

Quality Assessment of Gene Expression Data for Affymetrix GeneChips

François Collin^{1,2}, Julia Brettschneider², Ben Bolstad³, Terry Speed²

Introduction

Producing gene expression data using microarray technology is an elaborate process with many potential sources of variability. To maximize the scientific value of gene expression information derived from microarrays, it is essential to make rigorous quality assessments throughout the process.

Standard sample preparation protocols include a number of qualitative assessments meant to ensure that good quality RNA is used in the hybridization experiments. Following hybridization and image processing, each microarray chip provides a wealth of information that can be used to assess the quality of the data. Recommended post-hybridization quality assessment include general image quality assessment and analysis of intensity measures of specialized probes.

We propose some quality assessments derived from fitted models used to obtain expression values. We are using much of the data on the chip, to assess its quality and derive quality assessments which are more directly related to the quality of gene expression values than existing standards.

Model based expression data quality assessment

Consider the following model [1, 2]:

$$\log_2(\text{PM}_{kij}^*) = p_{ki} + c_{kj} + e_{kij} \quad (1)$$

where PM_{kij}^* is the background corrected, normalized probe intensity for probe set k , probe i , chip j ; p_{ki} represent the probe affinity effects; c_{kj} represents the \log_2 scale expression estimates; and e_{kij} an independent, identically distributed error term with mean zero and variance σ_k^2 . Many departures from quality standards attributable to processing failures will be reflected by inflated residuals from the fits to the models (1). Summarizing the residuals on a chip can therefore be expected to provide good discrimination among chips producing data of varying quality.

Normalized Unscaled Standard Errors (NUSE)

As we are ultimately interested in departures in quality to the extent that these affect the expression estimates it makes sense to combine residuals into estimated standard errors of expression estimates and summarize these at the chip level. Assuming that the models (1) were fitted robustly by iteratively reweighted least squares (IRLS), we can get an estimate of the unscaled standard error of the estimated \log_2 scale expression estimates as: **unscaled SE (c_{kj}) = $1/\hat{\alpha}_i w_{kij}$**

Thus for each chip, indexed by j , we get a vector of unscaled standard errors of estimated expressions, one component for each probe set, indexed by k . Note that the SEs of estimated expressions within a chip form a heterogeneous set as the value of σ_k^2 varies from probe set to probe set. Some heterogeneity across probe sets still remains in the unscaled SEs as $\sum_i w_{kij}$, the effective number of probes used in estimating the expression for probe set k , chip j , may vary from probe set to probe set. To remove this source of heterogeneity, we can normalize the unscaled standard errors by dividing by the average, or median, value of $1/\sum_i w_{kij}$ across chips. As a result we get a normalized, unscaled standard error (NUSE) of expression estimates for each probe set on a chip. We can summarize this vector of values for each chip to get an assessment of chip expression quality.

Spatial analysis of residuals

Residuals can be imaged on the chip in a manner similar to the way cell intensities are typically imaged. Spatial patterns of residuals themselves have proven difficult to visualize. The challenge is to capture spatial patterns of a dense scatter of numbers having both sign and amplitude. Each of these features are readily captured separately though. The weights used in the IRLS fit can be imaged to capture the magnitude of the residuals, highlighting residuals that deviate substantially from an overall estimated scale.

Relative expression summaries

We can also gage variability of expression measures obtained from a chip by summarizing the distribution of relative log expressions. For reference we use a virtual *median chip* constructed by taking, for each probe set, the median log expression from a set of chips. A vector of probe set relative log expressions for each chip can be summarized this by a measure of bias (median) and scale (IQR). These summaries are sensitive to technical sources of variability that are large compared to biological variation.

