of a reference sample in one channel aids normalisation and comparison across slides.[2,4,5] If it is likely that additional data will be collected, a reference design may be indicated simply by the increased flexibility it gives.

- Can identification and control of systematic experimental errors be aided? This could, for example, mean additional measurements, eg of collected sample amounts, or involve the liberal use of calibration points, such as control probes and exogenous spikes.[6]

- When a large number of probes are expected to differ between samples the use of 'housekeeping genes' and/or exogenous spikes will actually be critical for subsequent analysis (also see discussion in main text). This is likely, for example, when very different samples are assessed, or when small or 'boutique' arrays are used that feature a set of probes already preselected for their involvement in a particular biological process. A large number of control probes must then be placed randomly across the array to allow normalisation and subsequent analysis.

- An optimal number of biological or technical replicate samples[7] must be assessed, as dictated by the experimental situation, observed noise levels (eg from pilot experiments) and the available budget.[8] While the optimal replication strategy may strongly vary between different experiments, as a rule of thumb, a minimum of five biological replicates are typically required for robust inference.[9]

- Data collection in compliance with the Minimal Information About a Microarray Experiment (MIAME) is increasingly recommended or required for publication.[10,11]

Box 2: Platform options – too many flavours of chips

There are a variety of microarray platforms available, the technical properties of which affect choices in subsequent data analysis. In the remainder of this paper, microarray 'slide' shall mean the support of the microarray probes, no matter whether that is a traditional 'microscope glass slide', or not.

Platforms can be categorised by their probe type. There are three distinct types of microarray probes – cDNA probes, long oligonucleotide probes and short oligonucleotide probes. Long oligonucleotides (typically 50- to 70-mers) are thought to mimic the properties of cDNA probes offering high sensitivity and good specificity, while giving better probe homogeneity.[12] Oligonucleotide probes can target specific regions of a gene and hence allow the detection of splice variants.

For both cDNA and long oligonucleotide arrays, typically only one probe is designed for each gene that is to be probed (splice variants are an additional complication that will not be addressed here). Short oligonucleotide arrays are typified by the GeneChip™ platform (Affymetrix). The Affymetrix system involves the *in situ* manufacture of short oligos on a glass surface using photolithography. The initial method for doing this was limited to the production of 25-mers and this has meant that Affymetrix has adopted a design approach for their arrays specific to the use of such short probes.[13,14] For each gene, a unique region is identified, then a series of 11–20 complementary probes spanning this region are synthesised. These

complementary probes are referred to as 'Perfect Match' probes (PM). Each PM probe is then paired with a 'Mismatch' probe (MM), which has the same sequence as the PM except the central base is replaced with a mismatched nucleotide. The complete set of PM and MM probe pairs for each gene is referred to as a 'probe set'.

The process of inferring a single estimate of gene expression from a probe set has recently become an area of very active research.[15,16] Results of a standardised benchmark of competing methods are available.[17] Ideally, this process of probe summarisation should be done simultaneously with normalisation. In practice, for a lack of truly integrated methods, most laboratories normalise across chips before probe summarisation.

Microarray data analysis is often, for convenience, split into several distinct steps, such as: (i) image analysis, (ii) data preprocessing/normalisation and (iii) higher-level analysis and interpretation. While fully automated microarray image analysis is still an area of active research, most laboratories use the software that came with their scanner, with operators contributing to random error.[8] The myriad of tools available for higher-level analysis reflects the wide range of applications for microarray experiments. Approaches include exploratory techniques, such as all forms of clustering,[18,19] algorithms for classification,[20–22] statistical hypothesis tests (eg for differential expression[23] and fitting more complex probabilistic models – Bayesian networks,[24] hierarchical models,[8,25] . . .). All analyses, though, have in common the need to identify and remove experimental bias from measurements before further study. It should be emphasised that correct normalisation is essential for meaningful data analysis, as distortions through artefacts can be significantly larger than the biological signals of interest. This tutorial provides an introduction to the underlying concepts of common approaches to this problem, and gives a first guide to choosing between the available microarray data transformation and normalisation methods for this task. Where appropriate, we point to corresponding packages from the open-source Bioconductor software repository.[26]

THE CHALLENGE OF MICROARRAY DATA NORMALISATION

Experimental sources of unwanted signal variation

The experimental process leading to the signals that comprise microarray data is fairly complex. There is a large variety of sources of systematic global differences between measurements that reflect differences in:

- sample preparation, eg time from surgical removal until cryo-preservation;
- RNA extraction efficiency between samples;
- overall amplification yield, if an amplification protocol is employed;
- overall labelling yield;
- overall hybridisation efficiency and washing stringency;
- incorporation efficiency for different dyes during direct labelling;
- fluorescence gain of the excitation/detection system for different dyes.

Further sources of variation include specific effects of the manufacturing process, like differences in

- characteristics of individual printing pins;
- properties of specific probe source plates;

- batch effects (slides, buffers, etc).

More complex distortions can be caused by

- sequence-specific label incorporation bias;
- spatial variation of hybridisation efficiency and washing stringency;
- non-linear effects, such as non-linear dye fluorescence response and interaction between dyes (cross-talk and quenching) – these can give rise to complex confounding of effects, eg bias in the differential signal as a function of the average intensity.

