

Microarray Gene Expression: The Effects of Varying Certain Measurement Conditions

Walter Liggett, NIST

Jean Lozach, Illumina

Anne Bergstrom Lucas, Agilent

Ron L. Peterson, Consultant

Marc L. Salit, NIST

Danielle Thierry-Mieg, NCBI

Jean Thierry-Mieg, NCBI

Russ Wolfinger, SAS

Abstract

This chapter explores measurements from an experiment with a batch effect induced by switching the mass of RNA analyzed from 400 ng to 200 ng. The experiment has as additional factors the RNA material (liver, kidney, and two mixtures) and the RNA source (six different animals). We show that normalization can partially correct the batch effect, and that, after normalization, the size of the batch effect is comparable to the animal-to-animal variation. In addition, we present data analysis results that suggest other batch effects.

1. Introduction

In metrological terminology, conditions of measurement can be held constant, or certain conditions can be allowed to vary (ISO, 2007). Measurements made under constant conditions (also referred to as repeatability conditions) constitute a batch. A set of measurements may include several batches that are delineated by changes in certain conditions. Variations observed among the batches beyond what is observed under constant conditions are batch effects. In this paper, we illustrate batch effects in microarray measurements by examining variations related to two measurement conditions. The measurements were made with the Agilent Whole Rat Genome Microarray 4x44K. The two measurement conditions allowed to vary are the amplification and labeling input mass and the substrate on which four microarrays are mounted.

In contrast to the familiar univariate metrology, batch effects in microarray gene expression have many facets. The Agilent microarray considered here has 45018 probes, and, consequently, technical variation manifests itself in 45018 dimensions. For this reason, there are many choices to be made in the characterization of microarray batch effects. On one hand, one does not want to present a characterization that is incomprehensible because it is overwhelming in its detail. On the other hand, one does not want to hide manifestations of the batch effects that might be helpful in improving measurement protocols or methods for statistical inference. The purpose of this paper is to present a characterization that strikes a balance. We illustrate the various ways that batch effects show up in measurements and describe methods for analyzing batch effects. These ideas can be applied in other situations.

The structure of the experiment considered here provides many possibilities for data analysis. There are measurements on liver RNA and kidney RNA, which are very different, and on mixtures of these RNAs to allow insight into calibration-curve linearity.

Measurements were made on RNAs from six animals (*Rattus norvegicus*). Because these animals formed a control group for a previous study, their RNAs were expected to be similar. In addition, technical replicates are included. Guo, et al. (2006) performed an experiment with six animals as biological replicates but without mixtures or technical replicates.

The data set considered here consists of 96 microarray measurements. Seventy-two of these measurements were made on RNAs from six animals (*Rattus norvegicus*), twelve measurements from each animal. RNA was extracted from both the liver and the kidney of each animal. In addition to these RNAs, mixtures of the liver and kidney RNAs were prepared, one with 3 parts liver to 1 part kidney and the other with 1 part liver to 3 parts kidney. The four materials from each animal were each measured in triplicate. In addition to the individual animal materials, the liver RNAs from the six animals were pooled; the kidney RNAs were similarly pooled; and mixtures of these pools were prepared. These four materials were each measured in quadruplicate. Finally, the liver RNAs from the first three animals were pooled; the kidney RNAs were similarly pooled; mixtures were prepared; and the four materials were measured without replication. Moreover, the RNAs from the second three animals were pooled, mixed and measured. Affymetrix measurements of the same 96 materials have been reported as a preprint (Liggett et al., 2008).

This paper is organized as successive steps in uncovering the input mass effects, which are more pronounced than the substrate effects in our experiment. It should be noted that the two levels of input mass are completely confounded with choice of operator and with some other potential causes of batch effects. As is typical in microarray data analysis, we begin with normalization (Stafford, 2008). Our normalization method makes use of the mixture relations among the RNAs, which has as one result a partial correction of batch effects (Liggett, 2008). For each probe, we model the normalized intensities with a linear mixed model (Pinheiro and Bates, 2000). Such a model involves fixed effects such as the animal effects and the input mass effects, random effects such as the substrate effect, and the random error. Finally, we consider evidence of unrecorded changes in measurement conditions.

2. Input mass effect on the amount of normalization applied

In this section and the next, we present an analysis of the measurements on the individual animals with the Agilent platform (ArrayExpress E-TABM-555). These measurements consist of results from 72 arrays. Our analysis is based on background corrected intensities given by `gBGSubSignal` divided by `gMultDetrendSignal`, values of which can be found in the ArrayExpress data files. We eliminate intensities from control probes (`ControlType`) and from probes that exhibit feature nonuniformity (`glsFeatNonunifOL`) or saturation (`glsSaturated`) for any of the 72 individual-animal arrays. Thus, for each array, intensities from 42697 probes enter our analysis.

