Shiqin Helen Guan[1]**, Laura Bonnett**[2]**, Julia Brettschneider**[23*]

Using gene subsets in the assessment of microarray data quality for time course experiments

The application of established microarray data quality measures to time course experiments potentially gives misleading results. In particular, genuine biological variation my be mistaken for technical artifacts. We suggest tailoring standard methods to time course data by restricting the assessment to subsets of genes selected on the basis of the experiment. The method is tested on two different kinds of experimental data sets, one from a developmental process and one from a circadian process. With these modifications, quality assessment for microarray data can be tuned to appropriately address the special situation of time course experiments.

KEYWORDS: quality control, microarrays, time course data, Affymetrix chips, relative log expression, normalized unscaled standard errors.

1. INTRODUCTION

Nearly two decades after their introduction, microarray experiments are now standard practice in many modern molecular laboratories in both academia and the biomedical industry. Microarrays enable scientists to simultaneously assay the biochemical activity of many, even of all the genes in the genome in a biological sample at a given time.

Some laboratories make their own microarrays, while others use commercial platforms such as Affymetrix or Agilent. Whichever way, the scientific community has had to acknowledge that measurement errors can not be avoided entirely and may lead to inconclusive or non-reproducible experiments. Inter-lab and inter-platform comparisons have demonstrated this convincingly, see e.g. Hwang et al. (2004). A need for systematic quality control of the data from such experiments has by now been widely acknowledged.

How to optimally assess the quality of this new kind of data is not obvious and has emerged as a new research topic for statisticians. Brettschneider et al. (2008) characterize the collection of microarray data from a traditional quality assessment perspective in the spirit of Shewhart (see, e.g., Shewhart (1939)). They suggest several explicit quality measures for this new kind of data and compare them to the quality report provided by the commercial software.

In the case of time course experiments, the common quality assessment methods may give misleading information (see Guan et al. (2007)). In particular, there is a risk of mistaking genuine biological

[1]Queen's University, Department of Mathematics and Statistics, Kingston, Ontario, Canada;

[2]Warwick University, Department of Statistics, Coventry, United Kingdom;

[3]Queen's Cancer Research Institute, Division of Cancer Care Epidemiology and
  Department of Community Heath & Epidemiology, Kingston, Ontario, Canada;

*Corresponding author: `julia.brettschneider@warwick.ac.uk`

variation for apparatus noise or bias linked to experimental conditions. In fact, these methods are based on assumptions about the behavior of the the majority of the genes. Not too many of the genes are expected to change and the number of upregulated genes is expected to be approximately the number of downregulated genes. Time course microarray experiments, however, often differ in just these aspects from most other kinds of common microarray experiments. The fraction of differentially expressed genes is often much higher, and inhomogeneous over time. In order to avoid that this undermines the quality assessment, Guan et al tried to restrict it to temporarily stable genes in a developmental series. Their first approach uses 141 low-varying, genes selected using *a priori* biological knowledge. While this sounds like a reasonable approach, the results were not convincing. They also tried to base the quality assessment on genes that had been selected as temporarily low-varying genes in a first round of analysis and found that strategy more promising.

This paper aims at systematically developing a subset-based approach to quality assessment for microarray time course data. We focus on the case of short oligonucleotide microarrays. First, we assess the data quality using the RLE and NUSE measures introduced by Brettschneider et al. and compare the results with the scores supplied by the Affymetrix GCOS quality report. We then suggest a data-driven method to determine subsets of temporarily stable genes appropriate to the particular experiment. We consider subsets of various sizes. The approach is tested on data sets from two different types of microarray time course experiments: a developmental time series and a circadian rhythm series. Both data sets come from experiments on model organisms conducted in academic laboratories on short oligonucleotide microarrays manufactured by Affymetrix (see Section 4 for details).

2. BIOLOGICAL AND TECHNOLOGICAL BACKGROUND

The blueprint for the construction of a biological organism is its genome stored in DNA molecules in its set of chromosomes. By and large, genes are transcribed into RNA which is translated into proteins, the building blocks of the cells of the organism. These processes are substantially affected by many factors including cell type, developmental stage, genetic background and environmental conditions. As a result, the proteins vary in both shape and abundance.

A central aim in molecular biology is to relate the state of an organism to molecular observations. This includes the study of the gene and protein functions and interactions, and their alteration in response to changes in environmental and developmental conditions. Some explicit examples for such research questions are: Which genes are involved in the flowering of strawberries? Which molecular mechanisms guide the development of a fruit fly embryo? Does a certain protein inhibit human brain tumor growth? How does the activity of that protein affect other proteins and genes?

Such experiments used to focus on just one gene or protein at a time. With the introduction of high throughput methods in the 1990s, the abundance of huge numbers of such macromolecules could be measured in one experiment. The most common kind of high throughput molecular measurement technologies are *gene expression microarrays,* which measure the mRNA transcript abundance for tens of thousands of genes simultaneously.

A DNA microarray consists of a glass surface with a large number of distinct fragments of DNA

called *probes* attached to it at fixed positions. A fluorescently labelled sample containing a mixture of unknown quantities of DNA molecules called the target is applied to the microarray. Under the right chemical conditions, single-stranded fragments of target DNA will base pair with the probes which are their complements, with great specificity. This reaction is called *hybridization.* The raw data produced in a microarray experiment consists of scanned images of the microarrays after hybridization. The image intensity in the region of a probe is proportional to the amount of labelled target DNA that base pairs with that probe. In this way we can measure the abundance of thousands of DNA fragments in a target sample simultaneously. Microarrays are also referred to as (gene) chips or slides.

In this paper, we will focus on one of the standard kinds of microarrays, the *short oligonucleotide microarrays* produced, for example, by the company Affymetrix. They use a number of *probes* per gene (e.g. 14 for the Affymetrix drosophila array) randomly arranged over the array. Such a set is called the *probe set* for that gene. The probes are 25 nucleotides long and are successively synthesized to the chip using photolithography technology. The commercial software GCOS (see GCOS (2004)) includes a *quality report* with a dozen scores for each microarray (see Section 3). These scores are indirect indicators of data quality which are computed on the basis of individual arrays.

3. QUALITY ASSESSMENT METHODS FOR TIME COURSE DATA

In a first round of analysis, we will use three existing approaches for the assessment of short oligonucleotide microarray data. They are all based on the whole set of genes in the microarray. We summarize their definition here; more details and many data examples can be found in Brettschneider et al. (2008)l.