Last but not least, it certainly should be mentioned that both the operator and the date of the experiment are significant sources of variation. The general difficulty of combining data from different laboratories is well appreciated. Technical effects often dominate biologically relevant signals in multi-centre studies (Figure 1).

Control for technical variation by conservative normalisation

Ideally, the sources of these non-biological variations would be determined and accounted for, or minimised at the experimental stage. Often, this is only partly possible and, to allow comparison of different measurements, the remaining differences need to be controlled for during analysis, a process called *normalisation*.

Normalisation procedures remove unwanted variances from data by exploiting and enforcing *known or assumed invariances* of the data – in an exact, averaging or probabilistic manner, and with varying degrees of robustness towards outliers. Considering the complexity of the problem, it is not surprising that new normalisation methods are being introduced continuously. Instead of providing a list of the many methods available, we here restrict ourselves to a review of basic

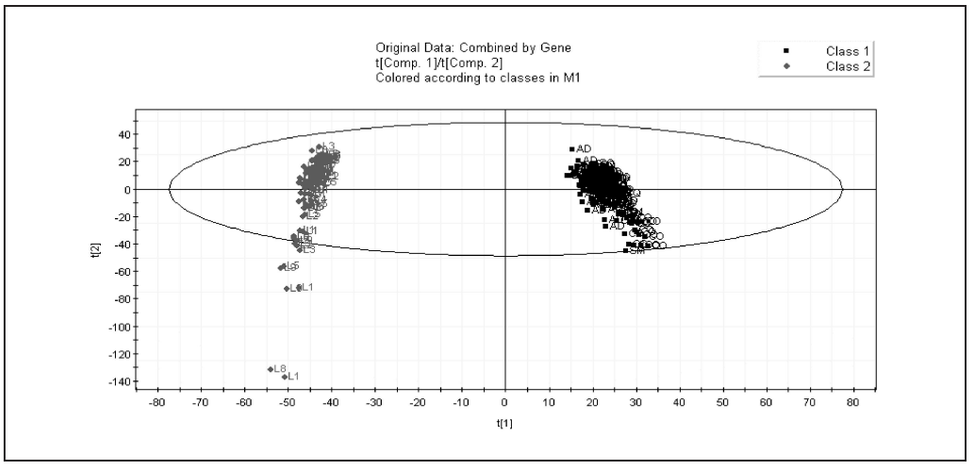

Figure 1: Affymetrix lung cancer data from Harvard (grey) and Michigan (black). The plot in principal components space[27] shows that, despite the laboratories' use of standardised commercial microarray platforms (HGU95a and HuGeneFL chips), the main source of variation maps to the laboratory in which the data have been generated, indicating that differences in sample collection protocols and the microarrays used dominated the biological effect of interest

Reproduced with permission from Kluwer Academic Publishers, 'Methods of Microarray Data Analysis IV', 2005, Shoemaker, J. S. and Lin, S. M. (Eds), 'Making sense of lung carcinomas gene expression data: Integration and analysis of two Affymetrix platform experiments', Xiwu Lin *et al.*, Fig. 3

concepts. In the next section, several well-known methods and the invariances they require are discussed. The fundamental assumptions of normalisation methods differ and are satisfied by real data sets to varying degrees. Consequently, a different normalisation method must be chosen appropriate for any specific experiment. Moreover, as the 'correct' result of a microarray experiment is usually not known, it is very difficult to assess the relative performance of normalisation methods on realistic data sets. In selecting an appropriate normalisation procedure for a particular experiment, a guiding principle is hence the conservative nature of the method – the lower the number of parameters, the less arbitrary. While removal of complex artefacts may require procedures with a higher number of parameters, it is easy to unwillingly introduce novel artefacts either by over-fitting or as a result of failure in separating biological from technical patterns. Satisfying all assumptions of the employed methods, the most conservative approach should hence be chosen that removes the dominant artefacts.

NORMALISATION METHODS – COMMON APPROACHES

In a discussion of several common normalisation approaches, we will highlight situations in which their assumptions are not met, and different methods need to be applied. We start with the most conservative procedures.

Rescaling, robustness, 'housekeeping genes' and spike controls

A very conservative, simple and popular approach is the rescaling of each sample by a constant. This controls for overall differences in signal strength as, for example, caused by different extraction or labelling yields. Instead of rescaling by the mean signal intensity, often the median signal intensity is used, giving considerable robustness to outliers. Even this simple method, however, makes a critical assumption, namely, that the median signal intensity will be invariant between samples. A popular variant of this approach assumes invariance of a *subset* of genes (typically called 'housekeeping genes'), rescaling by their median signal intensity. Although there is good reason to doubt that such a set of invariant genes exists in general (eg some studies have found the presumed 'housekeeping' gene *GAPDH* to be quite variable[28,29]), there may be sets of genes that are non-differentially expressed across the samples of a particular experiment. Nevertheless, there are cases where true biological differences between samples violate all of the above assumptions. Consider, for example, the massive generic down-regulation of gene expression in samples of dying tissue. This is a typical example of an experimental situation in which it is very difficult to find an appropriate normalisation procedure. It furthermore highlights the value of considering such issues at the stage of experimental design. If known quantities of spike RNA are added to the sample before extraction, standardised sample amounts are ensured, and the slides contain a *high number* of spike probes, then the spike signals can be used for normalisation, as these can now be assumed to be invariant. If such spikes are not available, all one can do is hope that the technical variation in signal intensity is smaller than the biological effect examined, and work on the unnormalised data.

