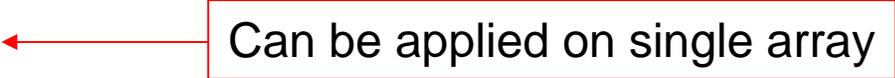
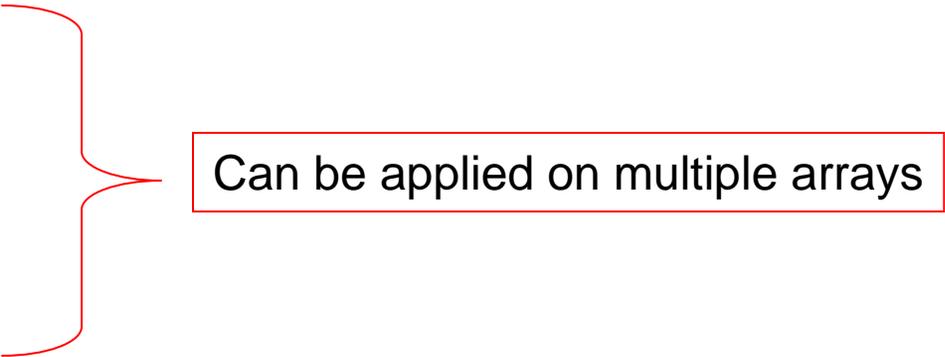


# Tutorial 7: Normalization Methods



# Normalization methods for Affymetrix data

- MAS 5  Can be applied on single array
  - RMA
  - DChip
  - Plier
  - Plier +16
- 
- Can be applied on multiple arrays

# Normalization methods for Affymetrix data - continued

Database Contents

- ANONYMOUS
  - Affy\_Rat\_MAS5\_only
    - Gene Lists
      - D0\_T12\_B\_a D0\_T12\_B[Biotin]
        - MAS5 {D0\_T12\_B\_a} [file: ...]
          - Mean/Median Scaling, ...
      - D0\_T12\_B\_b D0\_T12\_B[Bio...]
        - MAS5 {D0\_T12\_B\_b} [file: ...]
          - Mean/Median Scaling, ...
      - D0\_T12\_C\_a D0\_T12\_C[Bio...]
        - MAS5 {D0\_T12\_C\_a} [file: ...]
          - Mean/Median Scaling, ...
      - D0\_T12\_C\_b D0\_T12\_C[Bio...]
        - MAS5 {D0\_T12\_C\_b} [file: ...]
          - Mean/Median Scaling, ...
      - D0\_T12\_D\_a D0\_T12\_D[Bio...]
        - MAS5 {D0\_T12\_D\_a} [file: ...]
          - Mean/Median Scaling, ...
      - D0\_T12\_D\_b D0\_T12\_D[Bio...]
        - MAS5 {D0\_T12\_D\_b} [file: ...]
          - Mean/Median Scaling, ...
      - D2\_T12\_B\_a D2\_T12\_B[Bio...]
        - MAS5 {D2\_T12\_B\_a} [file: ...]
          - Mean/Median Scaling, ...
      - D2\_T12\_B\_b D2\_T12\_B[Bio...

Right-click the selected data, choose "Convert affy cel files to probe sets."

View data set(s) as wide spreadsheet - datasets side by side

Export

Convert affy cel files to probe sets

Mixed scatterplot

Virtual array images for data

Actual array images for data <<Dev. Only>>

Rank intensity plots for data

BarChart

Create gene list by data filtering...

Analysis

Quality Control

Normalize...

Duplicate data sets

Copy data sets for pasting elsewhere

Studies

Tree options...

