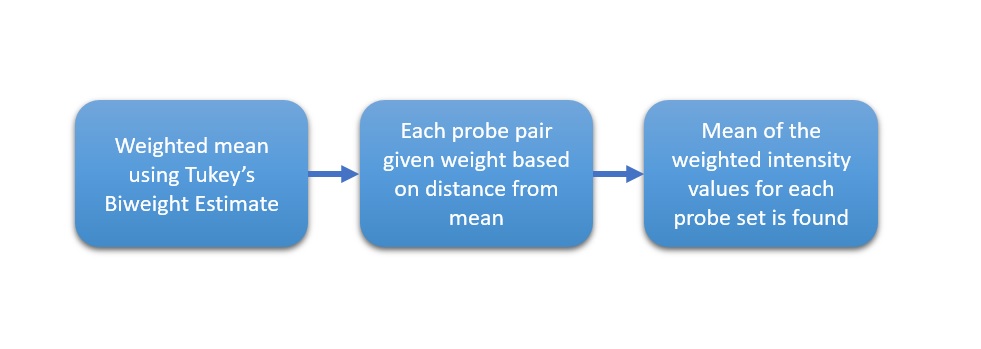
**Microarray Data Normalization Natalia Maksymchuk**

Normalization is a term that is used to characterize the process of eliminating of variations to allow appropriate comparison of data obtained from the two samples. In the case of microarray experiments (large-scale experiments) there are many sources of systematic variation that affect measurements of gene expression levels. In fact, average expression ratio of reference genes (genes that should not change in the two conditions) deviates from 1. This may be caused by different amounts of starting mRNA in two samples or by differential labelling efficiency of the two fluorescent dyes [1] or differential detection efficiencies.

One more reason for normalization is nonspecific binding for a given target. That’s why, Raw Affy data contains about twenty probes for the same RNA target and about half of these do not precisely match the target sequence. Thus, amount of nonspecific binding can be theoretically measured for a given target [2].

There are 3 techniques for microarray normalization:

1. Affymetrix MAS5 algorithm. MAS5 is a sensitive and selective algorithm for identifying differentially expressed genes [4]. It uses both perfect match and mismatch probes. MAS5 algorithm is following [2]:



When MAS5.0 algorithm is used alone for generating expression summaries it has high False Positive rates resulting from exaggerated variance at low intensities [5].

MAS5.0 algorithm has low computational requirements. It considers data on a perarray

basis, placing much lower demands on memory than other approaches that must access the entire dataset at once [5].

1. RMA (Robust Multi-array Average). It is an algorithm used to create an expression matrix from Affymetrix data. The raw intensity values are background corrected, log2 transformed and then quantile normalized. Next a linear model is fit to the normalized data to obtain an expression measure for each probe set on each array [3]. It summarize the perfect matches through median polish. The median polish algorithm, although robust, however it behaves differently depending on the number of samples analyzed.
2. FARMS (Factor Analysis for Robust Microarray Summarization) is a technique for summarizing array data at perfect match probe level. It is based on a factor analysis model for which a Bayesian maximum a posteriori method optimizes the model parameters under the assumption of Gaussian measurement noise. According to the Affycomp benchmark FARMS outperformed all other summarizations methods with respect to sensitivity and specificity.

* FARMS (Factor Analysis for Robust Microarray Summarization) is a summarization based on a factor analysis model for which a Bayesian Maximum a Posteriori method optimizes the model parameters under the assumption of Gaussian measurement noise[6].
* RNA concentration is estimated from the model. farms does not use background correction and uses either quantile normalization or cyclic loess[6].
* FARMS uses quantile normalization as default normalization procedure[6]
* It is computational efficient. According to the Affycomp benchmark FARMS[7] outperformed all other summarizations methods with respect to sensitivity and specificity[6].

**References**

1. M. Madan Babu An Introduction to Microarray Data Analysis

<https://www.mrc-lmb.cam.ac.uk/genomes/madanm/microarray/chapter-final.pdf>

1. Microarray analysis techniques <https://en.wikipedia.org/wiki/Microarray_analysis_techniques>
2. Robust Multi-array Average (RMA) <http://www.molmine.com/magma/loading/rma.htm>
3. The utility of MAS5 expression summary and detection call algorithms

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1950098/>

1. Stuart D Pepper, et. al. The utility of MAS5 expression summary and detection call

algorithms. BMC Bioinformatics, 2007.

1. Djork-Arné Clevert. Using FARMS for summarization using I/NI-calls for gene filtering. Software Manual, 2018.
2. http://affycomp.jhsph.edu