Generally, the observed intensities from a group of arrays exhibit inter-array differences that must be dealt with either by a preprocessing step referred to as normalization or by some other means. We normalize the 12 arrays from each individual animal separately

using a two-step process (Liggett, 2008). The first step is array-by-array global normalization of the intensities. For each array, we compute a scale factor by first adding 30 to each background-corrected intensity because some intensities are reported as negative and then by finding the geometric mean of these values. We then divide each intensity by this scale factor. The second step is adjustment of the globally normalized intensities so that the linear model implied by the mixing of the materials fits better. Letting the globally normalized intensity for array i of animal j and for probe g be y_{jig} , the adjusted intensity is $(y_{jig} - \eta_{0ji})/\eta_{ji}$. The values of η_{0ji} and η_{ji} are computed to improve the fit of the adjusted intensities to the linear model specified by the mixing.

In addition the estimates of η_{0ji} and η_{ji} , our two-step normalization process gives a fitted model for the intensities. As implied by the mixing, the intensities can be modeled as $x_{Aji}\theta_{A_jg} + x_{Dji}\theta_{D_jg}$. The fraction of the liver material in the mixture x_{Aji} is 1 for material A, $3/(3 + \varphi_j)$ for material B, $1/(1 + 3\varphi_j)$ for material C, and 0 for material D. The symbol φ_j denotes the ratio of the concentration of mRNA in the kidney RNA to the concentration of mRNA in the liver RNA. We have $x_{Dji} = 1 - x_{Aji}$. Liggett, et al. (2008) describe estimation of φ_j in the second step of the normalization. We use the estimates of φ_j , θ_{A_jg} , and θ_{D_jg} in specifying the analysis in the next section.

Figure 1 shows the estimated normalization parameters for each of the six animals. The points are labeled with the user designation A or B, which corresponds to input mass of 200 ng and 400 ng. We see that the second step of our normalization process generally scales the 200 ng arrays up in comparison to the 400 ng arrays. This can be viewed as our normalization providing a partial correction of the input mass batch effect.

3. Probe-by-probe modeling of the input mass effect

For each probe, there are 72 normalized intensities that are differentiated by the animal, the material (liver, kidney or a mixture), the input mass, and the substrate. By fitting a linear mixed model (Pinheiro and Bates, 2000) to these intensities, we can distinguish the contributions of these factors. Of particular interest is comparison of the animal-to-animal variation with the input mass effect. In case-control studies, the central statistical question is whether the amount of animal variation within the control group can account for the apparent differences between the cases and the controls. If the input mass effect is of the same size or larger than the amount of animal variation, then this effect has the potential for misleading researchers in the formulation of study conclusions.

It is, of course, not enough to compare animal variation with input mass effect for each probe. Summarization over the probes is necessary in the understanding of the input mass effect. In our approach, we confine the summarization to a subset of the probes and choose a model parameterization with coefficients that are comparable from one probe to another. One aspect of selecting the probe subset is choice of probes for which there is appreciable liver mRNA or appreciable kidney mRNA. Another aspect is selecting a probe subset for which there is a model parameterization appropriate for summarization.

In our approach, the model parameterization is based on representing the animal-to-animal variation and the input mass effect in terms of fractional deviations from the animal average intensities. Such an approach has the advantage of making the probes with strongest response comparable regardless of their liver-kidney difference or overall intensity response. The average intensities are obtained from the intensity estimates obtained in the normalization process. These are animal-by-animal estimates of the liver intensity $\hat{\theta}_{A_{jg}}$ and the kidney intensity $\hat{\theta}_{D_{jg}}$. We denote the average of the liver intensities by $\bar{\theta}_{A_g}$ and the average of the kidney intensities by $\bar{\theta}_{D_g}$. For array i of animal j , the animal average intensity is given by

$$\hat{x}_{A_{ji}}\bar{\theta}_{A_g} + \hat{x}_{D_{ji}}\bar{\theta}_{D_g} = (\bar{\theta}_{A_g} + \bar{\theta}_{D_g})/2 + (\hat{x}_{A_{ji}} - \hat{x}_{D_{ji}})(\bar{\theta}_{A_g} - \bar{\theta}_{D_g})/2.$$

In our notation, we apply a “hat” to $\hat{x}_{A_{ji}}$ and $\hat{x}_{D_{ji}}$ because they incorporate the estimate of φ_j obtained as part of the normalization. In modeling, we parameterize the animal effects with an intercept term $\alpha_{jg}(\bar{\theta}_{A_g} + \bar{\theta}_{D_g})/2$ and with a slope term $\beta_{jg}(\hat{x}_{A_{ji}} - \hat{x}_{D_{ji}})(\bar{\theta}_{A_g} - \bar{\theta}_{D_g})/2$. Similarly, we parameterize the input mass effect with the terms $\gamma_g(\bar{\theta}_{A_g} + \bar{\theta}_{D_g})/2$ and $\delta_g(\hat{x}_{A_{ji}} - \hat{x}_{D_{ji}})(\bar{\theta}_{A_g} - \bar{\theta}_{D_g})/2$.