Selection of methods

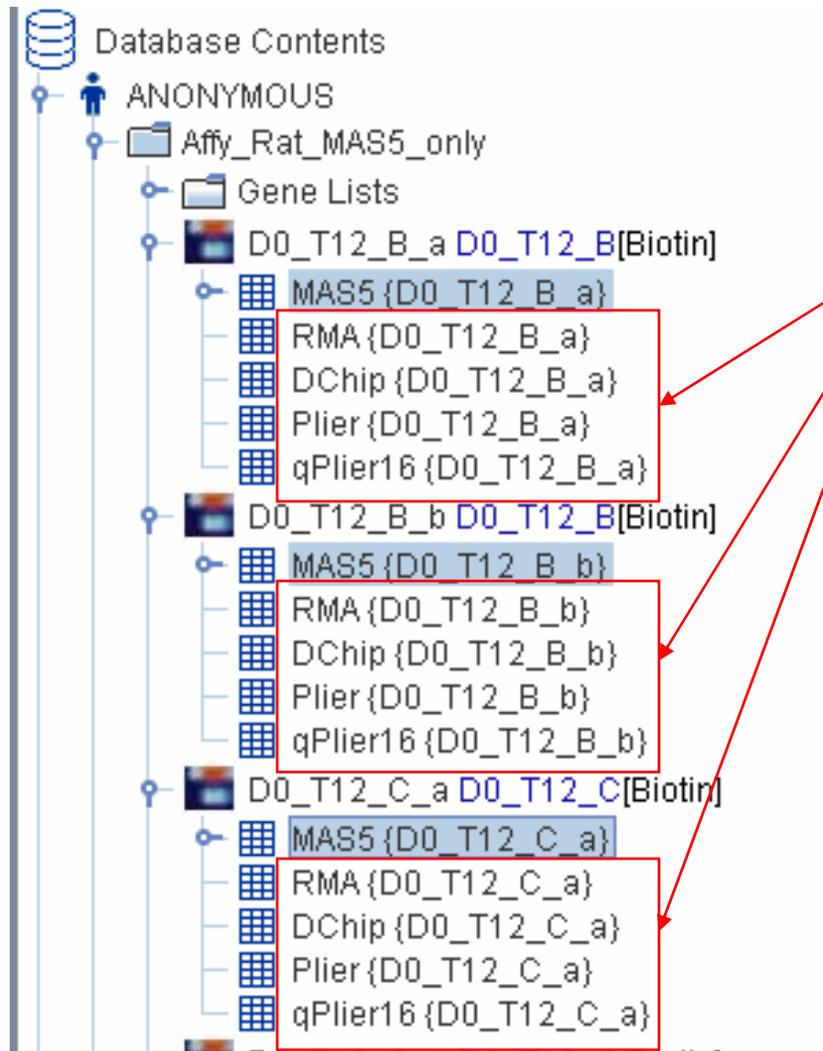
Select methods:

- MAS5
- RMA
- DChip
- Plier
- qPlier16

Quantile normalization for Plier

OK Cancel

# Normalization methods for Affymetrix data - continued



The converted probe sets files will be shown under the experiment.

