

Reading assignment on Normalization

The two papers for this assignment describe normalization methods for affy arrays (see Reading List for websites from which they can be downloaded)

The general idea is to perform normalization by making the distributions of “signals” in different arrays look similar – achieving this ought to be a guarantee of comparability.

Here we want to focus on normalization of multiple arrays.

- Sidorov *et al.* actually describe normalization for two arrays (model or baseline, and array to be normalized – these are the equivalent of numbers from green and red in one spotted array, but come from two separate arrays; sources of unwanted variation compromising comparability are different). However, the approach can be extended by normalizing multiple arrays to the same baseline.
- Bolstad *et al.* consider directly multiple arrays, comparing several approaches without and with selection of a baseline.

Pay attention to the criteria used to assess performance of a normalization approach.

Statistical technicalities:

Sidorov et al:

- Log-extreme largest value probability distribution function (pdf)
- Rules for constructing histograms

Bolstad et al.

- quantile-quantile plot
- trimmed mean
- latin square (mentioned, not crucial)
- variance versus bias trade-off

Instructions:

Work in groups. Read the two papers carefully and critically, and discuss them among yourselves. Write up a report of at most 3 pages (typed), highlighting features that you found interesting, points that you found confusing and open questions. Due date: **Tue March 2nd, in class.**