Explicit error models, robust fit by iterative trimmed least squares

For subsequent analysis, it is often desirable to decouple random error and signal intensity. It is very common to transform microarray data to a log-scale before further analysis. A logarithmic transform certainly collapses the original range of the signal, which can span over five orders of magnitude. Moreover, a log-transform decouples a random

multiplicative error e^{η} from the true signal intensity μ in the measurement y:

$$y = \mu e^{\eta} \Rightarrow \log y = \log\mu + \eta \quad (1)$$

It is well known that microarray data variance increases with signal intensity, which is in line with such a model. It is likely that the source of this variance is the hybridisation process itself, as it cannot be explained by instrument error (Figure 2).[30]

A purely multiplicative model, however, unrealistically predicts vanishing measurement error for very small signals, whereas in truth, there will always be some background noise. This can be seen in log–log plots of same–same comparisons (Figure 3). On a linear scale (left panel), the variance increases with signal intensity. On logarithmic scale (right panel), the variance is approximately constant for medium and high signal intensities, in line with the purely multiplicative noise model in equation (1). The measured variance for low-intensity signals, however, is larger than expected.

A more realistic error model allows for

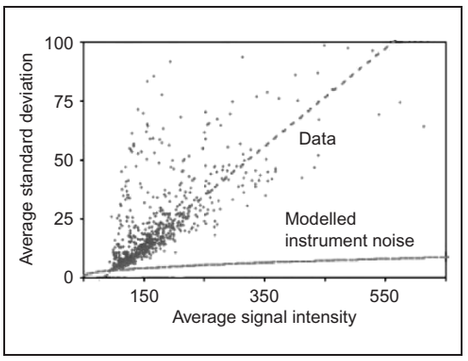

Figure 2: Increasing measurement variance with higher hybridisation signal[30] Reproduced with permission from Brown, C. S., Goodwin, P. C. and Sorger, P. K. (2001), 'Image metrics in the statistical analysis of DNA microarray data', *Proc. Natl Acad. Sci. USA*, Vol. 98, pp. 8944–8949. Copyright (2001) National Academy of Sciences, USA

both additive and multiplicative error terms:

$$y = a + b\mu e^{\eta} + \nu \quad (2)$$

Here, the constants a and b model a global background and gain, respectively. Just as the log-transforms decouples the true signal μ from the error term e^{η} in model (1), the asinh transform will decouple the

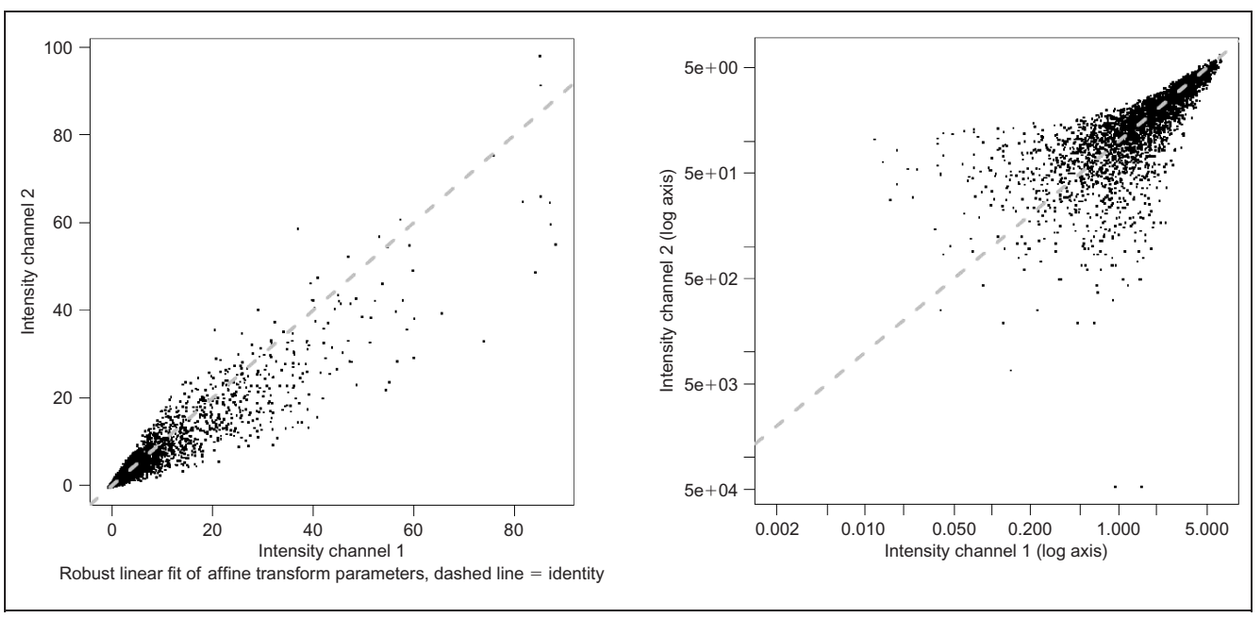

Figure 3: Data after affine transform normalisation. The two panels show the same data, which are from a same–same comparison. The left panel shows the data on a linear scale, while the right panel shows the effect of a $\log_2$-transform

true signal μ from the error terms e^{η} and ν in (2). As shown in Figure 4, it is approximately linear for small values; for larger ones it is well approximated by a logarithmic transform. Figure 5 shows how this model more successfully decouples variance and signal intensity in real data.