Intuitive approach: RLE. The microarray data needs to be pre-processed. This includes background adjustment, normalization and gene expression estimation. (We will use the RMA-based method described below, but any sensible pre-processing method can be used.) The *Relative Log Expression (RLE)* of a probe sets in array is computed as the log ratio of its expression value in that array and its median expression in a hypothetical reference array. The latter is usually obtained as the median of all arrays in that batch, constructed probe set by probe set. Under the biological assumption that the majority of gene is not differentially expressed, and that about the same amount of genes is upregulated as downregulated, the collective behavior of the RLE values can be used as a quality indicator. A good quality array has an RLE-distribution with a median around 0 and a small IQR.

RMA-based approach: NUSE. This approach uses the Robust Multichip Analysis (RMA) model by Irizarry et al. (2003). For a fixed probe set, the background adjusted and normalized intensities y_{ij} of probe j and microarray i are modeled as $\log_2 y_{ij} = \mu_i + \alpha_j + \varepsilon_{ij}$, with α_j a *probe affinity effect* (with zero-sum constraint), μ_i the log scale expression level for array i, and i.i.d. centered errors ε_{ij}. The use of an iteratively re-weighted least squares algorithms delivers a robust probe set expression estimate for this probe set for each array using the inverses w_{ij} of the residuals to down-weight malfunctioning probes. Let $W_i = \sum_j w_{ij}$ be the *total probe weight* (of the fixed probeset) in chip i. The *Unscaled Standard Error (USE)* of a fixed probe set is defined as $\sqrt{\sum_j w_{ij}^2}$ divided by the total probe weight $\sum_j w_{ij}$ in chip i. Finally, normalizing this expression by dividing by the median of all such expression over all chips is call the *Normalized Unscaled Standard Error (NUSE)*. Under the biological assumption that the

3

majority of gene is not differentially expressed, the collective behavior of the NUSE values measures the quality of the data from this array, relative to the quality of the data from the other arrays from the same experiment and jointly analysed. A good quality array has a NUSE-distribution with a median around 1 and a small IQR.

GCOS quality report. Affymetrix supplies the software GCOS for the analysis of their microarray data. Part of the output is a quality report which presents a number of microarray-wide quality scores. We use of the most common of these quality scores described in the following according to GCOS (2004) and Affymetrix (2001). The scores are typically used as relative scores to evaluate the quality of microarrays within a comparable batch of microarrays. Acceptance ranges for these scores are often not available, because they depend strongly on the experimental set-up and the biological process studied in the experiment. *Percent Present:* The percentage of probesets called *Present* by the Affymetrix detection algorithm. This value depends on multiple factors including cell/tissue type, biological or environmental stimuli, probe array type, and overall quality of RNA. Replicate samples should have similar Percent Present values. Extremely low Percent Present values indicate poor sample quality. In order to make comparisons with other scores more convenient, we also use *Percent Absent*. This is defined as $100\% -$ Percent Present. *Scale Factor:* Multiplicative factor applied to the signal values to make the 2% trimmed mean of signal values for selected probe sets equal to a constant. *Average Background:* Average of the lowest 2% cell intensities on the chip. Affymetrix does not issue official guidelines, but mentions that values typically range from 20 to 100 for arrays scanned with the GeneChip Scanner 3000. A high background indicates the presence of nonspecific binding of salts and cell debris to the array. *Raw Q (Noise):* Measure of the pixel-to-pixel variation of probe cells on the chip. The main factors contributing to Noise values are electrical noise of the scanner and poor sample quality. *GAPDH 3' to 5' ratio (GAPDH 3'/5'):* Ratio of the intensity of the 3' probe set to the 5' probe set for the gene GAPDH. It is expected to be an indicator of RNA quality.

For more details about RLE and NUSE see Brettschneider et al. (2008). They can be computed using the BioConductor R-package `affyPLM` described in BCBS05. For the computation of the GCOS quality scores we use the BioConductor R-package `simpleaffy` described in Wilson and Miller (2005).

Stable gene subset-based quality assessment . The second round of quality assessment is based on subsets of temporarily stable genes the rationale being that those genes would be less likely to undermine the quality assessment due to biological variability.

Stable genes are determined empirically from our data using a stratification of genes according to their degree of temporal variation using the empirical Bayes approach for microarray time course data developed by citetaiS06 and implemented in the R-package `timecourse`. The method required replicates. Not all microarray time course data are replicated, but in some cases, a surrogate for replication can be found. For example, in circadian data, time points with the same relative position in the circle can be identified. Another example would be to identify a minor mutation and a wild type. The justification for this is that only a small number of genes would be affected, which would not significantly affect the results of a quality assessment that is going to be based on a large set of genes. We first determining a subset of temporarily stable genes based on a Hotelling statistics ranking proposed in taiS06. Then, we compute RLE and NUSE distributions for the (hypothetical) microarray defined by a restriction of the

4

original microarray to the subset of the temporarily most stable genes. Such subsets typically contain between 5 and 30% of the genes, but we do not *a priori* specify this further, because the fraction of genes that vary over time depends strongly on the type of experiment. Often, there is not even enough biological knowledge available to provide a reliable estimate for that fraction. Instead of specifying the size of the subset beforehand, we perform the quality assessment based on a number of gene subsets of successively smaller sizes until no additional information can be obtained.

We have now a number of different one parameter summaries per chip: absolute value of Med(RLE), IQR(NUSE) and Med(NUSE). Can compare these, and we can compare them with the GCOS scores. Further, we compare the outcomes of the whole-chip methods with the outcomes of the subset-based assessment.

We compare the different approaches, with an emphasis on the identification of qualitative differences. This includes the observation of patterns and outliers. Using knowledge about both the biological phenomena studied in the experiment and the conditions during the experiment we try shed light on the potential origin of such patterns. Patterns observed in quality assessment may be related to the biological process itself or to experimental conditions. The most burning question whether a patterns is likely to indicate an aspect of data quality or to reflect genuine biological variation. This can in turn be used to compare and evaluate the quality assessment methods. In our study of the subset-based quality assessment we have a particular interest in understanding how relevant the size of the subset is.

4. EXPERIMENTAL DATA SETS

The drosophila developmental series data was created by Tomancak *et al.* (2002) and can be obtained from `www.fruitfly.org/cgi-bin/ex/insitu.pl`. Wild type drosophila (Canton S) were split into 12 population cages and allowed to lay eggs after aged for 3 days. Embryo samples were collected every hour for consecutive 12 hours starting at 30 minutes. Embryonic stages were examined by morphological markers. The same procedure was conducted on 3 days yielding 3 biological time course replicates (A, B, C). RNA was extracted, labeled and hybridized to short oligonucleotide microarrays (Affymetrix GeneChip DrosGenom 1.1) measuring 14,010 genes. Series A and C series were hybridized on a day other than the one for series B. Each chip was named as the combination of the replicate label and the time point label 01, 02,...,12.

