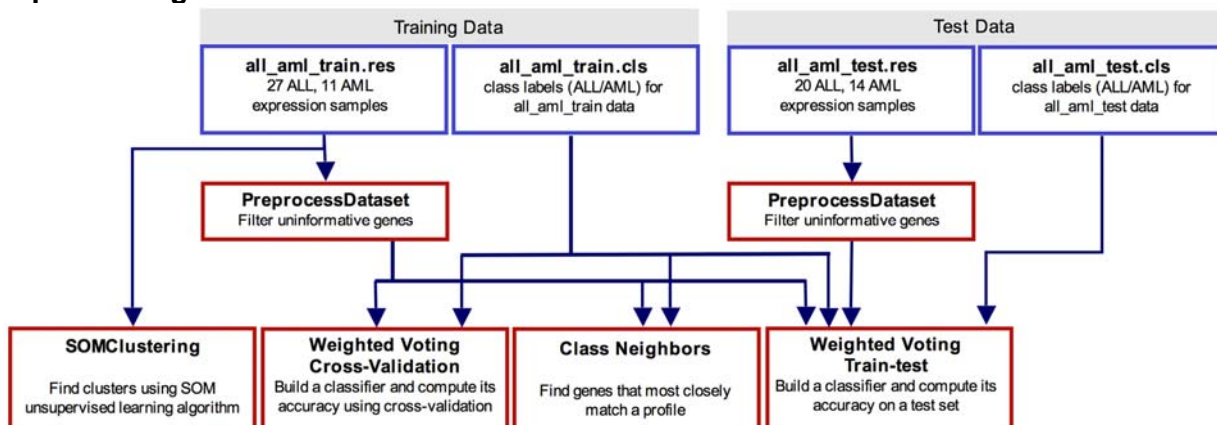


Golub.Slonim.1999.Science.all.aml

Module name: Golub.Slonim.1999.Science.all.aml
Description: Reproduce selected analyses from Golub and Slonim *et al.*, 1999
Author: GenePattern (Broad Institute),
 gp-help@broad.mit.edu

Summary: This pipeline reproduces several steps in the analysis done by Golub and Slonim in the molecular profiling and characterization of two leukemia subtypes, ALL and AML.

Pipeline Diagram:



References:

- Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D. & Lander, E. S. (1999) Molecular Classification of Cancer: Class Discovery and Class Prediction by Gene Expression Monitoring. *Science* 286, 531-7.
- http://www.broad.mit.edu/cancer/pub/all_aml

Modules:

1. PreprocessDataset

Perform several preprocessing options on a RES, GCT, or ODF input file

input filename	ftp://ftp.broad.mit.edu/pub/genepattern/all_aml/all_aml_train.res
output file	all_aml_train_filt
output file format	res
filter flag	filter
preprocessing flag	no disc or norm
minchange	5
mindelta	100
threshold	20
ceiling	16000
max sigma binning	1
prob thres	1
num excl	0

GenePattern

2. PreprocessDataset

Perform several preprocessing options on a RES, GCT, or ODF input file

input filename	ftp://ftp.broad.mit.edu/pub/genepattern/all_aml/all_aml_test.res
output file	all_aml_test_filt
output file format	res
filter flag	filter
preprocessing flag	no disc or norm
minchange	5
mindelta	100
threshold	20
ceiling	16000
max sigma binning	1
prob thres	1
num excl	0

3. SOMClustering

Self-Organizing Maps algorithm

dataset filename	Use 1st output from 1. PreprocessDataset
output stub	clustered_train_set
cluster range	2
seed range	42
iterations	50000
cluster by	columns
som rows	0
som cols	0
initialization	Random_Vectors
neighborhood	Bubble
alpha initial	0.1
alpha final	0.005
sigma initial	5.0
sigma final	0.5

4. SOMClusterViewer

Visualize clusters created with the SOM algorithm

som cluster filename	Use 1st output from 3. SOMClustering
----------------------	--------------------------------------

5. ClassNeighbors

Select genes that most closely resemble a profile

data filename	Use 1st output from 1. PreprocessDataset
class filename	ftp://ftp.broad.mit.edu/pub/genepattern/all_aml/all_aml_train.cls
marker gene list file	marker_results
marker data set file	data_results
num neighbors	2000
num permutations	250
user pval	0.5
mean or median	mean
ttest or snr	SNR
filter data	no

GenePattern

min threshold	10
max threshold	16000
min fold diff	5
min abs diff	50

6. GeneListSignificanceViewer

Views the results of marker analysis

Inputfilename	Use 1st output from 5. ClassNeighbors
---------------	---------------------------------------

7. WeightedVotingXValidation

Weighted Voting classification with Leave-One-Out Cross-Validation

data filename	Use 1st output from 1. PreprocessDataset
class filename	ftp://ftp.broad.mit.edu/pub/genepattern/all_aml/all_aml_train.cls
output file	all_aml_wv_xval
num features	50
filter data	no
thresh min	20
thresh max	16000
absolute diff	50
fold diff	5

8. FeatureSummaryViewer

Feature Summary Screen

feature filename	Use 1st output from 7. WeightedVotingXValidation
------------------	--

9. PredictionResultsViewer

Visualize prediction results

prediction results filename	Use 2nd output from 7. WeightedVotingXValidation
-----------------------------	--

10. WeightedVoting

Weighted Voting classification

train filename	Use 1st output from 1. PreprocessDataset
train class filename	ftp://ftp.broad.mit.edu/pub/genepattern/all_aml/all_aml_train.cls
test filename	Use 1st output from 2. PreprocessDataset
test class filename	ftp://ftp.broad.mit.edu/pub/genepattern/all_aml/all_aml_test.cls
pred results file	test_pred_results
num features	50
filter data	no
thresh min	20
thresh max	16000
fold diff	5
absolute diff	50

11. PredictionResultsViewer

Visualize prediction results

prediction results filename	Use 1st output from 10. WeightedVoting
-----------------------------	--