MAS 5 quality assessment procedures

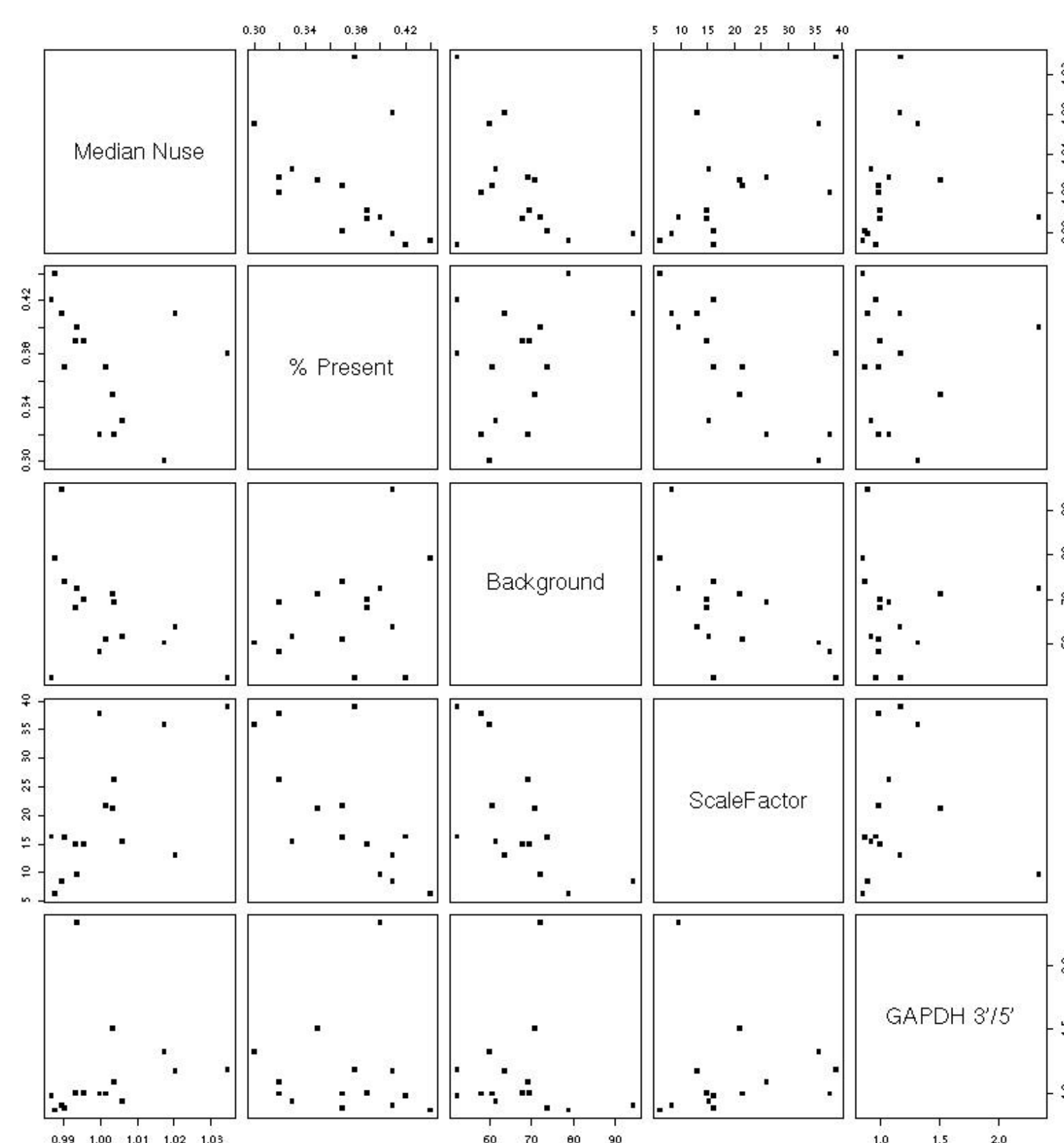
A number of quality checks can be performed following the analysis of the raw data by the Affymetrix software (MAS 5.0) [3]. Here we compare the model based chip expression quality assessments with quantitative assessments available through the MAS 5.0 analysis report, as well as an easily computed assessment based on housekeeping gene signal values:

- Scaling Factor
- Background
- Percent Present calls
- GAPDH 3'/5' ratio

Results

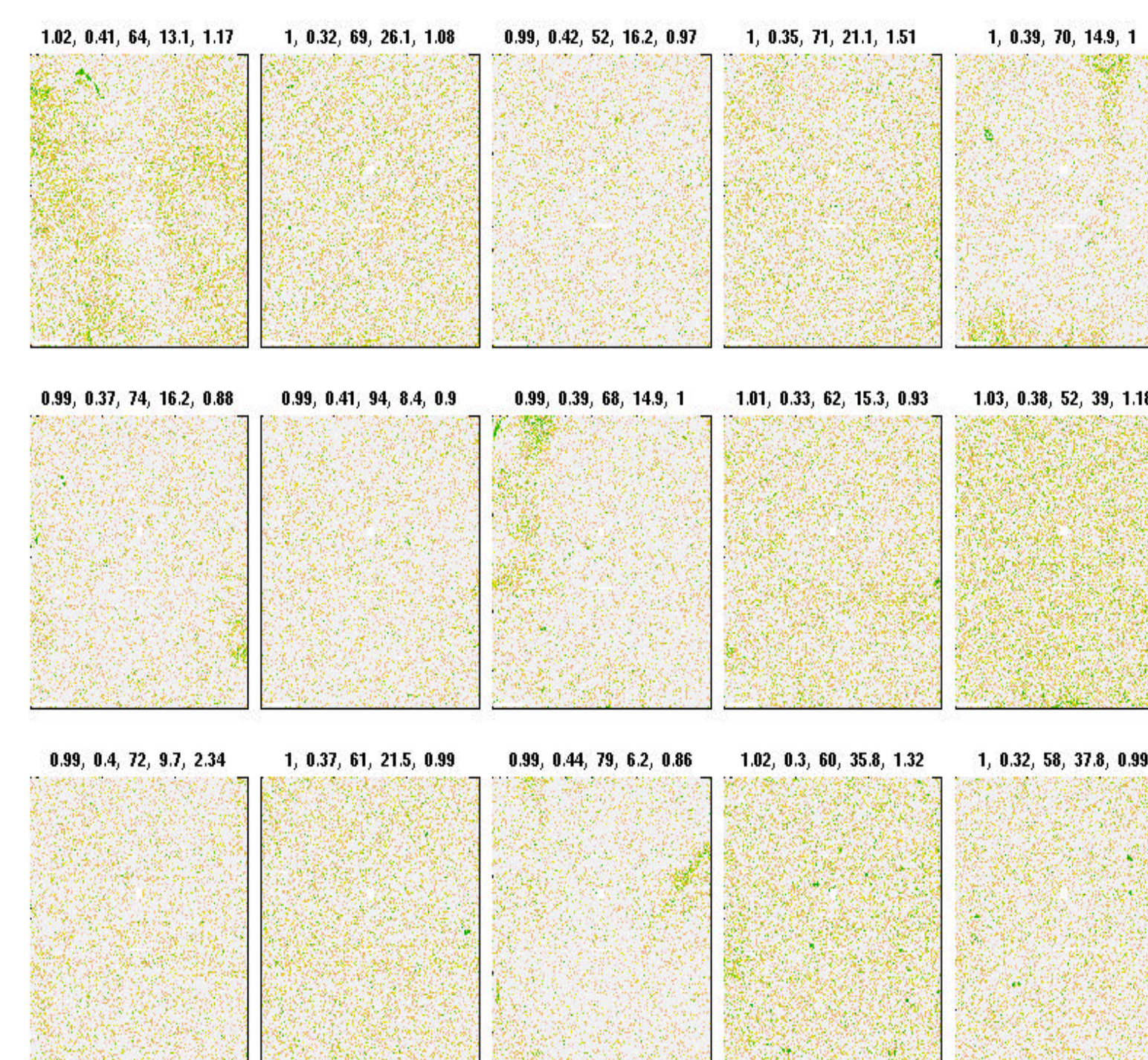
Relationships among quality indicators in 15 HG U133A chips

Three chips have slightly elevated expression variability as assessed by the chip median NUSE. There are no outliers in the % Present call, or Scale Factor statistics. There is one outlier in the Background and GAPDH 3'/5' ratio statistics – the corresponding chips do not have elevated assessed expression variability.