# Other normalization methods

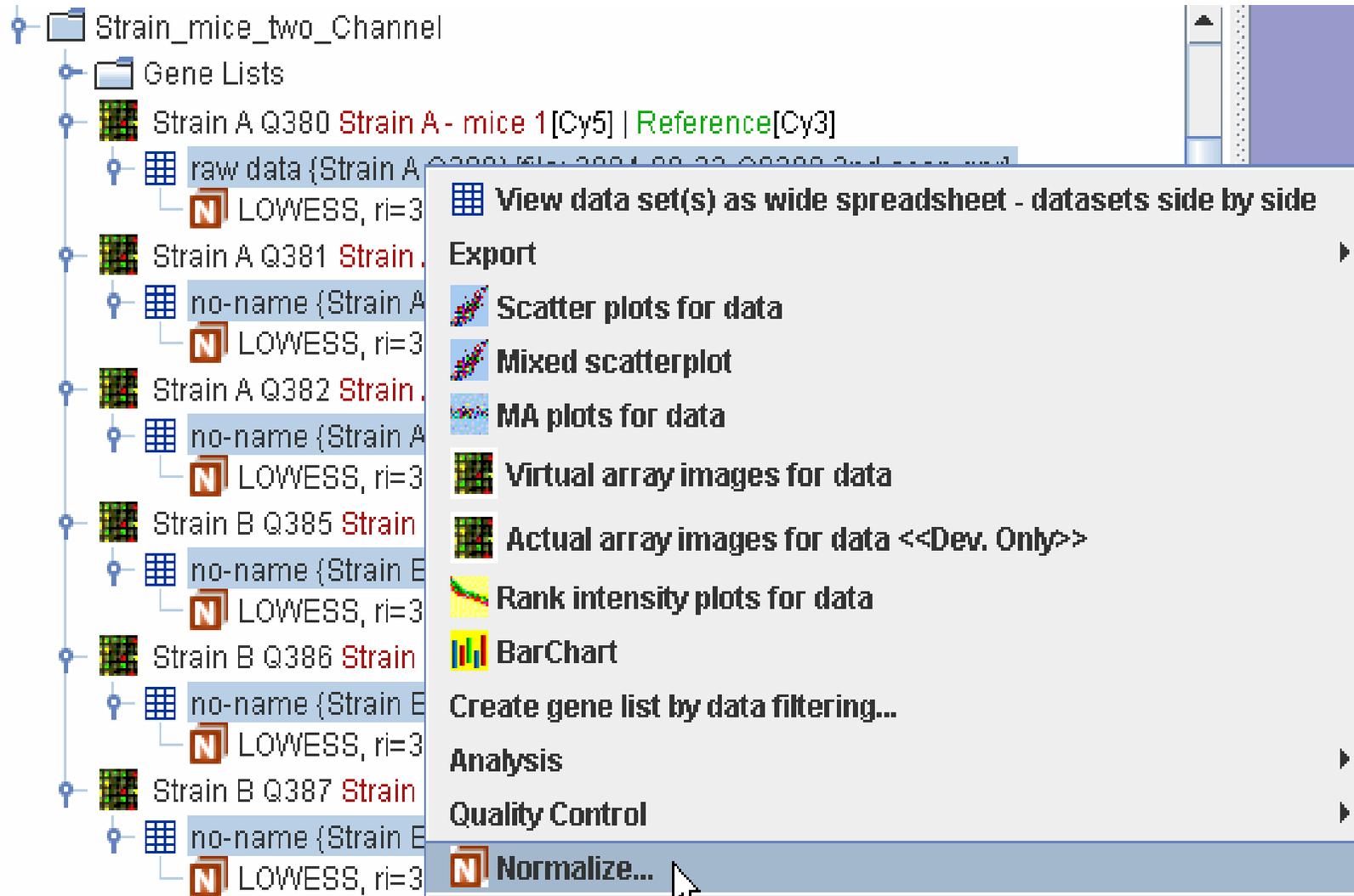
The following methods are for one and/or two channels

- Lowess
  - Total intensity norm
  - Linear Lowess
  - GenePix Mean Log ratio
  - Mean/Median scaling
  - Ref Avg Comp
  - Quantile
- For 2 channel only
- For either 1 or 2 channel
- For 1 channel only



# Normalize Data

Select the data, right-click, choose “Normalize...”



# Lowess

Lowess is for two channel data only.

**Select Normalization Method**

**LOWESS**

For 2 channel data only

The Lowess normalization method performs a robust locally weighted regression on the log ratio (M) vs. log geometric average (A) spot data, using each spot's locally estimated M value for spot by spot correction of log ratio values. Thus the Lowess method differs from many other normalization techniques because it is able to correct intensity (A) dependent ratio biases in an intensity-specific way. Visually this amounts to "straightening out" a curved A-M plot.