The probes we select for summarization satisfy two conditions. First, the probes satisfy

$$\left((\bar{\theta}_{A_g} > 3) \text{ AND } (\bar{\theta}_{D_g} > 0) \right) \text{ OR } \left((\bar{\theta}_{A_g} > 0) \text{ AND } (\bar{\theta}_{D_g} > 3) \right).$$

In general terms, this condition amounts to the requirement that either the liver mRNA or the kidney mRNA have appreciable intensity. As a way of gauging this condition, we note that there are 9620 probes for which $\bar{\theta}_{A_g} > 3$ and there are 11107 probes for which $\bar{\theta}_{D_g} > 3$. Second, the probes must exhibit a fold change of at least 2 between the liver and the kidney. In other words, the probes satisfy

$$\left(\left(\frac{\bar{\theta}_{A_g}}{\bar{\theta}_{D_g}} \right) > 2 \right) \text{ OR } \left(\left(\frac{\bar{\theta}_{D_g}}{\bar{\theta}_{A_g}} \right) > 2 \right).$$

These two conditions assure that the model parameterization we have adopted is not misleading.

With the substrate effect left out, the modeling equation for the normalized intensity u_{jig} corresponding to array i for animal j is given by

$$u_{jig} = (\alpha_{jg} + \gamma_{ig})(\bar{\theta}_{A_g} + \bar{\theta}_{D_g})/2 + (\beta_{jg} + \delta_{ig})(\hat{x}_{A_{ji}} - \hat{x}_{D_{ji}})(\bar{\theta}_{A_g} - \bar{\theta}_{D_g})/2 + \varepsilon_{jig},$$

where

$$\text{var}(\varepsilon_{jig}) = \sigma_g^2 \left[\left(\max(\hat{x}_{Aji} \hat{\theta}_{Aig} + \hat{x}_{Dji} \hat{\theta}_{Dig}, 0) \right)^2 + \omega^2 \right].$$

In this modeling equation, the unknown parameters are α_{jg} , β_{jg} , γ_{ig} , δ_{ig} and σ_g^2 . The other quantities are obtained from the normalization results. The parameters γ_{ig} and δ_{ig} are restricted to $\pm \gamma_g$ and $\pm \delta_g$ because the input mass has only two levels.

In Figure 2, the input mass effect is shown as a plot of the intercept coefficient γ_g versus the slope coefficient δ_g for the selected probes. The points lie close to the $x = y$ line. This means that most of the input mass effect shows up as a fractional deviation from the average intensity $\delta_g (\hat{x}_{Aji} \bar{\theta}_{Ag} + \hat{x}_{Dji} \bar{\theta}_{Dg})$, that is, as a multiplicative effect on the average intensity. Relative to this, the remainder of the input mass effect $(\gamma_g - \delta_g) (\bar{\theta}_{Ag} + \bar{\theta}_{Dg}) / 2$ is small. This leads us to thinking about the modeling results in terms of the fractional deviation from the intensity δ_g and the remaining fractional deviation $\gamma_g - \delta_g$.

Consider comparison of the input mass effect with the animal variation. On the basis of what is shown in Figure 2, we begin by comparing $\delta_g (\hat{x}_{Aji} \bar{\theta}_{Ag} + \hat{x}_{Dji} \bar{\theta}_{Dg})$ with the animal-by-animal changes in $\beta_{jg} (\hat{x}_{Aji} \bar{\theta}_{Ag} + \hat{x}_{Dji} \bar{\theta}_{Dg})$. To make this comparison, we compute, for each probe, the standard deviation of β_{jg} , $j = 1, \dots, 6$, the fractional deviations of the animal intensities. For the group of probes selected, the top two panels of Figure 3 show histograms for the fractional deviations from the average intensities. The top panel shows the input mass effect. The second panel shows the standard deviation of the six animal values. We see that these two histograms have roughly the same spread. To be more exact in the comparison of the two histograms, we must account for the fact that the animal variation is given as a single standard deviation although two standard deviations would come closer to covering all the animal variation. Moreover, the probe-to-probe variation in the animal standard deviation also involves sampling error. In other words, the standard deviations shown are spread out in part by the estimation error inherent in a sample of only six animals and in part by actual probe-to-probe variation in the standard deviation. A more exact comparison seems to require that we introduce a more definite context.

There is more to the animal variation than $\beta_{jg} (\hat{x}_{Aji} \bar{\theta}_{Ag} + \hat{x}_{Dji} \bar{\theta}_{Dg})$, the fractional deviation from the average intensity. There is $(\alpha_{jg} - \beta_{jg}) (\bar{\theta}_{Ag} + \bar{\theta}_{Dg}) / 2$. The bottom panel of Figure 3 shows a histogram of the standard deviation of $\alpha_{jg} - \beta_{jg}$, the remaining fractional deviation. Comparison of the second and third panels of Figure 3 shows that the remaining fractional deviation is generally smaller.