To apply model (2) for normalisation, the parameters a and b need to be obtained by fitting the model to the data. The measurements for all the genes of a sample share the same parameters in this model. Once the parameters have been obtained for each sample, the differences in global background and gain can be corrected for. Obviously, differentially expressed genes or technical outlier measurements should not affect this fit. The iterative trimmed least-squares fit implemented in the vsn package[31] (available from the Bioconductor repository) will identify a subset of non-differentially expressed non-outlier genes with a high breakdown point of 50 per cent. This means that up to 50 per cent of genes can be differentially expressed or technical outliers without this affecting the fit arbitrarily, making the result *robust*. In contrast to many other methods, this procedure even works reliably if the differentially expressed genes introduce a strong asymmetry, eg if they are mostly up-regulated (rather than having an even mix of up- and down-regulation).

It should be emphasised that this still is a very conservative approach, using only two parameters per sample. Alternatives

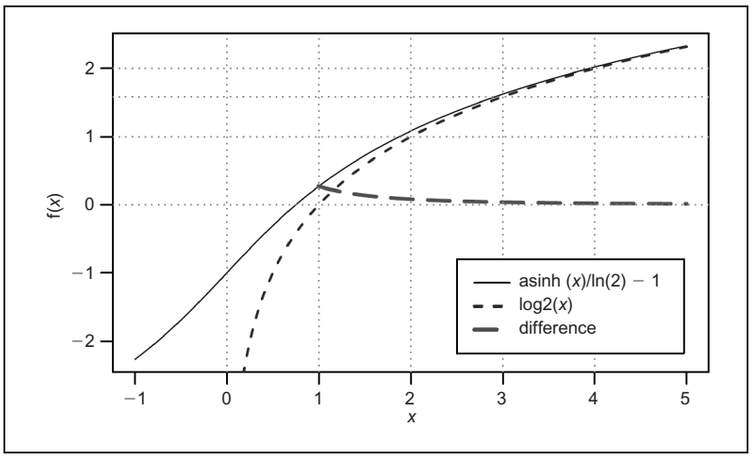

Figure 4: Comparison of an asinh- and a $\log_2$-transform. Note the deviation for values close to zero

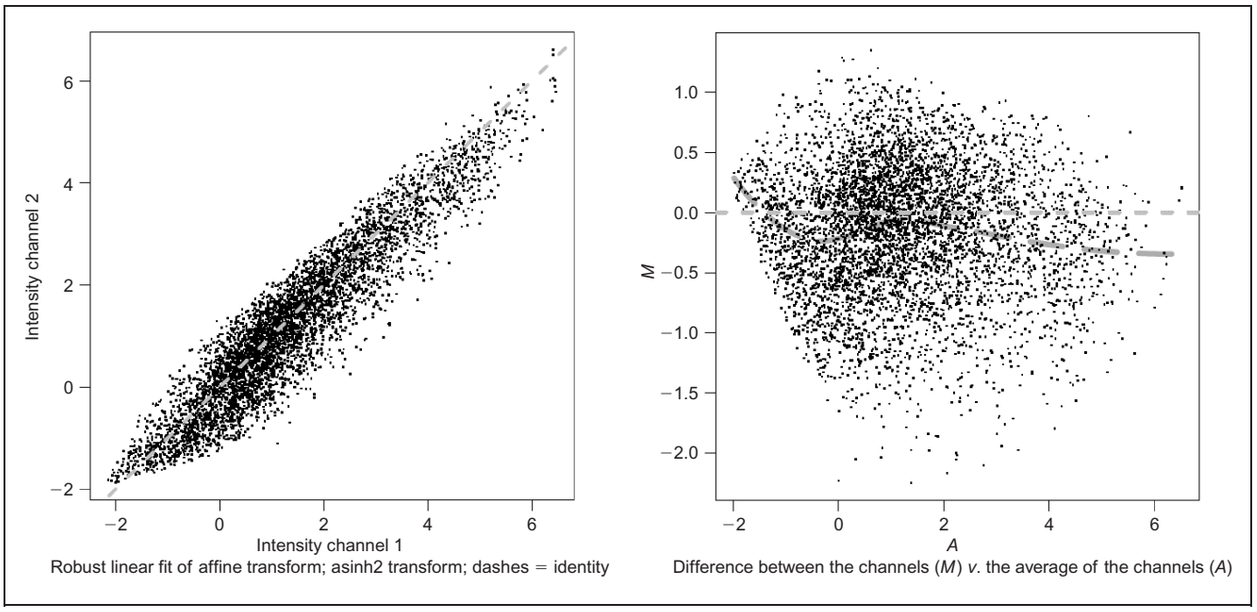

Figure 5: Data after variance stabilising normalisation (vsn). Both panels: on 'asinh-scale' the much-inflated fluctuations at low signal intensities have disappeared. The right panel shows the traditional '*M*(*A*)'-plot, with *M* the difference and *A* the mean of the two channels – artefacts are more obvious here. The dark-grey long-dashed line shows a Loess smoother, giving an indication of local trends in the plot. An 'S'-shaped distortion is clearly seen

