

I. Detailed Procedure:

A. General Workflow Diagram

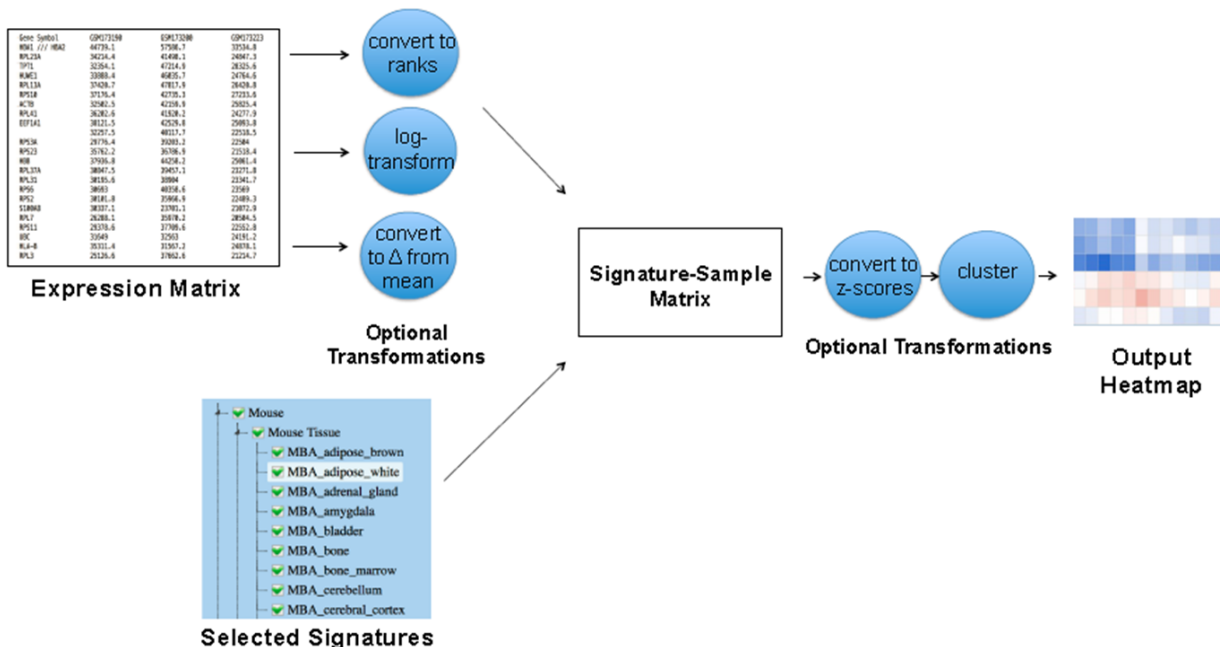


Figure 2. SaVant Pipeline

In the first step, an expression matrix containing values for genes in several samples is optionally converted to ranked lists of genes in samples and/or log-transformed. The expression matrix is then converted into a signature-sample matrix as described in Figure 1 using the selected signatures. Optionally, the signature-sample matrix is converted to differences from mean values, converted to z-scores, and/or clustered to produce a final heatmap.

B. Programs required (and how to install/build in Hoffman)

- None

C. Protocol:

1. Sample expression matrix format:

- Can be microarray or RNA sequencing data (properly normalized between all samples) Normalization can be done via RMA, fRMA, etc, depending on needs of the experiment.
- Tab-delimited file format (*.txt)
-

Example SaVant input matrix

SYMBOL	Sample1	Sample2	Sample3	Sample4	Sample...
SAVANT_GROUP	1	1	2	2	...
RFC2	7.0	7.0	6.7	6.9	...
HSPA6	7.8	6.9	8.5	7.2	...
PAX8	8.2	8.4	8.4	8.5	...