That the microarrays used in this experiment come from Agilent mounted four on a common substrate might be expected to introduce another batch effect. We model the substrate effect as a random effect given by $\zeta(\hat{x}_{Aji}\bar{\theta}_{Ag} + \hat{x}_{Dji}\bar{\theta}_{Dg})$, where ζ varies randomly from the arrays on one substrate to the arrays on another. We can perform a likelihood ratio test of the null hypothesis that there is no substrate effect. Were exact p values for this test available, we could combine the results from a set of probes using false discovery rate methodology (Storey and Tibshirani, 2003). Because this is not the case (Pinheiro and Bates, 2000), the current situation is somewhat more complicated.

Figure 4 shows the estimate of the standard deviation of ζ plotted versus the likelihood ratio test statistic for the 5014 probes selected. Values of the test statistic below 0.001 have been replaced by 0.001. Points for most of the probes lie in the lower right corner where the test statistic and the standard deviation are both nearly 0. We see that few of the probes have substrate-effect standard deviations that are comparable to the animal variation shown in Figure 3. Moreover, we see that few of the probes have values for the likelihood ratio test statistic greater than 1.

On the basis of theoretical results, it is generally believed that the likelihood ratio test statistic under the null hypothesis is distributed as a mixture of two distributions, a probability mass at 0 and a chi-square distribution with 1 degree of freedom. The probability mass at 0 can be seen in the probes shown in the lower right of Figure 4. The mixing proportion for the distributions must generally be determined by simulation (Pinheiro and Bates, 2000).

We conclude that the substrate effect should be of little concern. We have derived a mixing proportion from the observed distribution of the likelihood ratio test statistic. The resulting mixture distribution fits the observed distribution closely. Moreover, the random-effect standard deviations are relatively small compared to the animal variation.

4. Further evidence of batch effects

The choice of 200 ng of material for some of the Agilent microarray measurements was not part of the experimental design originally. Rather, two (microarray) users were assigned the task of making the 96 measurements, 72 on individual animal RNAs and the rest on RNAs made from pooling the individual animals RNAs. The user designated A in the ArrayExpress description of the measurements made a mistake in performing the 48 measurements assigned. In remaking these measurements, only 200 ng of material was generally available for each RNA sample. In that the 200 ng measurements were a recovery from a mistake, it would not be surprising if other batch effects showed up. The samples obtained by pooling the liver RNAs of the six animals and the kidney RNAs provide an opportunity to investigate this. From these pools, 3 to 1 and 1 to 3 mixtures were made as with the individual animal RNAs. The pooled materials were then measured in quadruplicate. User A made all these measurements.

Consider using the measurements on the six-animal pools for testing the linearity of the relation between concentration and the intensity. We begin with the same normalization

procedure that we applied to the measurements on the individual animal RNAs. The normalized intensities are given by u_{pig} . Linearity implies that

$$u_{pig} = \hat{x}_{Api} \theta_{Apg} + \hat{x}_{Dpi} \theta_{Dpg} + \varepsilon_{pig},$$

where the variance of ε_{pig} is given by

$$\text{var}(\varepsilon_{pig}) = \sigma_g^2 \left[\left(\max(\hat{x}_{Api} \hat{\theta}_{Apg} + \hat{x}_{Dpi} \hat{\theta}_{Dpg}, 0) \right)^2 + \omega^2 \right].$$

The parameters θ_{Apg} , θ_{Dpg} , and σ_g^2 are unknown, and values for the other parameters are taken from the output of the normalization. If the realizations of ε_{pig} are independent from measurement to measurement, then we can estimate σ_g^2 from the four sets of replicates regardless of the material-to-material relations among the intensities. This estimate provides the denominator for an F test. Otherwise, there are 4 average intensities, one for each material, and a model of the relation among these averages with 2 parameters θ_{Apg} and θ_{Dpg} . This gives another estimate of σ_g^2 . The ratio of the two estimates of σ_g^2 can be used as statistic for a lack of fit test. Under the null hypothesis, the test statistic is an F ratio with 2 and 12 degrees of freedom.

We consider all the probes for which

$$\left((\hat{\theta}_{Apg} > 3) \text{ AND } (\hat{\theta}_{Dpg} > 0) \right) \text{ OR } \left((\hat{\theta}_{Apg} > 0) \text{ AND } (\hat{\theta}_{Dpg} > 3) \right)$$

except for the control probes and those showing non-uniformity or saturation. There are 11530 such probes. Figure 5 shows a quantile-quantile plot for the log of the F ratio. If the null hypothesis were satisfied for every probe, the curve would fall on the $x = y$ line. We see that the curve is above this line and that there is further deviation from this line at the upper end. The deviation at the upper end is evidence that some probes exhibit saturation. Similar behavior has been seen in other expression microarray data. The separation of the curve from the $x = y$ line over the whole range is more puzzling.

That the curves in Figure 5 are above the $x = y$ line could be evidence of an unsuspected batch effect. That the F ratio is too large could be the result of the denominator of the F ratio being too small. Supposedly, the denominator is computed from 4 independent replicate measurements on each material. If there were an unsuspected batch effect, the independence assumption would not be valid. The F test is based on the hypothesis that the observed variation within the sets of replicates accounts for the observed deviation from a linear calibration curve. If there is batch structure that largely coincides with the replicate sets, then this hypothesis will not be true and the denominator of the F ratio will be too small.