When the final local regression estimate Mfit of the log ratio has been calculated for a spot having channel values c1 and c2, the corrected ratio c1'/c2' is determined by

$$\log(c1'/c2') = \log(c1/c2) - Mfit$$

*(fitted M value become new zero point of M values at this avg intensity)*

which we can rewrite as

$$c1'/c2' = (c1/c2) * 1/b^{Mfit}$$

*(where b is the logarithm base for computed M-values)*

To present this correction in two channel format, we spread the ratio correction factor reciprocally to both channels, so that the same ratio correction is achieved:

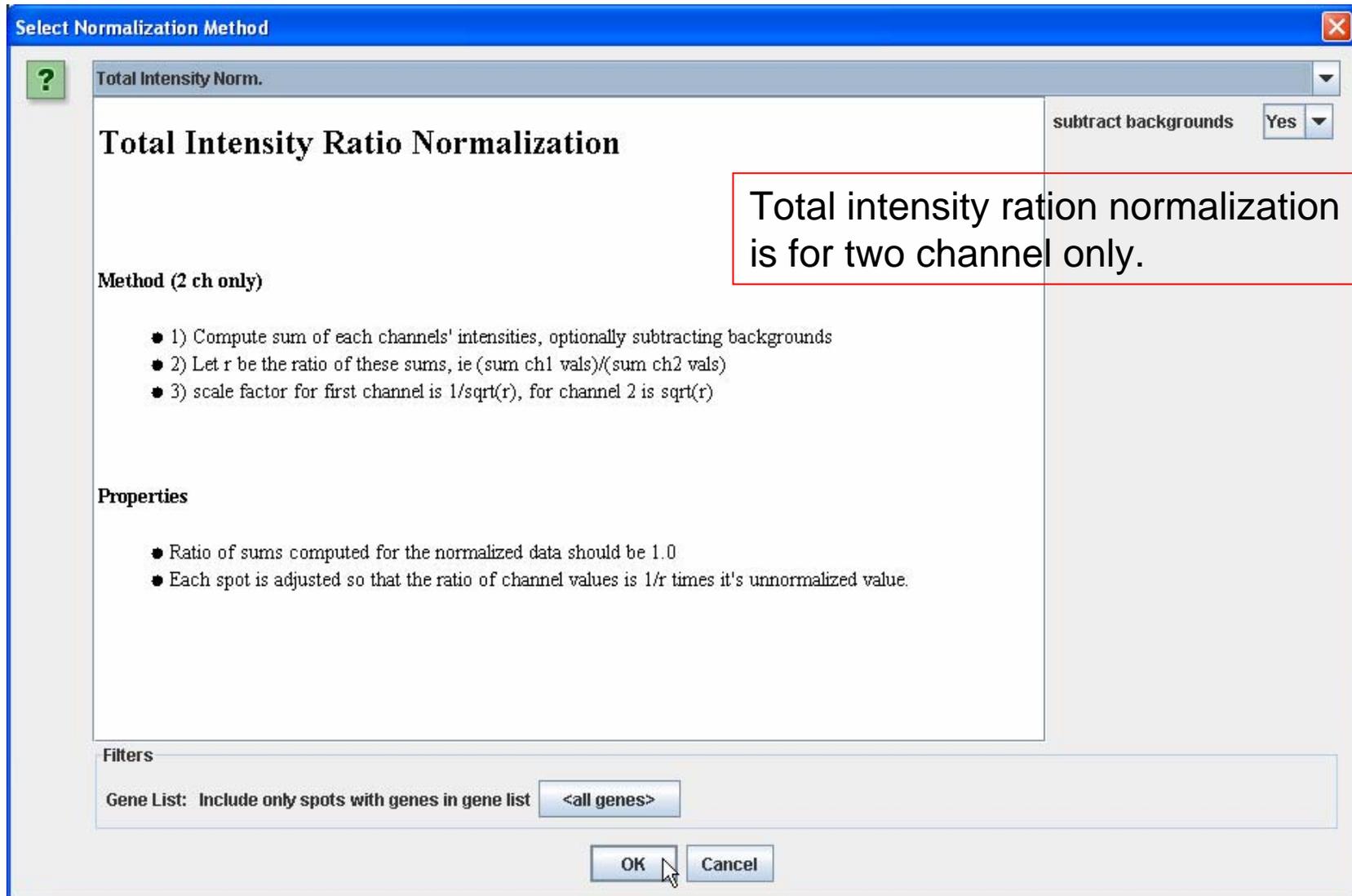
$$c1' = c1 * \sqrt{1/b^{Mfit}}$$
$$c2' = c2 / \sqrt{1/b^{Mfit}}$$