to global, constant background and gain terms could either fit a slowly varying function as background or gain terms or attempt to measure these directly. Modelling spatial variation without additional measurements has the disadvantage of greatly increasing the number of parameters of the model. Hence, there has been a strong interest in obtaining approximate measurements for 'local' background intensities, eg by measuring the fluorescence intensity in the slide area between the spotted probes. Unfortunately, there are two sources of background fluorescence observed on microarrays: (a) smears that cover both the probes and the area between probes, and (b) non-specific binding of fluorescent material to the area between the spotted probes, only. If the former dominates, then estimation of local background intensities by measuring fluorescence between probes makes a lot of sense, and correcting for this local background will improve data quality. On the other hand, if non-specific binding to only the area between spots dominates, such a correction is inappropriate and can create artefacts, eg 'negative intensities'. Many laboratories find that this is indeed the case, and measurements suggest that the background obtained from fitting a model such as (2) actually corresponds to the self-fluorescence of the substrate slide before hybridisation.[30]

The $M(A)$-plot of Figure 5 clearly shows an 'S'-shaped technical artefact which has not been removed by this transformation. The methods discussed next can sometimes be used to remove such complex distortions.

Detrending of the signal difference as a function of the signal average

'Smile'- or 'S'-shaped trends in the difference M between samples as a function of the average signal intensity A are commonly observed in microarray data (Figure 5). A possibly cause for such trends may be differences in the non-linear fluorescence response of dyes. In a comparison of similar samples, ideally, there should be no such distortions. A very flexible but rather aggressive approach hence enforces this invariance $M(A) = 0$ by computing an 'average trend' in order to subtract it, eg by means of a Loess smoother, which is available from the loess function in the statistics environment R. While most bias in the trend of $M(A)$ can indeed be removed this way, this normalisation procedure is not conservative, having a large number of parameters to be obtained from the data, as well as an arbitrary parameter setting the 'window length' for the smoothing algorithm. This flexibility of the transform can become a problem when a researcher fails to detect that the assumption

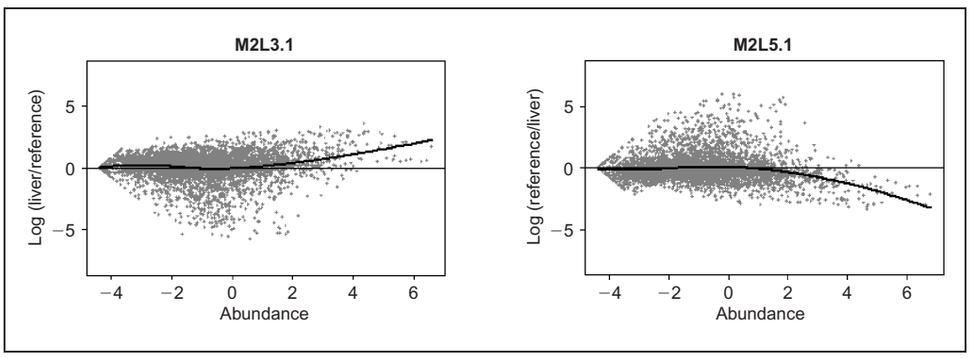

Figure 6: The left panel shows the $M(A)$ plot for a microarray experiment comparing liver tissue with a reference sample comprising a mix of liver, kidney and testis tissue.[32] The trend, shown by the solid line (a Loess smoother), reverses under dye swap (right panel). Two replicate experiments each confirmed this picture
Reproduced with permission from Coombes, K. (2002), 'Biology-driven clustering of microarray data: Applications to the NCI60 data set', CAMDA, Durham, NC, USA

$M(A) = 0$ does not hold for biological reasons. This can, for example, be the case if a large number of genes are asymmetrically differentially expressed (Figure 6). Simulations by Coombes and coworkers[32] have shown that such observed non-linear trends in $M(A)$ can be explained if 20 per cent of genes are specifically expressed in only one of the samples, which is, for example, easily perceivable for comparisons of rather different tissue types. Dye swap experiments can help detect such situations (Figure 6) to avoid indiscriminate application of, for example, a Loess smoother to detrend $M(A)$. Obviously, inappropriate detrending would both remove biological signal *and* create further technical artefacts.

Equalisation of signal distributions

While the $M(A)$ detrending described above can deal with a large variety of distortions, it is computationally expensive and its focus on expression difference does not lend itself well to situations with large numbers of (unpaired) samples.[33] A different approach, which is similarly generic, deals directly with the signal distributions of individual samples. For sufficiently similar samples, or if there are no strong asymmetries expected in differential expression, it is valid also to expect similar signal distributions for all samples, and to enforce this assumption by a normalisation transform. This process can be depicted by quantile–quantile (Q–Q) plots (Figure 7) – data are normalised by projection so that their Q–Q plot to a reference distribution gives a straight line. The reference distribution can be that of an arbitrary array, or a constructed/artificial one. A particularly efficient implementation[34] has made this approach very popular as part of the affy package for the analysis of Affymetrix slides in Bioconductor. This direct mapping, however, is not robust in the tails of distributions, where data tend to be sparse. An approach fitting natural cubic splines to quantile data[33] seems to resolve this issue in a computationally efficient way.