An investigation of an unknown batch effect on the basis of the 16 measurements from one probe would have limited possibility. Combining measurements from all the selected probes offers more possibilities. Let the residuals from the material means be denoted r_{pig} . Consider the standardized version of these residuals

$$r_{pig} / \sqrt{\left(\max(\hat{x}_{Api} \hat{\theta}_{Apq} + \hat{x}_{Dpi} \hat{\theta}_{Dpg}, 0)\right)^2 + \omega^2}.$$

Regarding each probe as a replicate, we can compute 16 by 16 covariance matrix from these standardized residuals. This covariance matrix might suggest the form of the unknown batch effect. One way to proceed is to perform a principal components analysis (PCA) on this covariance matrix. PCA suggests that the replicate-to-replicate variation for materials A and B is larger than that for materials C and D. This is consonant with the results shown in Figure 5 although the cause of the unknown batch effect is still not clear.

5. Conclusions

Examination of some particular batch effects suggests aspects of a general approach to dealing with batch effects. Not addressed in this chapter is the initial step of identifying potential batch effects. This chapter started with potential batch effects already identified. These came from the list of effects such as user (operator) effects and interlaboratory effects that are usually considered in metrology. Potential effects are often identified through remeasurement of reference materials such as the materials used in the MicroArray Quality Control (MAQC) project (MAQC Consortium, 2006). Because microarrays provide a multivariate response, clues can also come from single-array quality measures (Brettschneider et al., 2008).

Once potential batch effects have been identified, the methods illustrated in this chapter can be applied. Note that user was identified at the outset and before the 200 ng-400 ng protocol difference arose. The design and data analysis methods of the current study seem appropriate for use in the study of other batch effects. These methods have two advantages. First, the inclusion of liver, kidney, and mixture RNAs allow the batch effects to be portrayed in terms of slopes of calibration curves. Second, the inclusion of several animals allows comparison of biological variation with technical variation.

The batch effect caused by the change in input mass cannot be completely erased through data analysis. This shows that one cannot rely on data analysis to solve problems with batch effects. One solution is to reduce the batch effect through protocol modification. For example, stricter adherence to 400 ng as the input mass, which is the mass Agilent specified in its initial measurement protocol, would have been preferable to relying on the data analysis. Because complete elimination of a batch effect may not be possible, a complimentary solution is choice of experimental design. For example, the data analysis methods presented here were able to separate the animal variation from the user effect because the design specified the operator factor as orthogonal to the animal and material factors. In particular, for the individual animal measurements, one and only one of every three replicate measurements was made by user A with the 200 ng protocol. Were one

user to have measured four animals and the other user, two animals, the measurements would have been much harder to interpret.

Disclaimer

Certain commercial equipment, instruments, or materials are identified in this paper to foster understanding. Such identification does not imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the materials or equipment identified are necessarily the best available for the purpose.

References

- BRETTSCHEIDER, J., BOLSTAD, B. M., COLLIN, F. & SPEED, T. P. (2008) Quality assessment for short oligonucleotide microarray data. *Technometrics*, 50, 241-283.
- GUO, L., LOBENHOFER, E. K., WANG, C., SHIPPY, R., HARRIS, S. C., ZHANG, L., MEI, N., CHEN, T., HERMAN, D., GOODSID, F. M., HURBAN, P., PHILLIPS, K. L., XU, J., DENG, X., SUN, Y. A., TONG, W., DRAGAN, Y. P. & SHI, L. (2006) Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nature Biotechnology*, 24, 1162-1169.
- ISO (2007) *International vocabulary of metrology--Basic and general concepts and associated terms (VIM), ISO/IEC Guide 99:2007*, Geneva, Switzerland, International Organization for Standardization.
- LIGGETT, W. (2008) Technical variation in the modeling of the joint expression of several genes. IN STAFFORD, P. (Ed.) *Methods in microarray normalization*. Boca Raton, FL, CRC press.
- LIGGETT, W., PETERSON, R. & SALIT, M. (2008) Technical vis-a-vis biological variation in gene expression measurements. *Critical Assessment of Microarray Data Analysis (CAMDA)*. Vienna, Austria.
- MAQC CONSORTIUM (2006) The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements. *Nature Biotechnology*, 24, 1151-1161.
- PINHEIRO, J. C. & BATES, D. M. (2000) *Mixed-Effects Models in S and S-PLUS*, New York, Springer Verlag.
- STAFFORD, P. (2008) *Methods in Microarray Normalization*, Boca Raton, FL, CRC Press.
- STOREY, J. D. & TIBSHIRANI, R. (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, 100, 9440-9445.