subtract backgrounds Yes  
smoothing factor 0.2  
robustness iterations 3  
delta

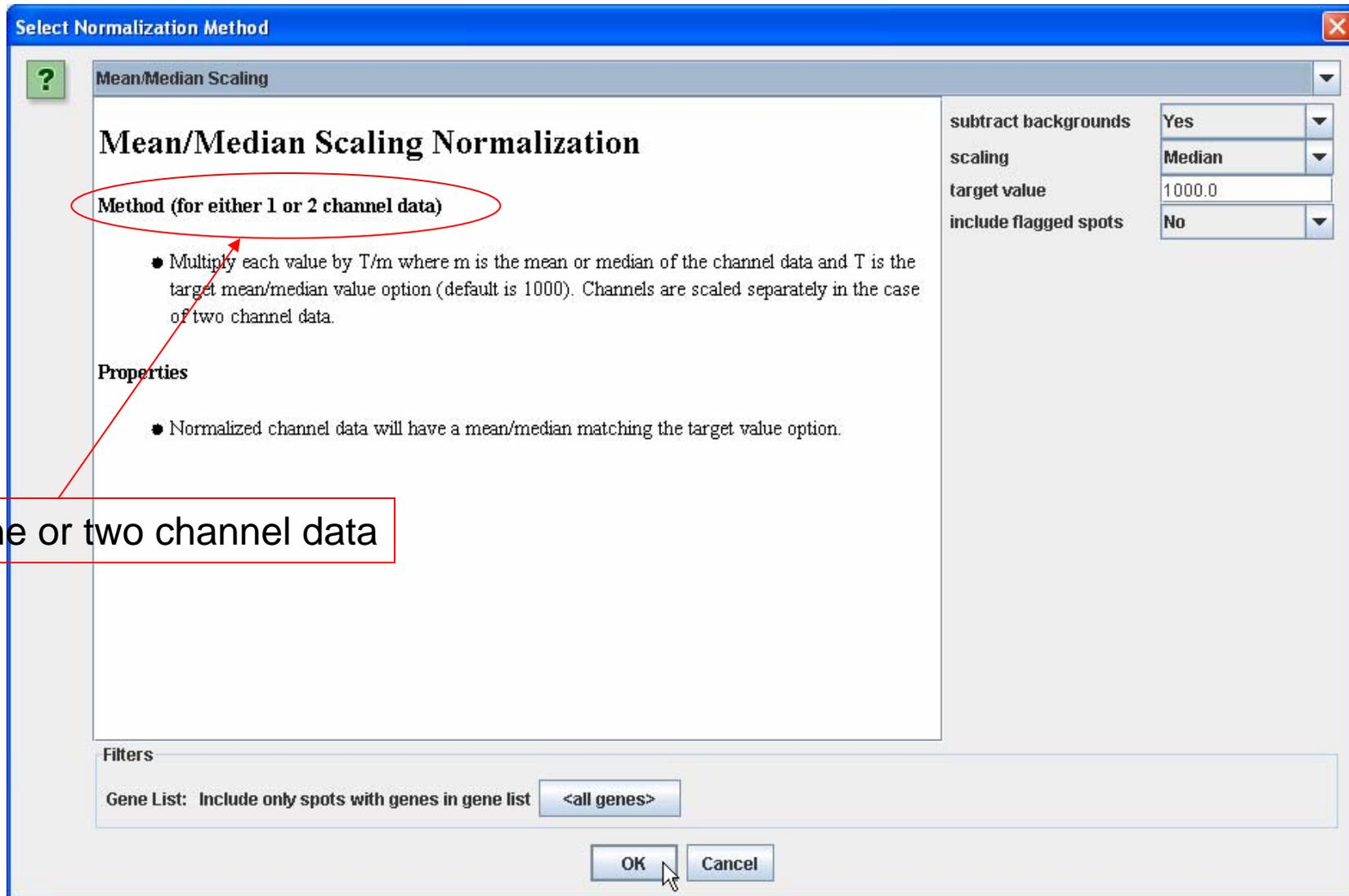
Filters  
Gene List: Include only spots with genes in gene list <all genes>