*This row optional

...
-----	-----	-----	-----	-----	-----

- d. Rows = genes in GENE SYMBOL format **make sure not to have duplicates, collapse redundant probes in to one GENE symbol beforehand. 1st row is column header for samples. 2nd row would be "SAVANT_GROUP" if ANOVA analysis desired (optional, see II.f.vii). For SAVANT_GROUP, label each group member with the same integer, starting with the number 1.
 - e. Columns= values for each sample, can be a microarray intensity value, sequencing count, DESEQ normalized count, or any ranking value of the genes. 1st column is gene symbol column.
2. Selecting SaVant Settings
- a. **Upload a gene expression matrix.** Click "Browse" and select your input tab-delimited input file as formatted above. Select which species your samples are from.
 - b. **Matrix Transformation Options. (Can select more than one)**
 - i. *Convert matrix to ranks.* Will rank genes for each sample independently based on the values in the matrix. Not as effective as using the actual gene expression values. Better for use for samples in which the normalization is very good or where the metric used for each gene is not related to the gene amount (ie. correlation value).
 - ii. *Log-transform matrix.* Not to be used with rank transformation above. Necessary if your expression values are not log-transformed so that they will have a normal distribution. Distribution of values can be checked after analysis by clicking "Distributions of values in input matrix samples" in output webpage.
 - iii. *"Delta":* Transform values to average difference from mean (for each gene). This setting helps to see differences between samples by dividing each gene's value by its average across all samples. Can be combined with either log-transformed or ranks transformations.
 - c. **Select Signatures.** Choose the signatures that you would like to visualize. Mouse or Human signatures can be chosen regardless if the input was mouse or human as SaVant will convert the gene nomenclature.
 - i. *Ranked Signatures:* These gene signatures have been ranked in terms of their specificity as described in SaVant publication (usually proportional median).

1. *Organize by type/datasets*: select to change organization of signatures (doesn't affect results)
 2. *Number of genes per signatures*. Choose number of genes to be used for signature visualization. Default and best choice usually is 50.
- ii. Unranked Signatures: These signatures are derived from the MsigDB database and no information as to which genes are more specific to signature is available. Usually all genes in the signature and this cannot be changed.
- d. **Clustering Options**. Select how and if you want your samples or signatures to be clustered in the output heatmap.
- e. **Statistical Options**.
- i. *Compute null distribution*. This selection will allow a p-value to be calculated for the signature score within your sample, using a permutation analysis. Default 10,000 iterations. Signature score heatmap will still be output and p-value will be seen with mouse over the particular cell.
 - ii. *Visualize p-values*. This selection will completely change heatmap output to the p-values and NOT the signature scores. Use this to determine which signatures may be significant in your signature.
- f. **Display options**.
- i. *Normalize output matrix to z-scores*. Will transform the entire heatmap matrix values (including across signatures and samples) in a z-score, then output the z-score heatmap.
 - ii. *Remove signatures with less than XX genes in submitted data*. Use this setting if you want to make sure that you have a sufficient number of genes used in each signature values calculation. Can be useful if your data does not have most of the genes represented from your organism.
 - iii. *Scale signature values by square root of signature genes*
 - iv. *Set color axes range*
 - v. *Report average value across signature instead of sum*
 - vi. *Only show signatures with min value*. Use this setting if you want to restrict the signatures analyzed to those that have a higher score in your output. Can be useful to filter out weak signatures in your dataset.
 - vii. *Only show signatures with a significant ANOVA p-value*. This setting will filter signatures to only those that have significantly different SaVant scores between the groups you indicated in your input matrix. (see "SAVANT_GROUP").

D. SaVant Output

E. Quality Control

- SaVant web program will tell you if there are issues with the formatting of your input file
- Select “distributions of values in input matrix” to see if expression values are normally distributed
- Select signature of interest to determine genes that are driving the signature score, ensure that there are many genes being used and are genes relevant to your study.

F. Settings/problems to avoid

- No pre-filtering of genes necessary other than to ensure there are no duplicate gene symbols.
- Be sure to check how many genes are used to measure your gene signatures
- Ideally do not overload SaVant heatmap analysis with more than 100 signatures (particularly possible with MsigDB signatures).

G. References

- <http://pathways.mcdb.ucla.edu/savant/>
- Version number July 2016
- Author of SCP contact info: Dennis Montoya (dmontoya@ucla.edu)
- Date of SCP version #7/5/16