Figure 1

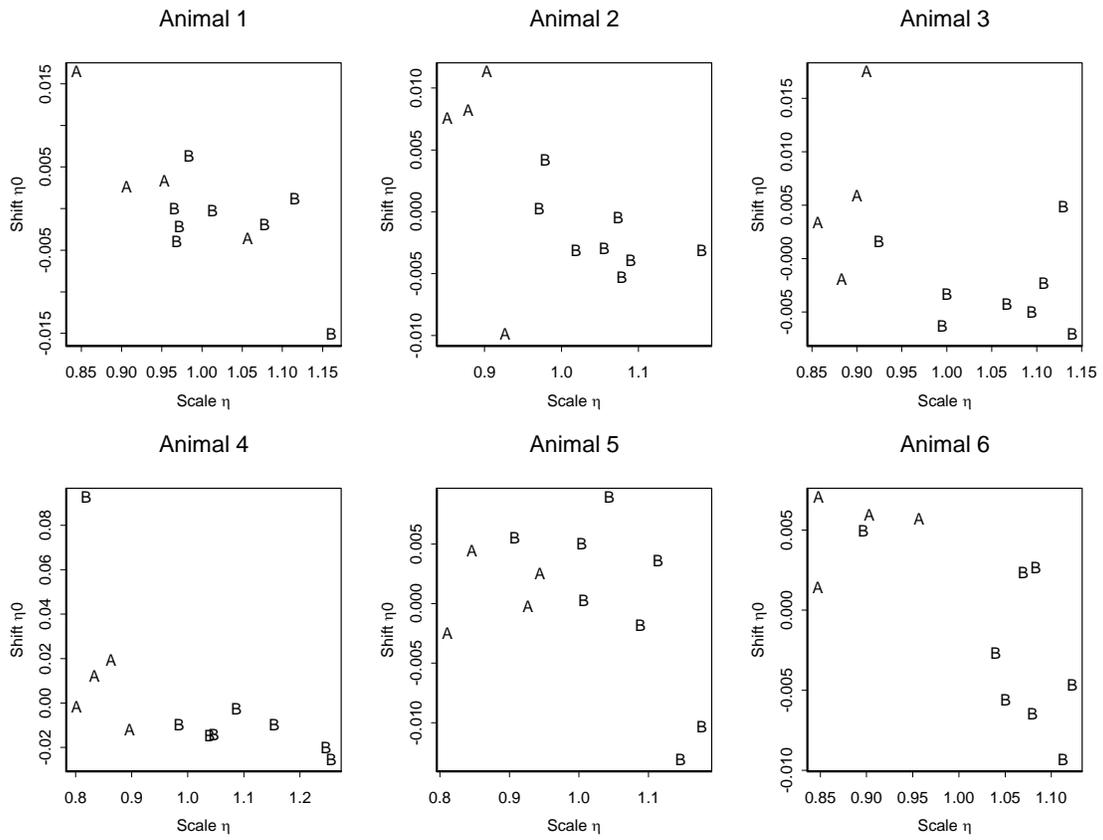


Figure 1. Normalization parameters for each array with input mass 200 ng (A) and 400 ng (B) indicated. Normalization consists of subtraction of the shift followed by division by the scale.

Figure 2

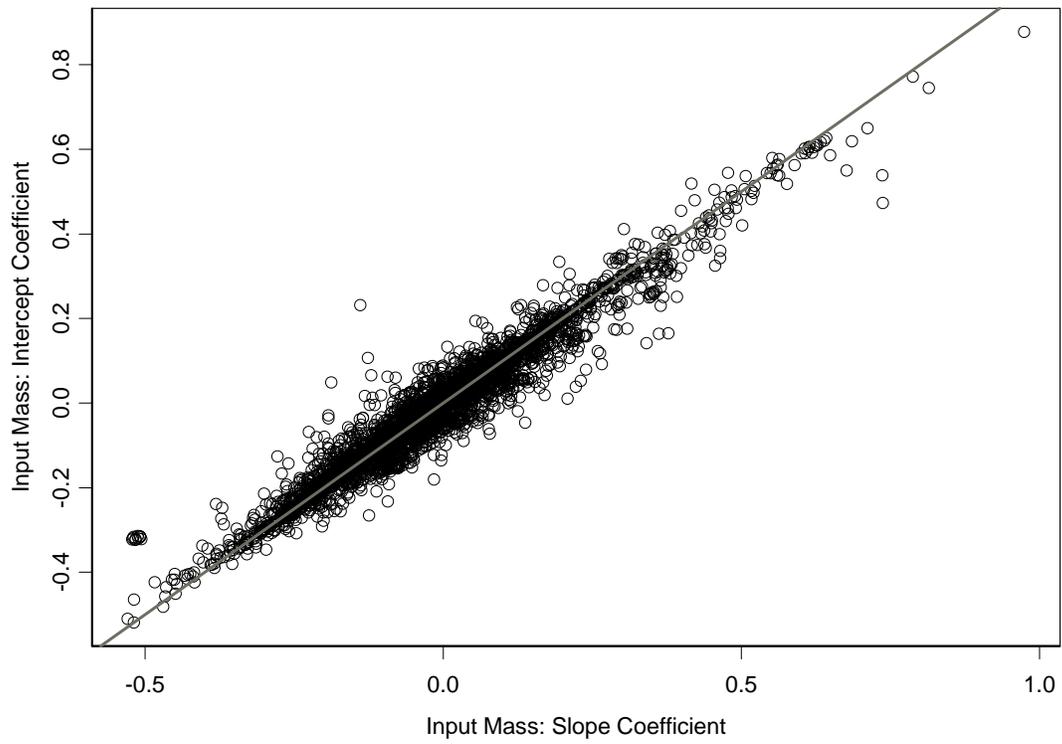


Figure 2. The input mass effect for 5014 selected probes. The effect is given as fractional change in the calibration curve intercept and the fractional change in the calibration curve slope.

