

Supplementary Table 2. Contents of GenePattern Module Repository

Module	Type	Description
GeneCruiser	Annotation	Retrieve gene annotations for Affy probe ids
ConsensusClustering	Clustering	Resampling-based clustering method
HierarchicalClustering	Clustering	Hierarchical Clustering
HierarchicalClustering.MATLAB	Clustering	Hierarchical Clustering
NMFConsensus	Clustering	Non-negative Matrix Factorization (NMF) Consensus Clustering
SOMClustering	Clustering	Self-Organizing Maps algorithm
ClassNeighbors	Gene List Selection	Select genes that most closely resemble a profile
ComparativeMarkerSelection	Gene List Selection	Computes significance values for features using several metrics
ExtractComparativeMarkerResults	Gene List Selection	Creates a dataset and feature list from Comp.MarkerSelection output
GeneNeighbors	Gene List Selection	Select genes that most closely resemble a continuous profile
SelectFeaturesColumns	Gene List Selection	Takes a 'column slice' from a .res, .gct, .odf, or .cls file
SelectFeaturesRows	Gene List Selection	Generate a .res, .gct, or .odf file containing only specified features
HeatMapImage	Image Creators	Creates a heat map from a dataset
HierarchicalClusteringImage.MATLAB	Image Creators	Hierarchical Clustering Image
ImputeMissingValues.KNN	Missing Value Imputation	Imputes missing values
Golub.Slonim.1999.Science.all.aml	Pipeline	ALL/AML methodology, from Golub and Slonim et al., 1999
Lu.Getz.Miska.Nature.June.2005.PDT.miRNA	Pipeline	Probabilistic Neural Network Prediction using mRNA
Lu.Getz.Miska.Nature.June.2005.clustering.ALL	Pipeline	Hierarchical clustering of ALL samples with genetic alterations
Lu.Getz.Miska.Nature.June.2005.clustering.ep.miRNA	Pipeline	Hierarchical clustering of 89 epithelial samples in miRNA space
Lu.Getz.Miska.Nature.June.2005.clustering.miGCM218	Pipeline	Hierarchical clustering of 218 samples from various tissue types
Lu.Getz.Miska.Nature.June.2005.mouse.lung	Pipeline	Normal/tumor classifier and kNN prediction of mouse lung samples
KNN	Prediction	K-nearest neighbors classification
KNNXValidation	Prediction	K-nearest neighbors classification with leave-one-out cross-validation
PNN	Prediction	Probabilistic Neural Network (PNN)
PNNXValidationOptimization	Prediction	PNN Leave-One-Out Cross-Validation Optimization
SVM	Prediction	Classification and Regression Tree
WeightedVoting	Prediction	Weighted Voting classification
WeightedVotingXValidation	Prediction	Weighted Voting classification with Leave-One-Out Cross-Validation
ConvertLineEndings	Preprocess & Utilities	Converts line endings to the host operating system's format
ConvertToMAGEML	Preprocess & Utilities	Converts a gct, res, or odf dataset file to a MAGE-ML file
DownloadURL	Preprocess & Utilities	Downloads a file from a URL
ExpressionFileCreator	Preprocess & Utilities	Creates a res or gct file from a set of Affymetrix CEL files
ExtractColumnNames	Preprocess & Utilities	Lists the sample descriptors from a .res file
ExtractRowNames	Preprocess & Utilities	Extracts the row names from a .res, .gct, or .odf file
GEOImporter	Preprocess & Utilities	Imports data from the Gene Expression Omnibus (GEO)
Hu68kHu35kAtoU95	Preprocess & Utilities	Converts Affymetrix Hu6800/Hu35KsubA to Affymetrix U95Av2 probes
MapChipFeaturesGeneral	Preprocess & Utilities	Map the features of a dataset
PreprocessDataset	Preprocess & Utilities	Preprocessing options on a res, gct, or Dataset input file
ReorderByClass	Preprocess & Utilities	Reorder the samples in an expression dataset and class file by class
SplitDatasetTrainTest	Preprocess & Utilities	Splits a dataset (and cls files) into Train and Test subsets
TransposeDataset	Preprocess & Utilities	Transpose a Dataset - .res .gct, .odf
mzXMLToCSV	Preprocess & Utilities	Converts a mzXML file to a zip of csv files
NMF	Projection	Non-negative Matrix Factorization
PCA	Projection	Principal Component Analysis
AreaChange	Proteomics	Calculates fraction of area under the spectrum that is attributable to signal
CompareSpectra	Proteomics	Compares two spectra to determine similarity
LocatePeaks	Proteomics	Locates detected peaks in a spectrum.
Peaks	Proteomics	Determine peaks in the spectrum using a series of digital filters.
PlotPeaks	Proteomics	Plot peaks
ProteoArray	Proteomics	LC-MS proteomic data processing module
ProteomicsAnalysis	Proteomics	Runs the proteomics analysis on the set of input spectra
GlobalAlignment	Sequence Analysis	Smith-Waterman sequence alignment
KSScore	Statistical Methods	Kolmogorov-Smirnov score for a set of genes within an ordered list
ComparativeMarkerSelectionViewer	Visualizer	View the results from ComparativeMarkerSelection
FeatureSummaryViewer	Visualizer	Feature Summary Screen
GeneListSignificanceViewer	Visualizer	Views the results of marker analysis
HeatMapView	Visualizer	Display a heat map view of a Dataset
JavaTreeView	Visualizer	Hierarchical clustering viewer that reads in Eisen's cdt, atr, and gtr files
MAGEMLImportViewer	Visualizer	A visualizer to import data in MAGE-ML format into GenePattern
PCAViewer	Visualizer	Principal Component Analysis Visualizer
PredictionResultsViewer	Visualizer	Visualize prediction results
SOMClusterViewer	Visualizer	Visualize clusters created with the SOM algorithm