The arabidopsis circadian rhythm data was created by Edwards et al. (2006) and is available at `http://millar.bio.ed.ac.uk/bibliogr.htm`. Wild type Col-0 seedlings were sterilized and grown over eight days under 12-hour-light/12-hour-dark cycles and transferred to constant light at 22°C. Circadian rhythms were measured by video imaging of leaf movement under constant light. Samples were taken over a period of a little more than two circadian cycles (52 hours) at 4-hour intervals beginning at hour 26 after the last dark-light transition. RNA was extracted, labeled and hybridized to short oligonucleotide microarrays (Affymetrix GeneChip ATH1) measuring 22,764 genes. Chips are named 'a' (for hour 26) through 'm' (for hour 74).

5. RESULTS

Standard approaches

The left side of Figure 1 shows boxplots of the distributions of the RLE values or all chips in the data set. There is a substantial amount of variation and there are no isolated outliers. The largest deviation of Med(RLE) from 0 (absolute value larger than 0.05) is observed in chips A1, B10, B2, C3, A9 and A2. Overall, it deviates more in chips from series B than in chips from series A and C. The chips with the widest IQR(RLE) (bigger than 0.4) are A1, A12, B1, B2 and B10. Overall, it is wider in chips from series B than in chips from series A and C. The left side of Figure 2 shows boxplots of the distributions of the NUSE values for all chips in the data set. There is a substantial amount of variation and there are no isolated outliers. The largest values for Med(NUSE) (larger than 0.01) are observed in chips B1, B10, B4, B5, A12, B11, B2, B12, B3, and A1. Generally, the values are larger in chips from series B than in chips from A and C.

Figure 3 shows that the four one-parameter summaries of the RLE and NUSE distributions reveal similar trends. Med(NUSE) and IQR(NUSE) are very strongly correlated. IQR(RLE) has a good correlation with both NUSE-based parameters. There is also a correlation between |Med(RLE)| and the NUSE-based parameters for series A and C, whereas this can not be said about series B.

In summary, all RLE- and NUSE-based measures indicate (i) lower quality in series B than in series A and C and (ii) lower quality at the beginning and at the end of each series than during the middle stages. We think that the observation (i) is indeed due to quality differences. The hybridization date of series B is different from the hybridization date of series A and C. Note that the data has been quantile normalized before the quality assessment, which should take care of the most drastic apparatus bias between the three different series. However, the remaining quality differences between the series are most likely to be related to different conditions during the hybridization or to differences in sample properties and handling. In contrast, we think that observation (ii) is not due quality difference, but to biological variation. There is often an increased genetic activity at the beginning and at the end of a developmental processes.

A comparison between the RLE- and NUSE-based quality measures with some of the GCOS quality scores is shown in the pair plots in Figures 4 and 4. The lowest Percent Present values are obtained for chips B9, B5, B6 and B4. Within the series A, the lowest values are obtained for chips A9, A5 and A6, and within the series C, the lowest values are obtained for chips C4, C5, C9, C3 and C8. These are all chips from the middle of the series. So, Percent Present behaves very differently from the RLE- and NUSE-based measures that typically show lower quality at the beginning and at the end of each series. Percent Present is generally lower for replicate series B than for A and C, which is consistent with the results observed for RLE and NUSE. However, there is hardly any linear relationship between Percent present and IQR(RLE), and it can be argued that the weak linear relationship between Percent present and Med(NUSE) is only a product of ecological correlation from the groups defined by the replicate series A, B and C. There is no linear relationship between any of the GCOS quality scores and

6

IQR(RLE) or Med(NUSE). There is no linear relationship between any two of the GCOS quality scores either. However, for series B only there is some correlation between Percent Present and Scale Factor, and for series C only there is some correlation between Scale Factor and Average Background.

Subset-based approach

In the time course analysis for the determination of temporarily stable genes we use triplicates for each stage provided by series A, B and C.

Figure 6 shows the behavior of RLE-distributions in response to restrictions of the gene set. The effect is remarkable. Recall that many of the boxplots of the RLE-distributions based on all genes were off-center (see Figure 1). For the subset of the 7,000 most temporarily stable genes, the boxplots are now well centered around 0 except for a little deviation in the case of chip A12. The latter also disappears when the set is further restricted to 2,000 genes. The restriction to subsets results in tighter RLE-distributions. The IQR(RLE) shrink to about half their sizes in the 7,000-gene-subset and even smaller in the 2,000-gene-subset. The difference between series B on the one hand and series A and C on the other hand are less obvious, but they remain. The differences between the early/late stages and the middle stages are far less pronounced than they were for the IQR(RLE) based on the whole set of genes. However, the relative positions of the chips remain roughly the same.

Figure 6 shows the behavior of NUSE-distributions in response to restrictions of the gene set. The distributions get slightly tighter, but beyond that the restrictions have nearly no effect. The relative positions of the chips with respect to Med(NUSE) and IQR(NUSE) remain basically the same.

RLE and NUSE react very differently to the restriction of the quality assessment to the subarrays defined by genes with low temporal variation. It makes almost no difference whether we choose a large subset of 7,000 genes or a smaller subset of 2,000 genes.

<center>CIRCADIAN RHYTHM DATA</center>

Standard approaches

The right side of Figure 1 shows boxplots of the distributions of the RLE values or all chips in the data set. They indicate fairly homogeneous quality. Med(RLE) is very close to 0 for all chips except the first one. IQR(RLE) varies a bit more with the largest values (bigger than 0.1) observed in chips h, c, a and b. The right side of Figure 2 shows boxplots of the distributions of the NUSE values for all chips in the data set. Med(NUSE) is very close to 1 in all chips except for somewhat elevated values in chips h, l and b. Figure 6 shows a linear association between IQR(RLE) and Med(NUSE).

In summary, the quality of the data seems quite homogeneous, and we canot detect any patterns related to experimental conditions or sample properties. We attribute the (small) existing differences to real quality issues rather than biological variation.

A comparison between the RLE- and NUSE-based quality measures with some of the GCOS quality scores is shown in the pair plots in Figure 6. There is a strong linear association between Scale Factor and Percent Absent. Both of them have a weak linear association with Average Background. The latter has a linear association with IQR(RLE) and, a bit weaker, with Med(NUSE). GAPDH 3'/5' does not show

<center>7</center>

a linear association with any of the GCOS scores nor with IQR(RLE) and only a weak and somewhat ambiguous association Med(NUSE).

Subset-based approach

In order to obtain a data set containing chips from exactly two successive circadian cycles we could drop either the first or the last chip from our data set. We decided to drop the first one (chip a), because it is an outlier with respect to Med(NUSE). The rest of the analysis and discussion refers to the data from the remaining set (chips b to m). In the time course analysis for the determination of temporarily stable genes any two chips from the same circadian stage were treated as replicates (b=h, c=i and so on).