Like the transform discussed in the previous section, quantile normalisation is not a conservative transform and should not be used in situations where the signal distributions of samples are expected to differ, eg when asymmetric differential expression of many genes. Figure 8 shows that the signal distribution in male and female fruitflies differs, reflecting true

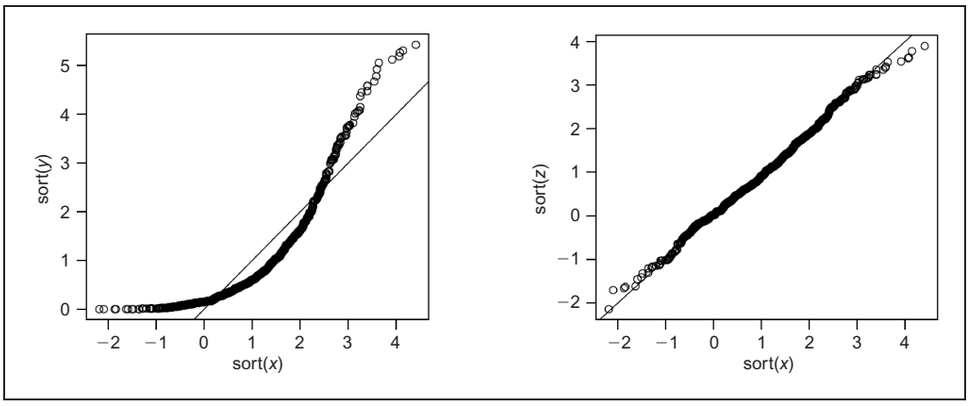

Figure 7: Quantile–quantile (Q–Q) plots for rather different distributions (left panel) and for similar distributions (right panel). In Q–Q plots, the quantiles of one sample are graphed as a function of the same quantiles of the other, allowing direct comparison of, eg, the 25th percentile of the two samples (which marks the value which is larger than 25 per cent of the data in each sample, respectively). Identical distributions will hence give a straight line in Q–Q plots. For samples of equal size, a Q–Q plot can easily be generated as scatter plot of the sorted samples to be compared

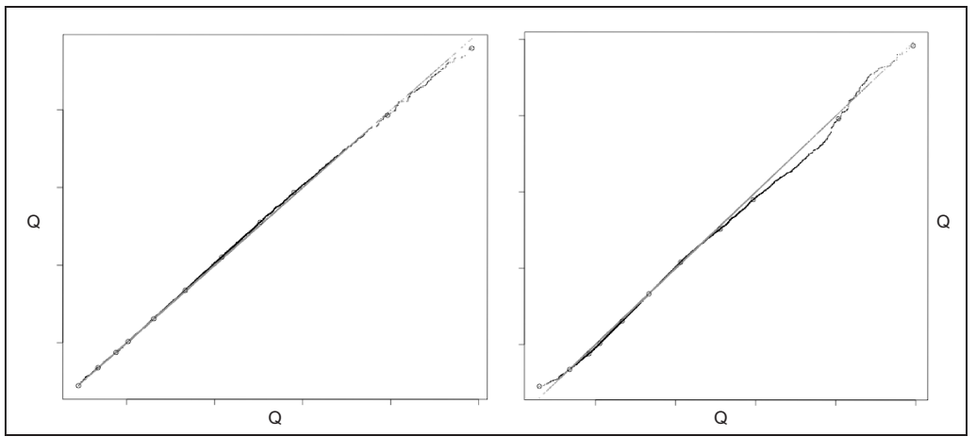

Figure 8: Q–Q plots of two female fruitfly samples (left) show similar signal distribution, while a comparison of samples from a male and a female fly consistently show distinct differences (right)

biological differences. Quantile normalisation of these samples would not only remove this biological signal but also introduce further technical artefacts.

Spatial detrending

Well-designed microarrays have replicate probes distributed in a random order across the slide surface to allow separation of spatial artefacts from biological signal. Unfortunately, many early arrays (from both academic and commercial providers) have not followed this design criterion for reasons of manufacturing convenience. For these older-generation slides – which are still widely employed – computational approaches are prone to fail, separating spatial artefacts and biological signal. Where possible, one may try to alleviate this problem by hybridising every other replicate slide rotated by 180°. The most stringent designs are currently implemented in the most recent high-density arrays with multiple redundant probes per gene randomly distributed across the slide (eg as available from Affymetrix or Nimblegen).

In modern microarrays, spatial trends in signal intensity are usually a sign of technical artefacts, such as caused by unequal hybridisation conditions or washing. For example, one can sometimes clearly identify the areas beneath the probe inlet and outlet holes of an Affymetrix slide (Figure 9). Traditionally, Loess smoothers are computed in order to estimate and subtract spatial trends. While this can reduce variance between replicates, recent calibration experiments suggest that spatial variation of *unspecific* binding cannot be corrected this way (Kreil *et al*., manuscript in preparation).