OK Cancel

# Total Intensity Ratio Normalization

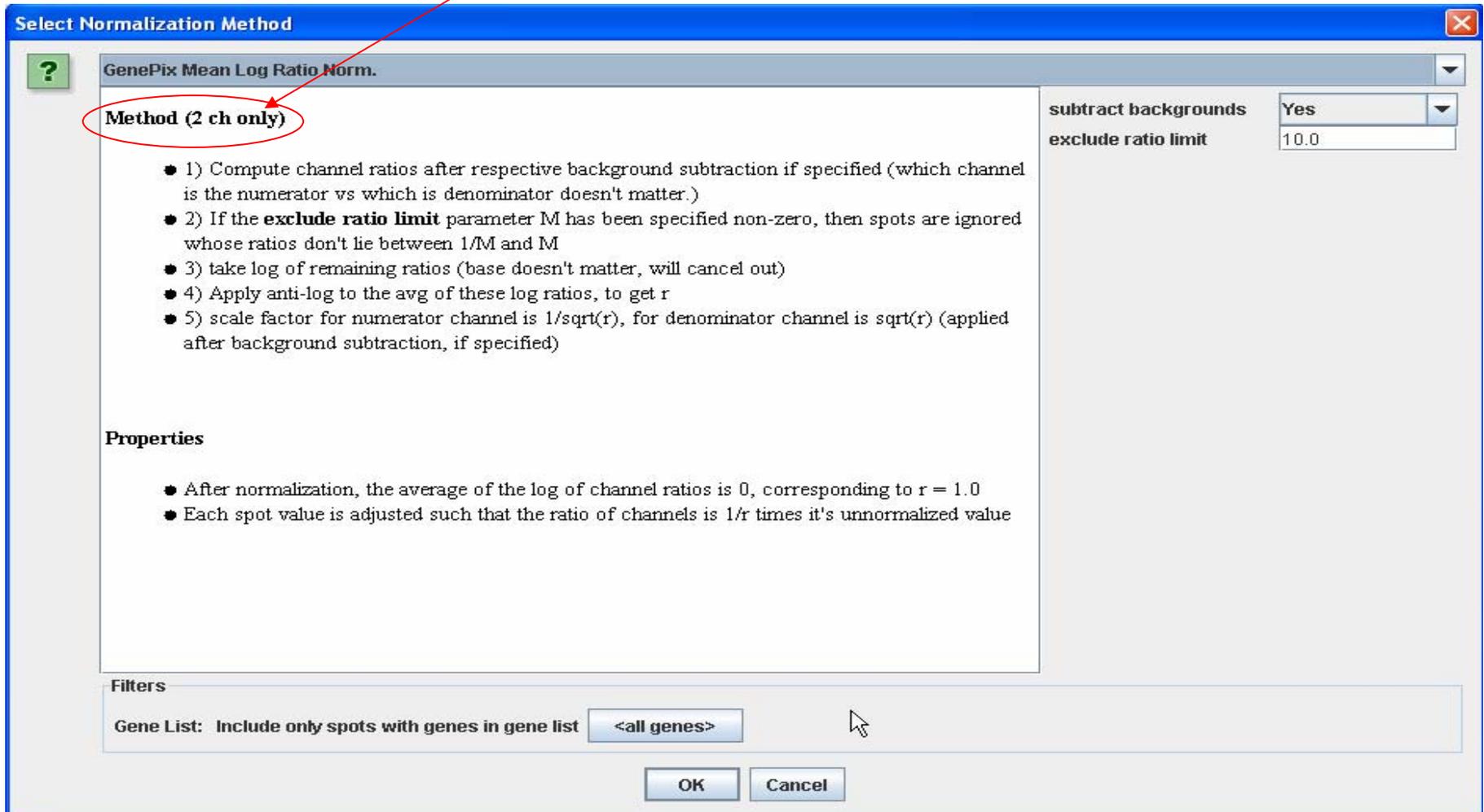


# Mean/Median Scaling Normalization



# GenePix Mean Log Ratio Normalization

This method is for two channel data only



# Linear & Lowess Normalization

Select Normalization Method

Linear&Lowess

## Linear&Lowess Normalization

Method (for 2 channel data only)

- First, values for each channel are multiplied by  $T/m$  where  $m$  is the mean or median of the channel data and  $T$  is the target mean/median value option (default is 1000).
- Then a Lowess normalization is performed on the resulting scaled channel data.

subtract backgrounds Yes

smoothing factor 0.2

robustness iterations 3

delta

scaling Geometric Mean

target value 1000.0

include flagged spots for scaling No

Filters

Gene List: Include only spots with genes in gene list <all genes>

OK Cancel

For two channel data only

# Quantile Normalization

**Select Normalization Method**

Quantile

## Quantile Normalization

**Method (1 ch only)**

All datasets are first sorted by ascending intensity value, after removing flagged values if the option is set to exclude them. All datasets are then trimmed to the length of the shortest dataset involved, by trimming the smallest (front) values from each. Then the least (front of list) value is taken from each dataset, and the arithmetic or geometric mean of these least values is written as the normalized intensity value for each of the front spots. Thus the same normalized intensity value is written for each dataset, however the spot (gene) that this intensity value is written for depends on the dataset. This process is repeated until the end of the datasets are reached.

