

Readings for Module 3: Normalization, Missing Values, Transfo's and Filtering

- Affymetrix, Microarray Suite version 5.0 (2002). Statistical Algorithms Reference Guide (two array comparison analysis. Normalization proposed in Affymetrix algorithms).
- Schuchhardt, J., Beule, D., et al. (2000). Normalization strategies for cDNA microarrays. *Nucleic Acids Research* 28, 10.
- Finkelstein D.B., Ewing R., Gollub J., Sterky F., Somerville S., Cherry J.M. (2002). Iterative linear regression by sector. *Methods of Microarray Data Analysis*. SM Lin, KF Johnson eds (Kluwer Academic), 57-68. Preprint: http://afgc.stanford.edu/afgc_html/CAMDA_11_update.pdf
- Kepler T, Crosby L., and Morgan K.T. (2002). Normalization and analysis of DNA Microarray Data by Self-Consistency and Local Regression. *Genome Biology* 3(7), research0037. SFI tech report at: <http://www.santafe.edu/sfi/publications/Abstracts/00-09-055abs.html>
- Hartemink A.J., Gifford D.K., Jaakkola T.S., Young R.A. (2001). Maximum likelihood estimation of optimal scaling factors for expression array normalization. SPIE BIOS 2001, San Jose, California. Preprint: <http://www.psrc.lcs.mit.edu/publications/Papers/normabs.htm>
- Yang Y.H., Dudoit S., Luu P., Speed T.P. (2001). Normalization for cDNA microarray data. SPIE BIOS 2001, San Jose, California. Preprint: <http://www.stat.Berkeley.EDU/users/terry/zarray/Html/>
- Yang Y.H., Dudoit S., Luu P., Lin D.M., Peng V., Ngai J., Speed T.P. (2002) Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. *Nuclei Acid Research* 30(4), e15.

Normalization for affy arrays:

- Sidorov IA, Hasack DA, Gee D, Yang J, Cam MC, Lempicki RA, and Dimitrov DS (2002). Oligonucleotide micorarray data distribution and normalization. *Information Sciences* 146:67-73. Available at: <http://www-lmmb.ncifcrf.gov/~dimitrov/bibliogr/pdf/Sidorov2002InformSci.pdf> (can use for multiple arrays selecting a baseline array)
- Bolstad, B.M., Irizarry R. A., Astrand, M., and Speed, T.P. (2003), A Comparison of Normalization Methods for High Density Oligonucleotide Array Data Based on Bias and Variance. *Bioinformatics* 19(2),185-193. Preprint: <http://stat-www.berkeley.edu/users/terry/Group/index.html> new site, Speed's group (multiple arrays, without and with baseline)

Missing value imputation:

- Troyanskaya O., Cantor M., Sherlock G., Brown P., Hastie T., Tibshirani R., Botstein, D., and Altman, R. B. (2001). Missing value estimation methods for DNA microarrays. *Bioinformatics* 17, 520-525. Available at: http://ismb00.sdsc.edu/9_8_pdfs/SunMethods/Troyanskaya.pdf
- ZhouX., Wang X., Dougherty E.R. (2003). Missing-value estimation using linear and non-linear regression with Bayesian gene selection. *Bioinformatics* 19, 2302-2307.

For both normalization, and following module on differentially expressed genes

- Tsodikov A., Szabo A., Jones D. (2002) Adjustments and measures of differential expression for microarray data. *Bioinformatics* 18(2), 251-260.