The original design of a two-channel DNA microarray is actually quite ingenious, as it already provides approximate self-normalisation for spatial variation of hybridisation and washing efficiency by performing the measurements of samples to be directly

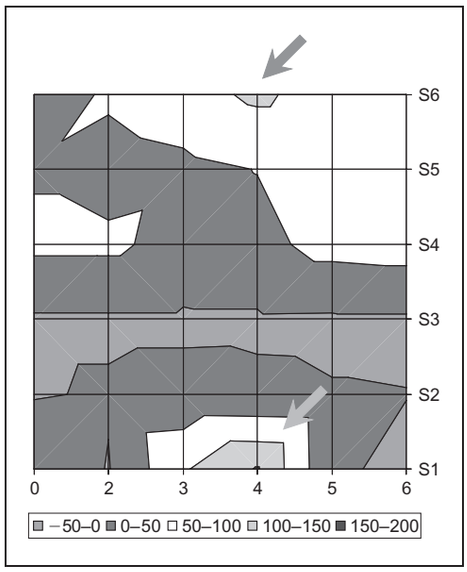

Figure 9: Spatial trends in the signal intensities measured on an Affymetrix slide. The brighter areas beneath the probe inlet and outlet holes (arrows) can be clearly identified.[35] Reproduced with permission from Ptitsyn, A. A. (2002), 'Topological adjustments to the Genechip expression values', CAMDA, Durham, NC, USA

compared in a single probe that is subject to the same local reaction conditions. To extend the sensitivity of screens, however, replicate slides are needed. In addition, more complex experiments require the joint analysis of multiple samples. In either case, modern work needs normalisation between slides, not just between channels. It hence seems natural to expect future microarrays to feature novel experimental designs that support between slide normalisation directly, rather than rely on purely computational approaches.

SUMMARY AND OUTLOOK

We have explained how modern microarray experiments, which require the combination of data from multiple slides, rely on removing unwanted technical variation between sample measurements, a process called normalisation. We have then given an introduction to common concepts and principles employed by current normalisation methods, including the assumptions that they exploit and that they rely on. It is essential that a normalisation/data-transformation method appropriate to the properties of the experiment in question is chosen. We have discussed the strengths and limitations of several popular approaches, providing guidance for their use. The normalisation of microarray data is still a very active area of research, with frequent novel interactions between microarray design and analysis methods (cf. probe summarisation, see Box 2). Just as the original design of two-channel microarrays had technical features reducing the impact of unwanted experimental variance, it can be expected that future microarray designs together with advanced models will further improve the ability to normalise data across samples. Such progress will allow an even more quantitative and sensitive approach to modern applications in the clinical and biological sciences.

David P. Kreil
Dept of Genetics,
Univ. of Cambridge,
Downing Street,
Cambridge CB2 3EH, UK
Tel: +44(0)1223 764107
Fax: +44(0)1223 333992
E-mail: D.Kreil@gen.cam.ac.uk
and
Inference Group,
Cavendish Laboratory,
Univ. of Cambridge,
Madingley Road,
Cambridge CB3 0HE, UK
and
Roslin R. Russell
Dept of Genetics,
Univ. of Cambridge,
Downing Street,
Cambridge CB2 3EH, UK
E-mail: Ros@gen.cam.ac.uk

References

1. Duggan, D. J., Bittner, M., Chen, Y. *et al.* (1999), 'Expression profiling using cDNA microarrays', *Nat. Genet.*, Vol. 21, pp. 10–14.
2. Yang, Y. H. and Speed, T. (2002), 'Design issues for cDNA microarray experiments', *Nat. Rev. Genet.*, Vol. 3, pp. 579–588.
3. Auburn, R. P., Kreil, D. P., Meadows, L. A. *et al.* (2005), 'Spotting high quality DNA microarrays', in preparation.
4. Churchill, G. A. (2002), 'Fundamentals of experimental design for cDNA microarrays', *Nat. Genet.*, Vol. 32 Suppl., pp. 490–495.
5. Simon, R. M. and Dobbin, K. (2003), 'Experimental design of DNA microarray experiments', *Biotechniques*, March Suppl., pp. 16–21.
6. van de Peppel, J., Kemmeren, P., van Bakel, H. *et al.* (2003), 'Monitoring global messenger RNA changes in externally controlled microarray experiments', *EMBO Rep.*, Vol. 4, pp. 387–393.
7. Pan, W., Lin, J. and Le, C. T. (2002), 'How many replicates of arrays are required to detect gene expression changes in microarray experiments? A mixture model approach', *Genome Biol.*, Vol. 3, p. R22.
8. Wernisch, L., Kendall, S. L., Soneji, S. *et al.* (2003), 'Analysis of whole-genome microarray replicates using mixed models', *Bioinformatics*, Vol. 19, pp. 53–61.
9. Pavlidis, P., Li, Q. and Noble, W. S. (2003), 'The effect of replication on gene expression microarray experiments', *Bioinformatics*, Vol. 19, pp. 1620–1627.
10. Brazma, A., Hingamp, P., Quackenbush, J.

et al. (2001), 'Minimum information about a microarray experiment (MIAME) – toward standards for microarray data', *Nat. Genet.*, Vol. 29, pp. 365–371.

