

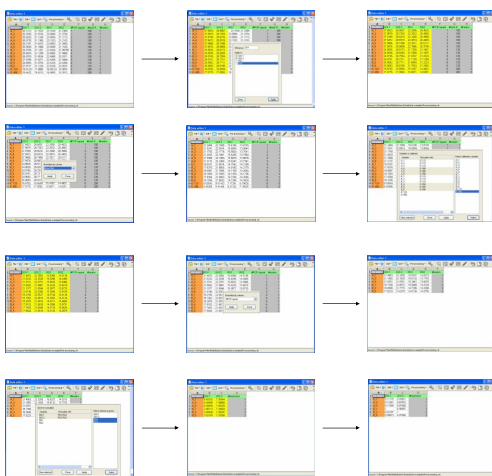
Mikael Kubista, José Manuel Andrade, Björn Sjögreen and Amin Forootan  
MultiD Analyses AB, Lotsgatan 5A, Gothenburg, Sweden (www.multid.se)  
Department of Analytical Chemistry, University of A Coruna, A Coruna, Spain

## INTRODUCTION

Real-time PCR has rapidly become the preferred technique for quantitative analysis of nucleic acids. Its superior sensitivity, reproducibility, and dynamic range make it the preferred choice for both absolute and relative quantification of nucleic acids, as well as expression profiling. Correct interpretation of real-time PCR data requires appropriate experimental design, accurate data pre-processing and analysis of the data using proper statistical and multivariate methods. For this process we have developed the GenEx software.

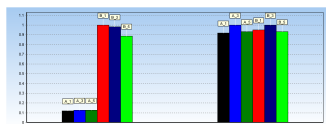
## DATA PRE-PROCESSING

1. Correct for primer-dimers
2. Correct for PCR efficiency
3. Normalize to sample amount
4. Normalize to Interplate Calibrators
5. Normalize technical repeats.
6. Normalize to reference genes
7. Convert to relative quantities
8. Convert to fold changes
9. Mean center/Autoscale data.

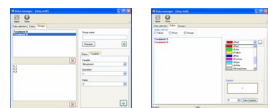


## COMPARISON OF GROUPS

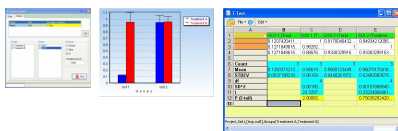
Expression of genes in multiple samples is readily presented in bar graphs



Create groups with treatment (biological) repeats using the index in the #treatment classification column, and assign plot colors and plot symbols

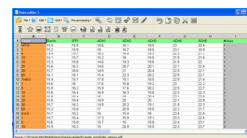


Compare means and variances of the treatment groups.

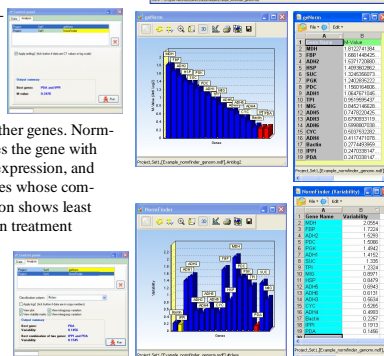


## FINDING OPTIMUM REFERENCE GENES

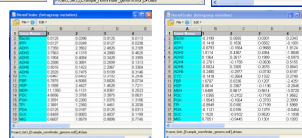
Data are entered as CT values. A classification column is added to identify groups of like samples for Normfinder analysis



geNorm identifies the two genes with the most constant expression ratio when compared to other genes. Normfinder identifies the gene with least variable expression, and the pair of genes whose combined expression shows least variation within treatment groups.

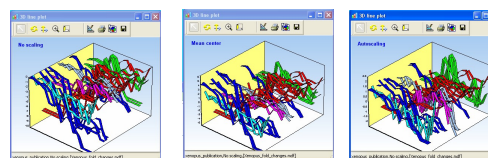


Normfinder also calculates the variability within each treatment group and bias between treatment groups using the information in the classification column.

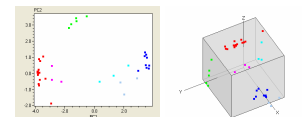


## EXPRESSION PROFILING

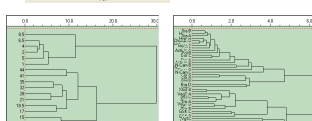
Classification of genes/samples based on similarities in expression profiles



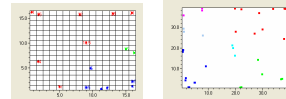
Classification of genes in *Xenopus laevis* during development by Principle Component Analysis



Classification of samples (mean centered) and genes (autoscaled) of *Xenopus laevis* during development by Hierarchical clustering



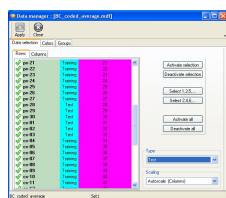
Self-organized map (SOM) constructed from expression profiles for *Xenopus laevis* developmental stages (samples) and genes.



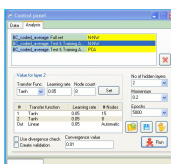
## CLASSIFICATION

Based on a training data set a model is developed based on artificial neural network (ANN) for classification. The model can then be used to classify unknown test samples.

GenEx has been selected classification tool in EU FP6 SMARTEALTH (www.smarthealthip.com)



Adjustable parameters defining ANN architecture



Convergence of ANN training (left) and classification of breast cancer samples (below)

