COPA plugin

Note: this analysis can be performed “per gene” or “per reporter ID”. To do either or the user will have to merge on genesymbol or similar BEFORE running this plugin. Below, when referring to genes I actually refer to positions (in BioAssaySet). In tables below, e.g. section 5 each row should be annotated with three columns; reporter ID, GeneSymbol, and LocusLink.

Note2: Implement section 1-6 first. Plotting and report functions (7-10) can come later.

COPA:

1 - Median center gene expression values (across assays), setting each gene’s median expression value to zero, i.e. for each gene calculate median value (across assays) and from each expression value subtract the median value (for corresponding gene). Have this step be an option (yes, no) so that users can choose to perform the centering separately if desired.

2 - Calculate the median absolute deviation (MAD) using the centered gene expression data.

$$MAD = median(|Y_i - \tilde{Y}|)$$

where $\tilde{Y}$ is the median of the data (across assays) and $|\boldsymbol{Y}|$ is the absolute value of $\boldsymbol{Y}$.

3 - Scale MAD to 1 by dividing each gene expression value by its MAD. For each gene we have now created a set of transformed expression data (across assays). Create a new BioAssaySet using this data. Have this as an option, e.g. Create new BioAssaySet, Yes/No.

4 - For each gene get the 75th, 90th, and 95th percentile score, i.e. value, from the transformed expression values.

5 - Tabulate percentile score for all genes. Sort descending on 95th percentile score (attach this table as a tab separated list)
i.e.

	75th	90th	95th
GeneA	0.45	2.1	10.5
GeneB	0.56	2.4	2.8
GeneC	0.67	2.2	2.6

..
..

6 - Create reporter list

7 - Create report/plots

8 - Plot transformed expression data for assays (assays sorted acending), i.e., one plot per gene (see fig 1)

a) NOTE: do this for top 10 (setting?) genes.
b) Do it for all selected reporter ID alternative genes (setting i.e., H2000045|H3000352, or BRCA1|ESR1).

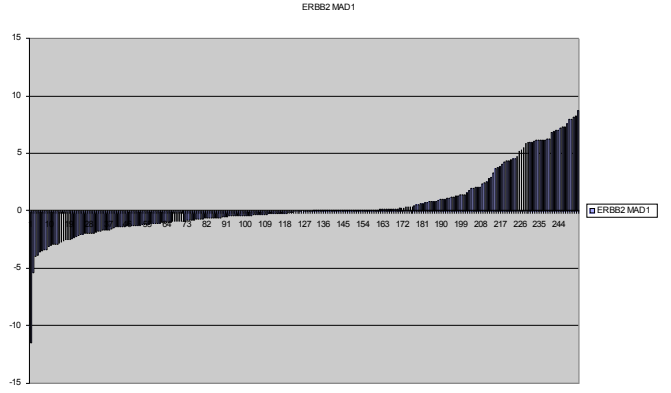

Fig1

9 - Plot 70th, 90th, and 95th percentile for genes (genes sorted acending), i.e., similar to fig 1 but one plot for entire BioAssaysSet including all genes.

10 - Annotation plots, i.e., redo plots from section 8, but separated on selected annotation (if selected), e.g. settings annotation type = s_er_pos_neg, annotation values = er_pos|er_neg.
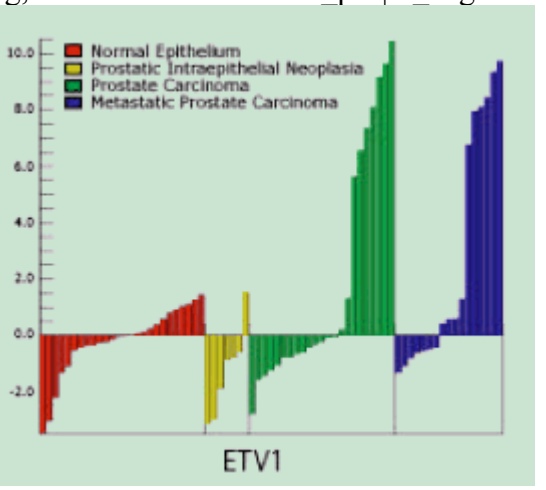
Fig2