Figure 6 shows the behavior of RLE-distributions in response to restrictions of the gene set. The boxplots of the RLE-values of chips b to m based on all genes are well centered around 0. The restriction to the sets of temporarily stable genes does not change that. The RLE-distributions are getting a bit tighter with each restriction. The rank order of the chips with respect to the IQR(RLE) remains unchanged.

Figure 6 shows the behavior of NUSE-distributions in response to restrictions of the gene set. The distributions get slightly tighter, but beyond that the restrictions have nearly no effect. The relative positions of the chips with respect to Med(NUSE) and IQR(NUSE) remain basically the same, except that the two outliers, chip m and chip l, swap their positions.

Both RLE- and NUSE-based quality measures are little affected by the restriction to temporarily stable genes.

6. DISCUSSION

Standard microarray data quality assessment based on RLE- and NUSE-distributions from whole microarrays is potentially problematic in the case of time course data, because the fraction of active genes as well as the magnitude of their activity may vary over time. Apparent quality differences may turn out to be driven by the biological process of interest. Stratification of genes by a statistical time course analysis sheds some light on the issue. We have suggested and tested an alternative quality assessment based on subarrays defined by restriction of the original microarray to subsets of temporarily stable genes.

In this paper, we use the time course analysis proposed by Tai and Speed (2006) to select the genes with low temporal variation. However, our methods could also be based on a different time course analysis method suited to determine such gene subsets.

We showed on a developmental series that the subset-based method can help minimizing this confusion in the case of the RLE-baset quality measures. The NUSE, however, turned out to be very little effected by restrictions to subarrays. In our second data example, a circadian rhythm series, the problem did not occur. However, a restriction to a subarray defined by temporarily stable genes resulted in the same quality assessment as for the whole microarray. The size of the subset, within reasonable bounds, was not relevant for the results in either data set.

This inconsistency between NUSE and RLE is remarkable. Usually, despite their different construction, they give very similar results in quality assessment (see Brettschneider et al. (2008)). Inconsistency

8

between RLE- and NUSE-based measures and the GCOS quality scores, however, is not surprising. In particular, we do not expect the time structure to play any role in such a quality assessment, because the quality assessment is done for on a one-by-one basis rather than evaluating the quality of each microarrays within the whole series of microarrays.

The developmental data set demonstrates that it is necessary to underpin the "mechanical" application of any quality assessment tools with knowledge about the experiment to reveal the whole quality story, especially in a situation as complex as time course data. The circadian rhythm data set shows that the alterations we suggested to make to the standard microarray data quality assessment do not have to have any negative effects even when not necessary. It would be desirable to test the subset-based approach on additional data sets and to use it also in combination with microarray time course analysis methods other than Tai and Speed (2006). The use of any such subset-based approaches, with careful monitoring of the size of the subsets, is a strategy that potentially refines the quality assessment for any microarray time course data.
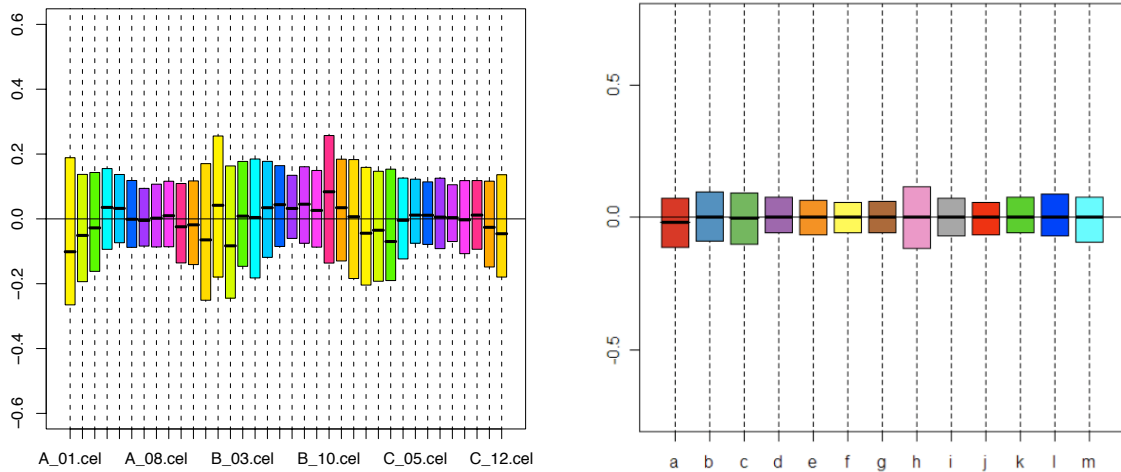
Figure 1. Relative Log Expression (RLE) for drosophila embryo developmental series (left) and for arabidopsis circadian rhythm series (right). Each boxplots shows the distribution of the RLE values of one microarray. The drosophila microarrays are ordered by series (A, B, C), and by stage (1,2,...,12) within each series. The arabidopsis microarrays are ordered chronologically .
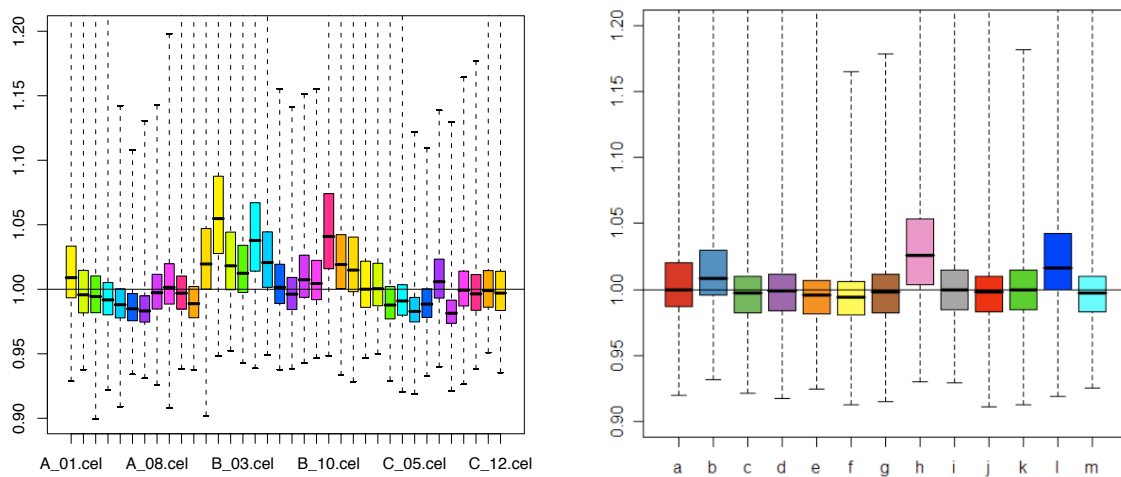