Figure 3

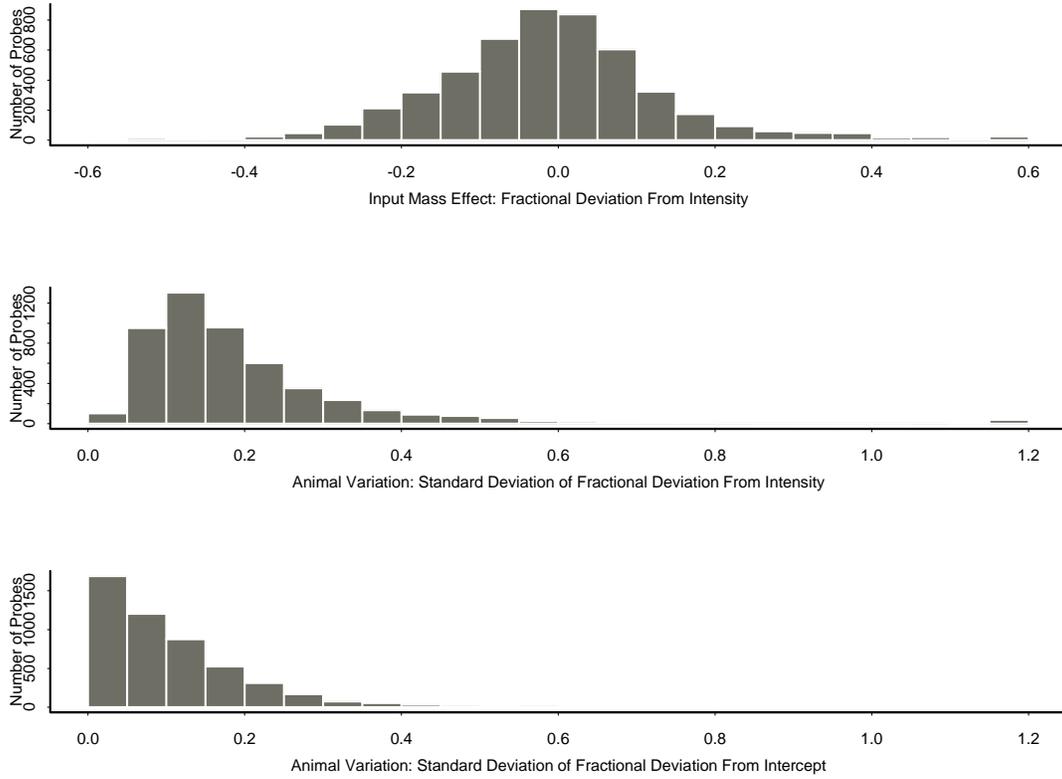


Figure 3. The input mass effect compared to the animal-to-animal variation and the animal intensity variation compared to the animal intercept variation. Histograms for 5014 selected probes.

