BIOS 7659 Journal Club:

A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. (Bolstad et al., 2003)

Tim Vigers

9/22/2020

Introduction

- ▶ The goal of normalization is to separate the interesting biological variation from the variation that is a result of sample preparation, array production and processing, etc.
- ▶ Affymetrix proposes scaling the arrays so that each one has the same mean expression summary measure.
    - ▶ This does not work well when there are non-linear relationships between arrays.

Alternatives

- Other approaches, such as non-linear smooth curves or transforming data to standardize the distribution of intensities across arrays, rely on picking a "baseline" array.
- Bolstad et al. compare three different approaches, all of which combine data from every single array rather than relying on a baseline.

Cyclic loess

- ▶ Basically an extension of the M vs. A plots discussed in class, but applied to pairwise combinations of Affymetrix arrays.
- ▶ M is the difference in log expression values and A is the average (a Bland-Altman plot).
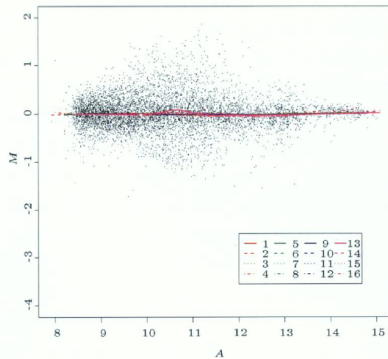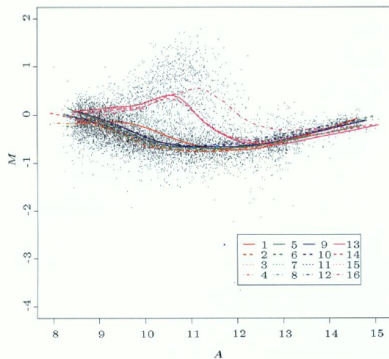

Figure 1: Dudoit et al., 2002

Cyclic loess

1. Take two arrays i and j, each with probes $k = 1, ..., p$.

2. Create an MA plot for these two arrays, and fit a loess curve through these data:

$$M_k = log_2(\frac{x_{ki}}{x_{kj}}), A_k = \frac{log_2(x_{ki}x_{kj})}{2}$$

3. Subtract the normalization curve fits $M'_k = M_k - \hat{M}_k$ and obtain adjusted probe intensities:

$$x'_{ki} = 2^{A_k + \frac{M'_k}{2}}, x'_{kj} = 2^{A_k - \frac{M'_k}{2}}$$

4. Take each of these adjustments (one for each pairwise comparison between arrays) and weight them equally across the set of arrays.

Contrast method

References

1. Dudoit, S., Yang, Y. H., Callow, M. J., & Speed, T. P. (2002). STATISTICAL METHODS FOR IDENTIFYING DIFFERENTIALLY EXPRESSED GENES IN REPLICATED cDNA MICROARRAY EXPERIMENTS. Statistica Sinica, 12(1), 111–139. JSTOR.