Figure 2. Normalized unscaled standard error (NUSE) for drosophila embryo developmental series (left) and for arabidopsis circadian rhythm series (right). Each boxplots shows the distribution of the NUSE values of one microarray. The order of the microarrays is the same as in Figure 1.
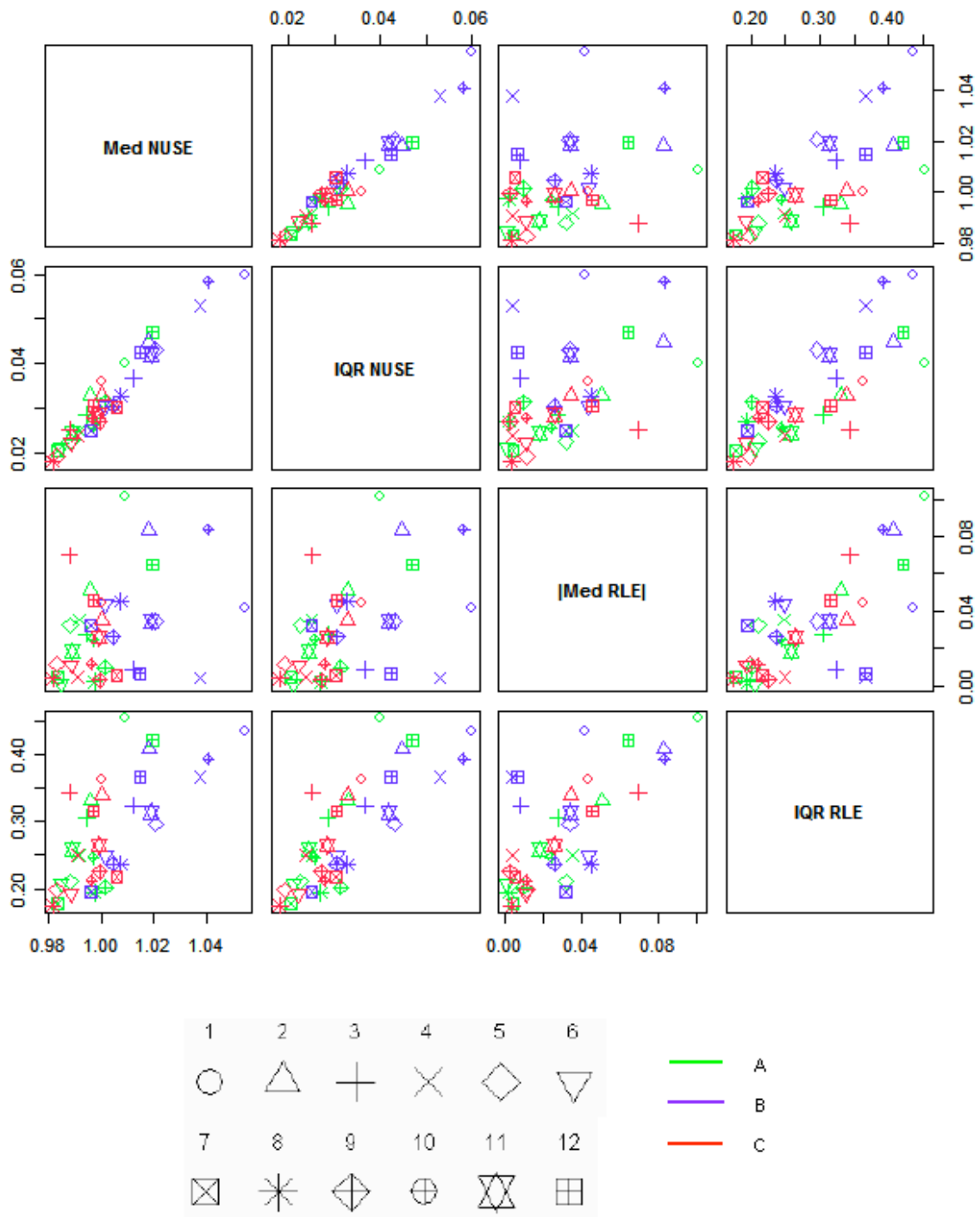
Figure 3. Relationship between RLE- and NUSE-based one-parameter summaries for the drosophila embryo development data .
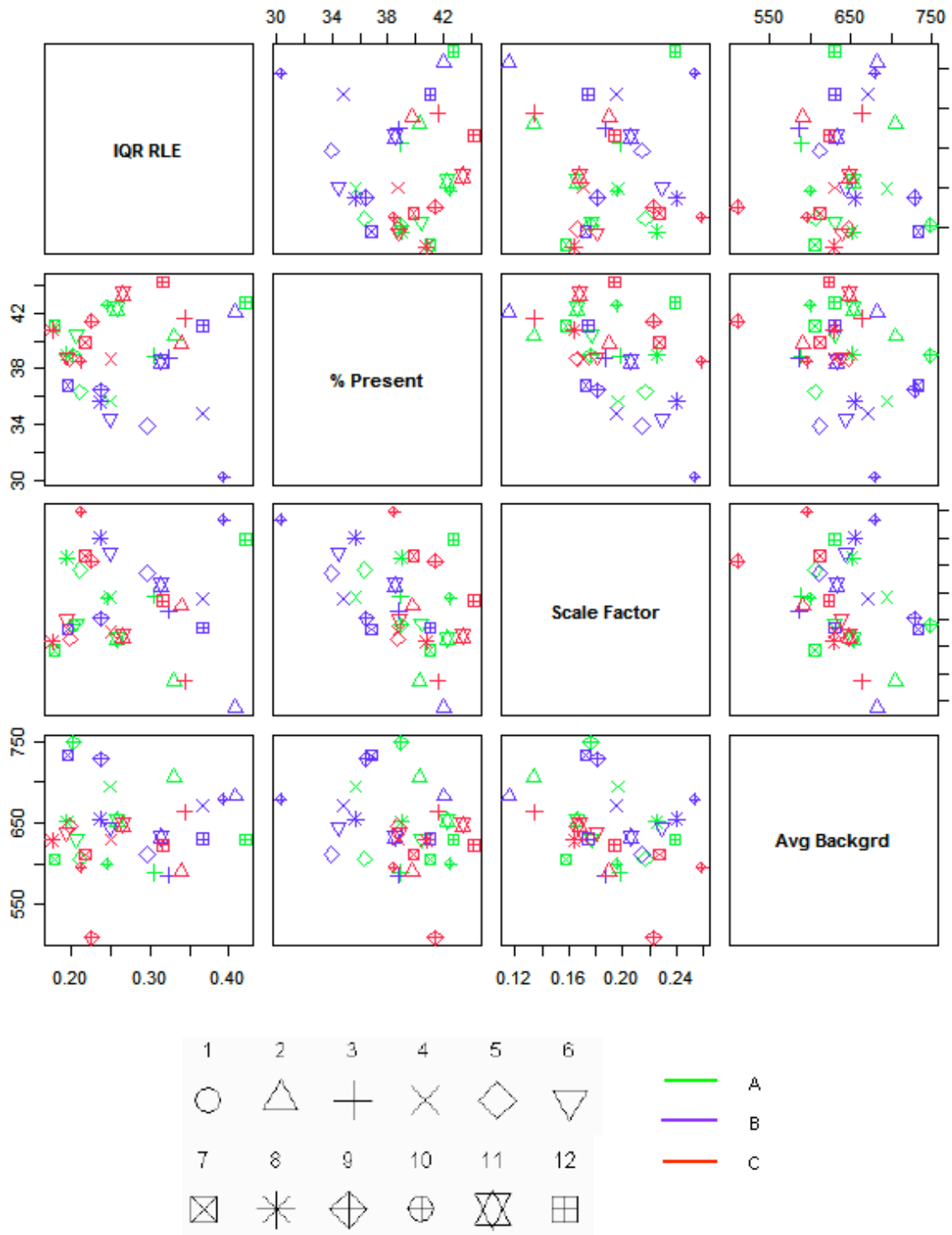
Figure 4. Relationship between IQR(RLE) and GCOS quality report scores for the drosophila embryo development data.
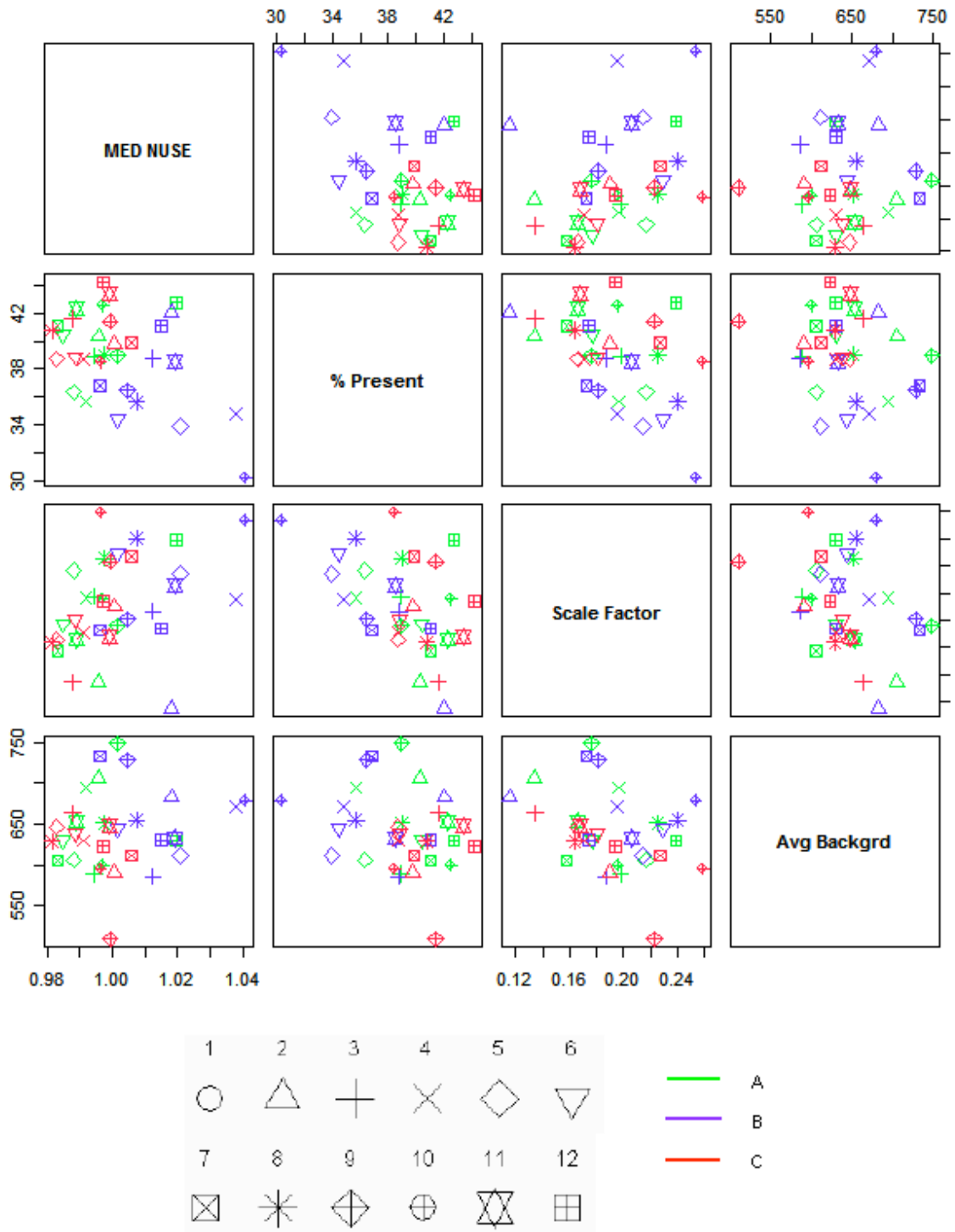
Figure 5. Relationship between Med(NUSE) and GCOS quality report scores for the drosophila embryo development data.
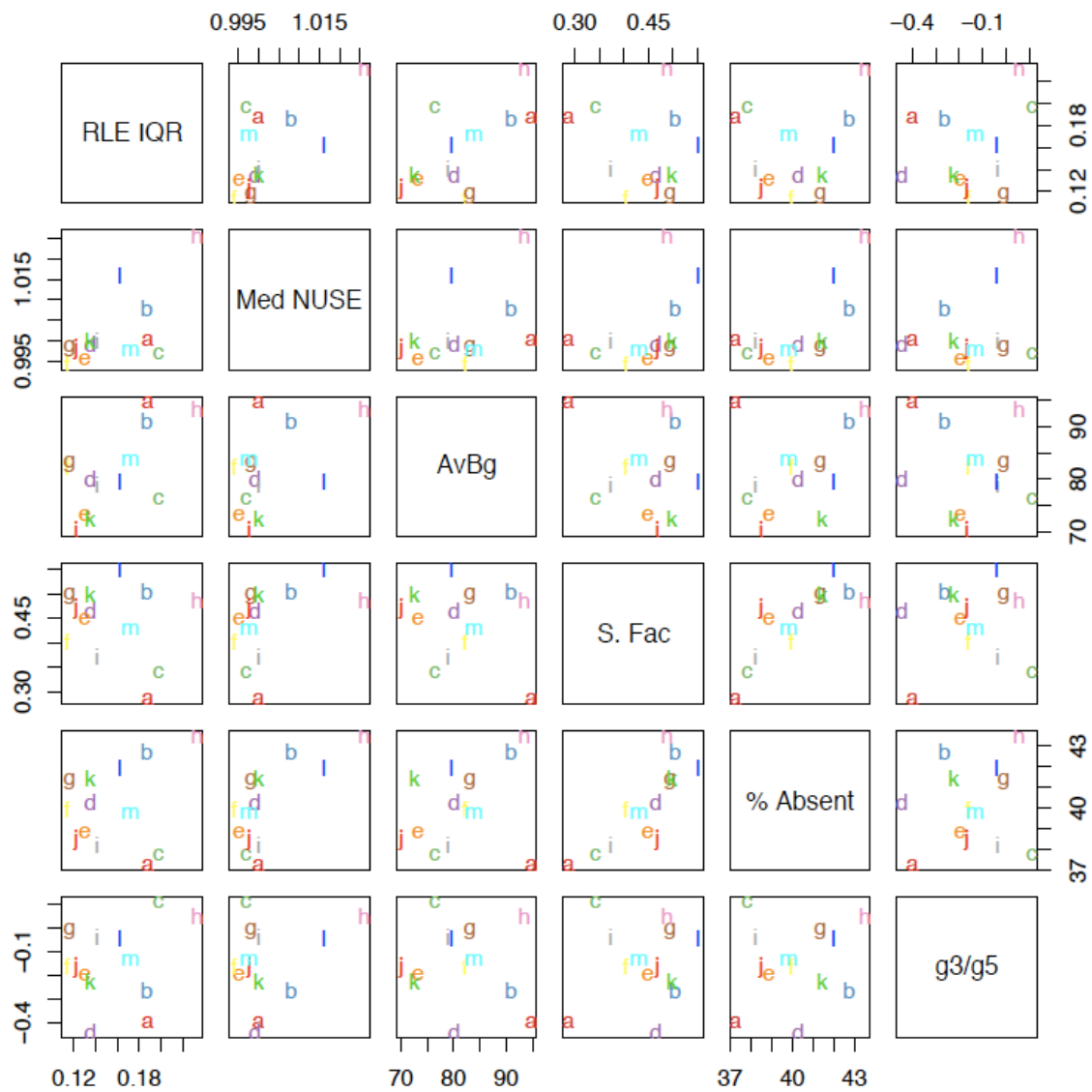
13

