Microarray Cheat Sheet

by Nimisha (lemonbuzz) via cheatography.com/132716/cs/40278/

Introduction

Principle	Amount of hybridisation detected is proportional to no. of fragments in sample
Microarray	DNA probes bound to glass slide, to which sample DNA fragments can be hybridised
Probes	Oligonucleotides, ink-jet printed onto slides (Agilent) or synthesised in-situ (Affymetrix)
Sample	labelled ssDNA or antisense RNA

Replicates

Technical replicate	Repeated measurements or procedures using the same biological sample to evaluate precision and reproducibility
Biological replicate	Use of multiple independent biological samples to account for biological variability

One and Two Color Arrays

One Color Array	Each sample loaded into a separate microarray
Two Color Array	Two samples, labelled differently, loaded onto same microarray in same amounts. Competitive hybridisation.

Data Preprocessing

Background Correction	- Adjust for non-specific hybridisation - mismatch probes (Affymetrix) - exogenous negative control spots - remove features with lower intensity than background
Log Transformation	Improves data distribution for classical statistical analysis

Data Preprocessing (cont)

Normalisation	Removes systematic effects due to technical differences, which aren't due to biological differences

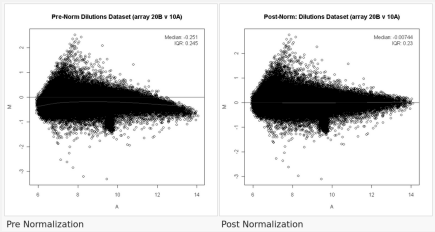
Normalisation

Within-Array Normalisation	Two-color microarrays; align two channels for each array
Between-Array Normalisation	One-color microarrays; Single channel platforms; Quantile normalisation, Cyclic Loess
Sources of Bias	Dye bias, Array bias, Spatial bias
Dye Swap Design	pair of samples compared twice with reversed dye
Reference Design	Each sample hybridised against a common reference sample

Within Array Normalisation

Sources of Bias	- Differential dye incorporation - Diff. emission response to excitation - Non-uniform focusing across the array
Correction of diff responses of Cy3 and Cy5 channels	1. LR of Cy3 vs Cy5 intensity 2. LR of log ratio against avg. intensity (**MA plot**) 3. Non-linear (**Loess**) regression of log ratio against avg. intensity

MA plots



MA plot - Interpretation

Vertical axis	Log ratio (R/G) = M i.e. log fold change; M = log2 (condA exp / condB exp)
Horiontal axis	average log-intensity between two cond; mean(condA exp + condB exp)
Non-zero intercept	One channel is consistently brighter than the other
Slope not equal to 1	One channel responds more strongly at high intensities than other
Slope not straight line	Non-linear relation b/w intensities of two channels

Loess Regression

Working	- local reg in overlapping windows of data - join the regression to form a smooth curve

Between Array Normalisation

Types	Scaling, centering, distribution normalisation
Scaling	Ensure mean/median of all distributions are equal; Subtract overall mean log intensity from each log intensity
Centering	Ensure that all the distribution's mean and SD are equal

By **Nimisha** (lemonbuzz)

cheatography.com/lemonbuzz/

Published 13th September, 2023.
Last updated 16th September, 2023.

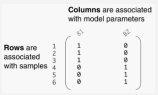
Between Array Normalisation (cont)

Quantile normalisation	- sort each array in order - average across rows - sort avg values in original order
Explanatory variables	**Covariates** (quantitative measurements) and **factors** (categorical variables)
Levels	Unique values within a factor

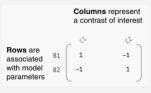
Matrices

Design or Model matrix	Describes the experimental design of the microarray experiment
Contrast matrix	Defines specific comparisons of interest b/w different experimental conditions i.e. defines the hypotheses to be tested
Types of design matrices	1. Mean Reference Model (with intercept) 2. Means model (wihout intercept)
Means model	Mean gene expression levels compared independently for each sample group
Mean reference model	One sample group set as baseline or reference; gene exp in other groups compared relative to reference

Design Matrix



Contrast Matrix



Microarray Overview

Applications	- find expressed genes in a sample/cond - find diff exp across samples/cond - exp signatures of sample/cond - exp signature of set of genes
Limitations	- rely on prev knowledge of genome seq - high BG levels due to cross-hyb - complex normalisations needed - limited range of detection due to BG and saturation signals
Not good for	- determining exp or diff exp of single or small set of genes - accurately studying protein levels and functional activity (measure steady-state levels of RNA transcripts) - absolute level of exp of a gene (true conc of mRNA)

By **Nimisha** (lemonbuzz)

cheatography.com/lemonbuzz/

Published 13th September, 2023.
Last updated 16th September, 2023.