Figure 4

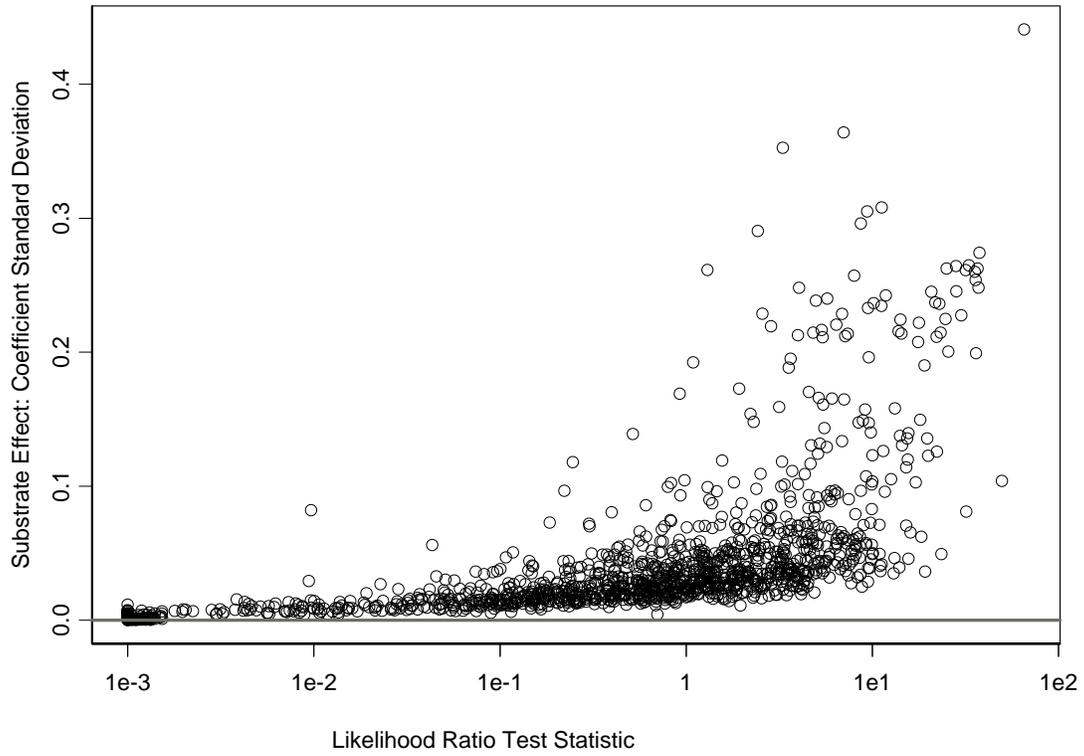


Figure 4. Aspects of the substrate effect for 5014 probes: effect standard deviation and the statistic for the likelihood ratio test of the hypothesis of no substrate effect.

Figure 5

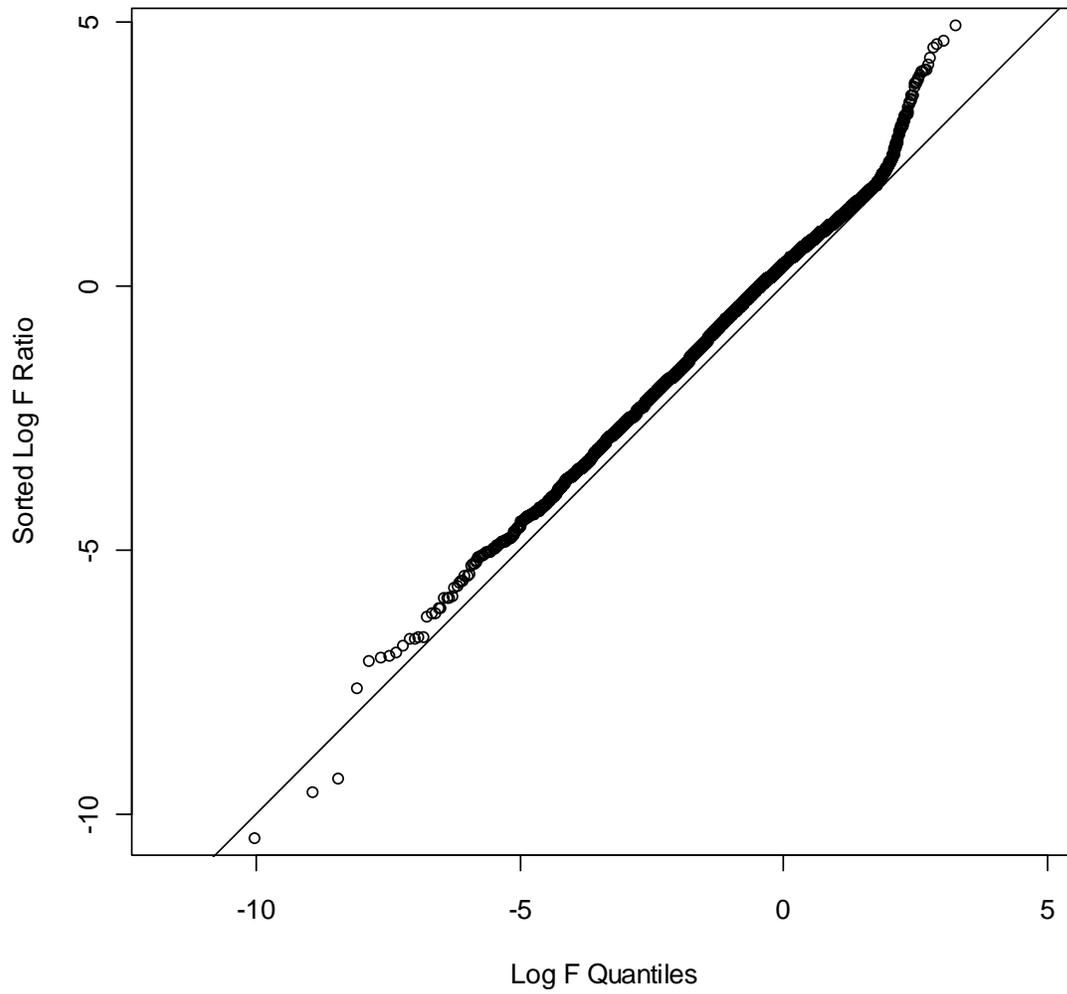


Figure 5. Measurements of the pooled materials from 11530 probes: quantile-quantile plot of the F test for linearity of the calibration curve.