Figure 6. Relationship between IQR(RLE), Med(NUSE) and GCOS quality report scores for the arabidopsis circadian rhythm data. The microarrays are ordered as in Figure 1 .
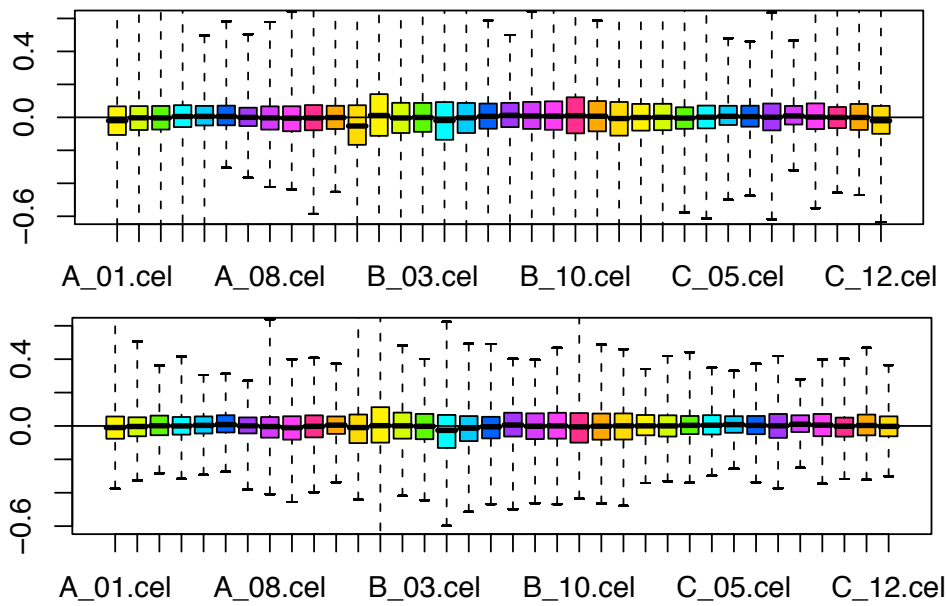
CRiSM Paper No. 09-24, www.warwick.ac.uk/go/crism