**Properties**

- Each dataset normalized together via quantile normalization will have the same set of output intensity values.

**subtract backgrounds** Yes

**include flagged spots** Yes

**mean type** arithmetic

**Filters**

Gene List: Include only spots with genes in gene list <all genes>

OK Cancel

# Reference Average Comparison Normalization

For one channel only

**Ref Avg Comp**

## Reference Average Comparison Normalization

**Method (1 ch only)**

The normalized intensity  $I'$  is given in terms of the original intensity  $I$  and the reference intensities  $R(k)$  by

$$I' = I / g$$

where

$$g = (R(1) * R(2) * \dots * R(N))^{(1/N)}$$

is the geometric mean of the reference intensities.

**Properties**

- $\log(I') = \log(I) - \text{Avg}(\log(R(k)))$ , where Avg is the usual arithmetic average.

subtract backgrounds Yes

include flagged spots No

Highlight Reference Data Sets (Use Shift and/or Control keys for multiple selection)

- Affy\_Rat\_MAS5\_only/D0\_T12\_B\_a[r10649]
- Affy\_Rat\_MAS5\_only/D0\_T12\_B\_b[r10650]
- Affy\_Rat\_MAS5\_only/D0\_T12\_C\_a[r10651]
- Affy\_Rat\_MAS5\_only/D0\_T12\_C\_b[r10652]
- Affy\_Rat\_MAS5\_only/D0\_T12\_D\_a[r10653]
- Affy\_Rat\_MAS5\_only/D0\_T12\_D\_b[r10654]
- Affy\_Rat\_MAS5\_only/D2\_T12\_B\_a[r10655]

Exclude reference datasets from being normalized

Filters

Gene List: Include only spots with genes in gene list <all genes>

OK Cancel