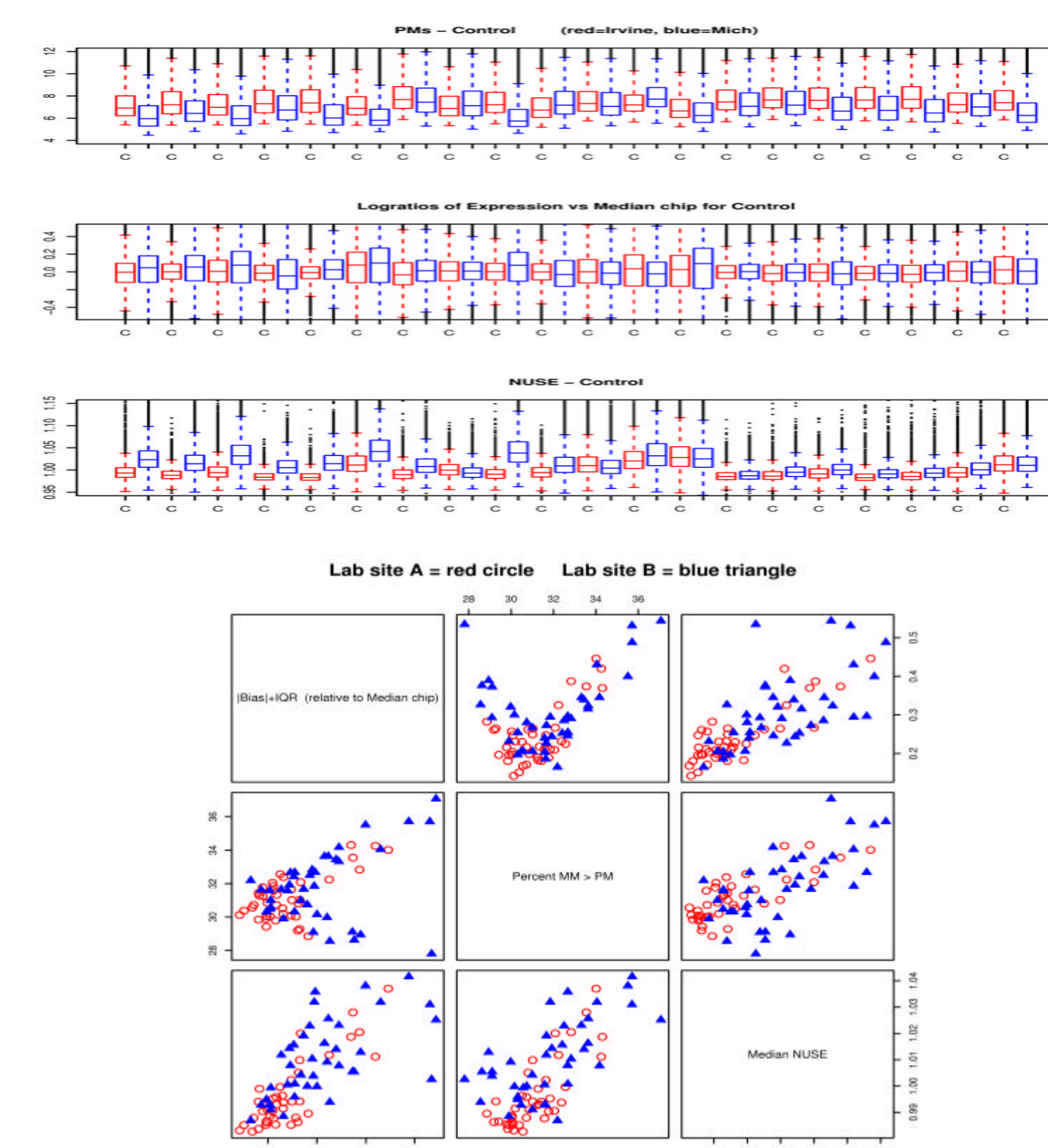
Spatial distribution of large absolute residuals

Pseudo images of the IRLS weights show pattern of residuals with large absolute value (the green dots). Headings for each chip is a 5 point quality summary: Median NUSE, % Present, Background, Scale Factor, and GAPDH 3'/5', respectively. Among the chips with elevated residuals, one shows a local pattern, chip (1,1), and 2 show a more evenly distributed pattern of elevated residuals, chips (2,5) and (3,4).



Dataset II: Quality difference between lab sites

The boxplots for the NUSE show the differences in hybridization quality most clearly, in magnitude as well as variability. A high NUSE corresponds - most of the time, but not always - to a low signal (PM). The IQR of the log-ratio of the expression versus the median chip also detects this difference. The scatterplots of |bias|+IQR and median NUSE show that these scores are correlated, even when looking at both lab sites separately. Plotting these measures against the MM>PM rate shows outlier groups. These chips have a low MM>PM rate, but they do not score as well under the other measures.



Data

To illustrate the use of assessed expression measure variability and chip images of residuals, we use set of cel files that is publicly available and accessible through the web [4].

Our second dataset is a part of a case-control study including replicates done in different lab sites. Most of the hybridizations are of good quality, but the refined probe level based quality measures make a systematic difference in the chip quality between the labs very obvious.

Discussion

The model based quality assessment proposed shows greater sensitivity to departures from standards than existing standards. Reflecting variability in expression measures, it also provides a better basis for judging quality. Chip quality measures can be used to look for systematic differences due to experimental conditions, as shown in the detection of quality differences between lab sites in dataset II.

Detecting departures from quality standards is only the first step toward improving data for analysis. Following detection, an action needs to follow – should one reject a chip from future analysis, or adjust the analysis to account for increased variability. One would also like to diagnose the cause of the departure from quality standards to improve production processes. These are some important open questions.

References

- [1] R. A. Irizarry et. al. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Research*, 31(4 e15), 2003.
 - [2] R.A. Irizarry et. al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*, 2003. to appear.
 - [3] *Affymetrix Microarray Suite User's Guide version 5.0.*, Affymetrix, Inc, Santa Clara, CA, 2001.
 - [4] Mary E. Ross et. al. Supplemental Information for: Classification of pediatric acute lymphoblastic leukemia by gene expression profiling, *Blood*, 2003 <http://www.stjude-research.org/data/ALL3/rawFiles.html>
1. **Gene Logic, Inc., Data Analysis Group**
 2. **Department of Statistics, University of California, Berkeley**
 3. **Division of Biostatistics, University of California, Berkeley**