Figure 7. Relative Log Expression (RLE) for drosophila embryo developmental series for the 7000 least changing genes (first row) and for the 2000 least changing genes (second row). The microarrays are ordered as in Figure 1 .
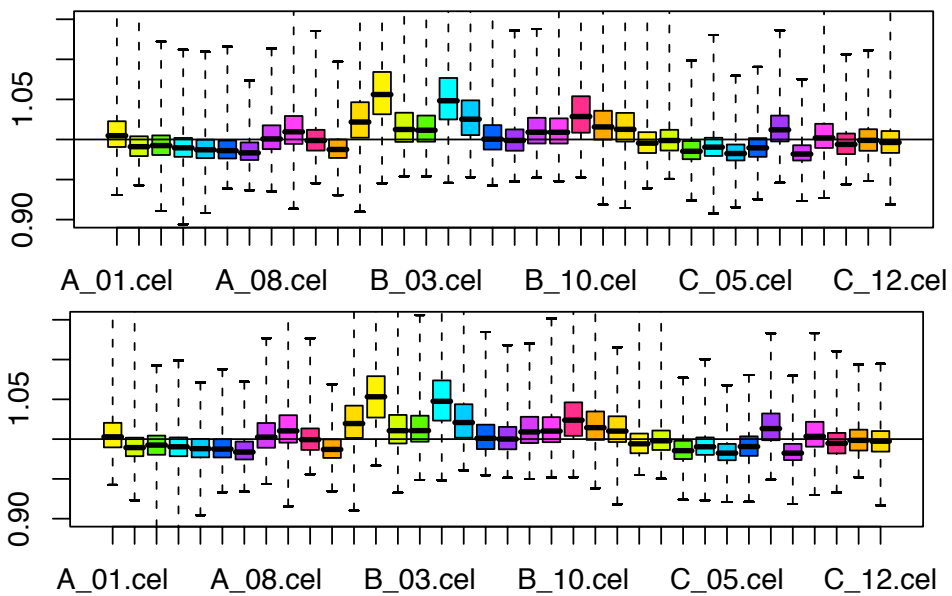

