

Overview

I Preprocessing & quality check

II Group annotation & stability

III Portraying analyses

IV Sample similarity analyses

V Pseudotime estimation

Overview

Dataset name:

Macro Showcase

Number of samples:

60

Number of groups:

4

Number of genes:

10009

ID type of genes:

ensembl_gene_id

Dimension of the SOM:

15 x 15

Date:

MI Sep 21 11:06:33 2016 CEST

Analyst:

wirth

scrat version:

0.9.9

Overview

I Preprocessing & quality check

II Group annotation & stability

III Portraying analyses

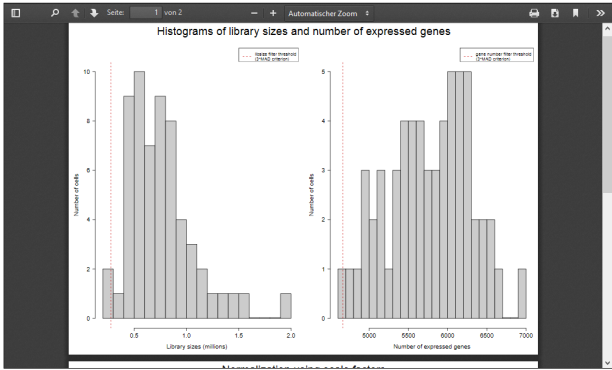
IV Sample similarity analyses

V Pseudotime estimation

Block I: Preprocessing & quality check

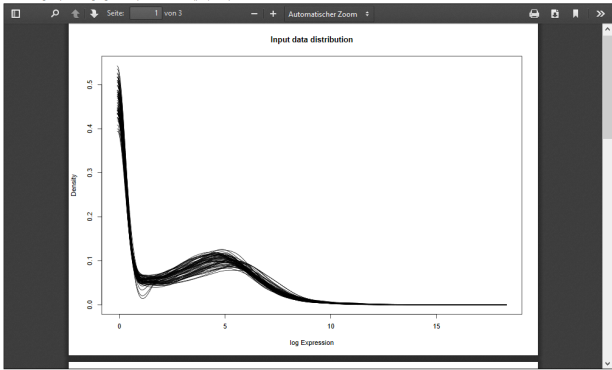
Read count preprocessing

Read counts are preprocessed using standard workflow:
Cells with library size 3 MAD below median are discarded, as well as cells with number of expressed genes 3 MAD below median (see p.1 in PDF below). Read counts are then translated to expression values using scale factor normalization (p.2 shows the factors).



Input data overview

Data distribution (p.1 in PDF below), mean expression vs. variance plot (p.2; inter-quartile ranges (IQR) are shown to identify potential outliers), and individual cells' and group-wise aggregated expression level (p.3) are provided.



Overview

I Preprocessing & quality check

II Group annotation & stability

III Portraying analyses

IV Sample similarity analyses

V Pseudotime estimation

Block II: Group annotation & stability

Data driven group annotation

skipped

Group silhouette coefficients

Silhouette coefficient (S) represents correlation of a cell's expression state to the centroid of its group, in relation to correlation to the nearest of the centroids of the other groups (p.1 in PDF below, individual cell and group-wise aggregated S). Positive S represents compact groups, whereas negative S indicate overlapping groups.

The figure shows a 'Correlation Silhouette' plot. The top part is a bar chart with the y-axis labeled 'S' ranging from -0.2 to 1.0. The bars are colored by group: blue, orange, green, and grey. The bottom part is a box plot with the y-axis labeled 'S' ranging from -0.2 to 1.0. The box plots are colored by group: blue, orange, green, and grey. The x-axis labels are 'Group 1', 'Group 2', 'Group 3', and 'Group 4'.

Overview

I Preprocessing & quality check

II Group annotation & stability

III Portraying analyses

IV Sample similarity analyses

V Pseudotime estimation

Block IV: Sample similarity analyses

Different sample similarity metrics and methods are applied to the dimension reduced expression module data and metadata, respectively. Application of downstream analyses on aggregated data instead of single gene data was shown to increase representativeness and reduce noisiness.

Supervised and clustered heatmaps

Heatmaps of module (p.1-2) and metagene (p.3-4) expression data with supervised and hierarchically clustered sample ordering.

The figure shows a heatmap titled 'Heatmap on module data (kmeans)'. The x-axis is labeled 'Clustering heatmap on module data (kmeans)'. The y-axis is labeled 'H', 'E', 'C', 'D', 'B', 'F', 'G', 'A'. The color scale ranges from -3 (blue) to 3 (red). The heatmap shows a clear pattern of expression across the samples.