11. Brazma, A., Parkinson, H., Sarkans, U. *et al.* (2003), 'ArrayExpress – a public repository for microarray gene expression data at the EBI', *Nucleic Acids Res.*, Vol. 31, pp. 68–71.
12. Wang, H. Y., Malek, R. L., Kwitek, A. E. *et al.* (2003), 'Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays', *Genome Biol.*, Vol. 4, p. R5.
13. Lockhart, D. J., Dong, H., Byrne, M. C. *et al.* (1996), 'Expression monitoring by hybridization to high-density oligonucleotide arrays', *Nat. Biotechnol.*, Vol. 14, pp. 1675–1680.
14. Wodicka, L., Dong, H., Mittmann, M. *et al.* (1997), 'Genome-wide expression monitoring in *Saccharomyces cerevisiae*', *Nat. Biotechnol.*, Vol. 15, pp. 1359–1367.
15. Cope, L. M., Irizarry, R. A., Jaffee, H. A. *et al.* (2004), 'A benchmark for Affymetrix GeneChip expression measures', *Bioinformatics*, Vol. 20, pp. 323–331.
16. Irizarry, R. A., Hobbs, B., Collin, F. *et al.* (2003), 'Exploration, normalization, and summaries of high density oligonucleotide array probe level data', *Biostatistics*, Vol. 4, pp. 249–264.
17. URL: http://affycomp.biostat.jhsph.edu/
18. Tamayo, P., Slonim, D., Mesirov, J. *et al.* (1999), 'Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation', *Proc. Natl Acad. Sci. USA*, Vol. 96, pp. 2907–2912.
19. Slonim, D. K. (2002), 'From patterns to pathways: Gene expression data analysis comes of age', *Nat. Genet.*, Vol. 32 Suppl., pp. 502–508.
20. Brown, M. P., Grundy, W. N., Lin, D. *et al.* (2000), 'Knowledge-based analysis of microarray gene expression data by using support vector machines', *Proc. Natl Acad. Sci. USA*, Vol. 97, pp. 262–267.
21. Vohradsky, J. (2001), 'Neural network model of gene expression', *FASEB J*, Vol. 15, pp. 846–854.
22. Theilhaber, J., Connolly, T., Roman-Roman, S. *et al.* (2002), 'Finding genes in the C2C12 osteogenic pathway by *k*-nearest-neighbor classification of expression data', *Genome Res.*, Vol. 12, pp. 165–176.
23. Baldi, P. and Long, A. D. (2001), 'A Bayesian framework for the analysis of microarray expression data: regularized *t*-test and statistical inferences of gene changes', *Bioinformatics*, Vol. 17, pp. 509–519.
24. Friedman, N., Linial, M., Nachman, I. and Pe'er, D. (2000), 'Using Bayesian networks to analyze expression data', *J. Comput. Biol.*, Vol. 7, pp. 601–620.
25. Lönnstedt, I. and Britton, T. (2005), 'Hierarchical Bayes models for cDNA microarray gene expression', *Biostatistics* (in press) (URL: http://www.math.uu.se/~ingrid/bio6.pdf).
26. Gentleman, R. C., Carey, V. J., Bates, D. M. *et al.* (2004), 'Bioconductor: Open software development for computational biology and bioinformatics', *Genome Biol.*, Vol. 5, p. R80.
27. Lin, X., Park, D., Eslava, S. *et al.* (2003), 'Making sense of human lung carcinomas gene expression data: Integration and analysis of two Affymetrix platform experiments', in 'Proceedings of the Fourth International Conference for the Critical Assessment of Microarray Data Analysis (CAMDA)', 12th–14th November, Durham, NC.
28. Ke, L. D., Chen, Z. and Yung, W. K. (2000), 'A reliability test of standard-based quantitative PCR: Exogenous vs endogenous standards', *Mol. Cell Probes*, Vol. 14, pp. 127–135.
29. Suzuki, T., Higgins, P. J. and Crawford, D. R. (2000), 'Control selection for RNA quantitation', *Biotechniques*, Vol. 29, pp. 332–337.
30. Brown, C. S., Goodwin, P. C. and Sorger, P. K. (2001), 'Image metrics in the statistical analysis of DNA microarray data', *Proc. Natl Acad. Sci. USA*, Vol. 98, pp. 8944–8949.
31. Huber, W., Von Heydebreck, A., Sultmann, H. *et al.* (2002), 'Variance stabilization applied to microarray data calibration and to the quantification of differential expression', *Bioinformatics*, Vol. 18 Suppl. 1, pp. S96-104.
32. Stivers, D., Wang, J., Rosner, G. and Coombes, K. (2002), 'Organ-specific differences in gene expression and UniGene annotations describing source material', in 'Proceedings of the Second International Conference for the Critical Assessment of Microarray Data Analysis (CAMDA)', 14th–15th November, Durham, NC.
33. Workman, C., Jensen, L. J., Jarmer, H. *et al.* (2002), 'A new non-linear normalization method for reducing variability in DNA microarray experiments', *Genome Biol.*, Vol. 3, p. R48.
34. Bolstad, B. M., Irizarry, R. A., Astrand, M. and Speed, T. P. (2003), 'A comparison of normalization methods for high density oligonucleotide array data based on variance and bias', *Bioinformatics*, Vol. 19, pp. 185–193.
35. Ptitsyn, A. (2002), 'Topological adjustments to the Genechip expression values', in 'Proceedings of the Second International Conference for the Critical Assessment of Microarray Data Analysis (CAMDA)', 14th–15th November, Durham, NC.