Figure 8. Normalized unscaled standard error (NUSE) for drosophila embryo developmental series for the 7000 least changing genes (first row) and for the 2000 least changing genes (second row). The microarrays are ordered as in Figure 1.
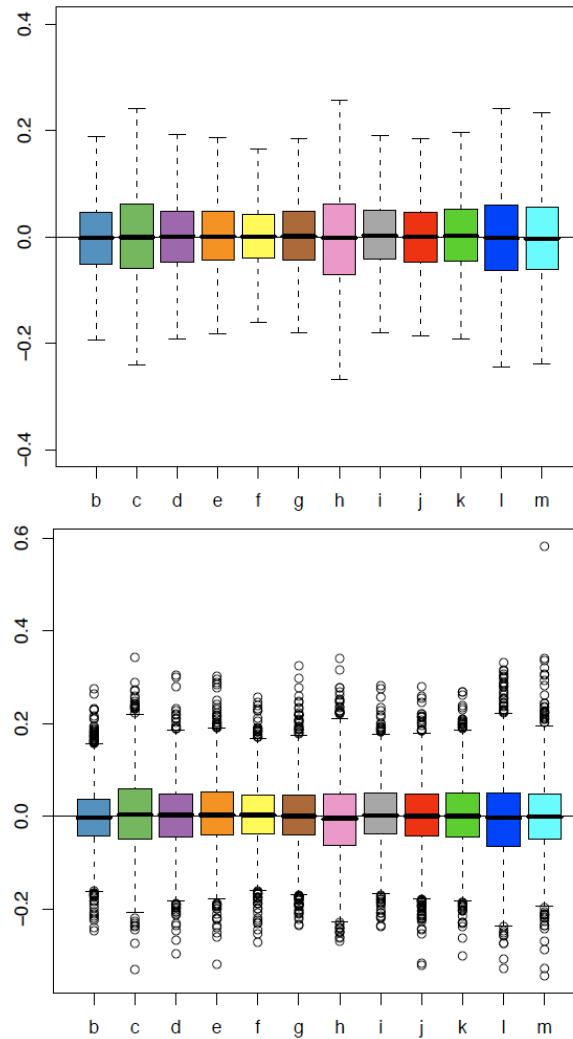
15

Figure 9. Relative Log Expression (RLE) for arabidopsis circadian rhythm series for the 7000 least changing genes (first row) and for the 2000 least changing genes (second row). The microarrays are ordered as in Figure 1..

Figure 10. Normalized unscaled standard error (NUSE) for arabidopsis circadian rhythm series for the 7000 least changing genes (first row) and for the 2000 least changing genes (second row). The microarrays are ordered as in Figure 1.
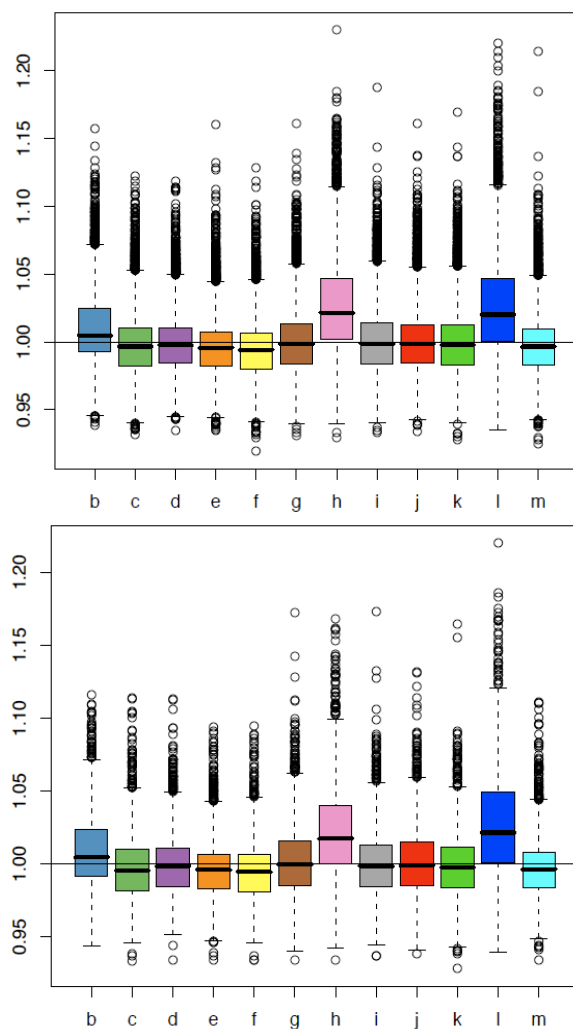
CRiSM Paper No. 09-24, www.warwick.ac.uk/go/crism

REFERENCES

Affymetrix (2001), *Guidelines for assessing data quality*, Affymetrix Inc, Santa Clara, CA.

GCOS (2004), *GeneChip Expression Analysis – Data Analysis Fundamentals*, Affymetrix, Inc, Santa Clara, CA.

Bolstad, B., Collin, F., Brettschneider, J., Simpson, K., Cope, L., Irizarry, R., and Speed, T. (2005), Quality Assessment of Affymetrix GeneChip Data, in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, eds. Gentleman, R., Carey, V., Huber, W., Irizarry, R., and Dudoit, S., New York: Springer, Statistics for Biology and Health, pp. 33–48.

Brettschneider J., Collin F., Bolstad B., and Speed T. (2008), Quality assessment for short oligonucleotide arrays (with 5 commentaries and rejoinder), in *Technometrics*, 50(3): 241-264 (article), 279-283 (rejoinder).

Guan S., Zheng J., and Brettschneider J. (2007) Microarray data quality assessment for developmental time series, in *LASR Proceedings*, eds. Barber S., Baxter P., and Mardia K., Systems Biology Statistical Bioinformatics, Leeds University Press.

Edwards, K., Anderson P., Hall A., Salathia N., Locke J., Lynne J., Straume M., Smith J., Millar A. (2006), Flowering locus C mediates natural variation in the high temperature response of the Arabidopsis circadian clock, Plant Cell, 18: 639-650.

Hwang, K., Kong, S., Greenberg, S., and Park, P. (2004), Combining gene expression data from different generations of oligonucleotide arrays, *BMC Bioinformatics*, 5, e159.

Irizarry, R., Bolstad, B., Collin, F., Cope, L., Hobbs, B., and Speed, T. (2003), Summaries of Affymetrix GeneChip probe level data, *Nucleic Acids Res*, 31, e15.

Shewhart, W. (1939), *Statistical Method from the Viewpoint of Quality Control*, Lanceser, Pennsylvania: Lancester Press, Inc.

Tai, Y. and Speed, T. (2006), A multivariate empirical Bayes statstic for replicated microarray time course data, *Ann Statist*, 34, 2387–2412.

Wilson, C. and Miller, C. (2005), Simpleaffy: a BioConductor package for Affymetrix quality control and data analysis, *Bioinformatics*, 21, 3683–3685.