

Using MBatch Assessments: SupervisedClustering_Pairs_Structures

Tod Casasent

2023-10-06

1 Introduction

These instructions are aimed at people familiar with R and familiar with TCGA/GDC platforms and data types. They are intended to introduce the reader to producing the given assessment. These instructions will only rarely, if ever, touch on the appropriateness of the assessment algorithm or interpretation of output. See MBatch_01_InstallLinux for instructions on downloading test data.

2 Algorithm

SupervisedClustering_Pairs_Structures is a function used to perform batch effects assessments using the supervised clustering algorithm for each pair of batch types provided.

3 Output

The primary output method for MBatch is to view results in the Batch Effects Website, described elsewhere. The PNG files are rough versions of the website output.

Graphical output is a heatmap of the correlation values, topped by a covariate bar with the batch information, and at the top dendrograms for the clustering. The columns are batch values for a single batch type. The rows are sample ids.

4 Usage

SupervisedClustering_Pairs_Structures(theData, theTitle, theOutputPath, theListOfBatchPairs, theBatchTypeAndValuePairsToRemove=list(), theBatchTypeAndValuePairsToKeep=list())

5 Arguments

##theData An instance of BEA_DATA.

BEA_DATA objects can be created by calls of the form new(“BEA_DATA”, theData, theBatches, theCovariates). If you have no covariate data, use an empty data.frame created with data.frame()

mData: Object of class “matrix” A matrix where the colnames are sample ids and the rownames are gene equivalents. All names should be strings, not factors.

mBatches: Object of class “data.frame” A data.frame where the column “names” are batch types. The first batch “type” is “Sample”. All names and values should be strings, not factors or numeric.

mCovariates: Object of class “data.frame” A data.frame where the column “names” are covariate types. The first covariate “type” is “Sample”. All names and values should be strings, not factors or numeric.

##theTitle A string title to use in PNG files.

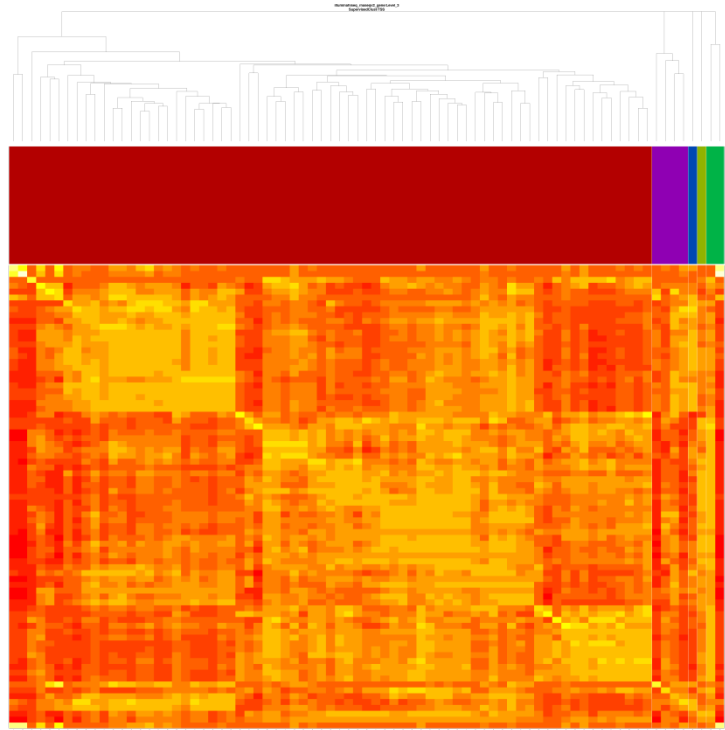


Figure 1: Supervised Clustering Example

##theOutputPath String giving directory in which to place output PNG files.

##theListOfBatchPairs

A vector of strings, where pairs of strings give batch types to use for pairs assessment.

##theBatchTypeAndValuePairsToRemove A list of vectors containing the batch type (or * for all types) and the value to remove. list() indicates none while NULL will cause an error.

##theBatchTypeAndValuePairsToKeep A list of vectors containing the batch type (or * for all types) and a vector of the the value(s) to keep. list() indicates none while NULL will cause an error.

6 Example Call

The following code is adapted from the tests/SupervisedClustering_Pairs_Structures file. Data used is from the testing data as per the MBatch_01_InstallLinux document. In the future, we plan to make the output from MBatch more user friendly, but currently, this produces the following output at the command line.

This output can generally be skipped as very long and generally obscure. After the output is an explanation of files and directories created.

```
{
  require(MBatch)

  inputDir <- getTestInputDir()
  outputDir <- getTestOutputDir()
  compareDir <- getTestCompareDir()

  # set the paths
  theGeneFile=cleanFilePath(inputDir, "matrix_data-Tumor.tsv")
```

```

theBatchFile=cleanFilePath(inputDir, "batches-Tumor.tsv")
theOutputDir=cleanFilePath(outputDir, "SupervisedClustering_Pairs_Structures")
theRandomSeed=314

# make sure the output dir exists and is empty
unlink(theOutputDir, recursive=TRUE)
dir.create(theOutputDir, showWarnings=FALSE, recursive=TRUE)

# load the data and reduce the amount of data to reduce run time
myData <- mbatchLoadFiles(theGeneFile, theBatchFile)
myData@mData <- mbatchTrimData(myData@mData, 100000)

# here, we take most defaults
SupervisedClustering_Pairs_Structures(theData=myData,
  theTitle="Test Data Title",
  theOutputDir=theOutputDir,
  theListOfBatchPairs=c("PlateId", "TSS", "BatchId", "TSS"),
  theDataVersion="DATA_2022-09-09-1600",
  theTestVersion="TEST_2022-10-10-1300",
  theBatchTypeAndValuePairsToRemove=list(),
  theBatchTypeAndValuePairsToKeep=list() )
}

```

```

## 2023 10 06 12:31:28.072 DEBUG qcprludev10 Changing LC_COLLATE to C for duration of run
## 2023 10 06 12:31:28.073 INFO qcprludev10 \ / \ / \ / \ / \ / \ / \ / \ / \ /
## 2023 10 06 12:31:28.073 INFO qcprludev10 Starting mbatchLoadFiles
## 2023 10 06 12:31:28.074 INFO qcprludev10 MBatch Version: BEA_VERSION_TIMESTAMP
## 2023 10 06 12:31:28.074 INFO qcprludev10 read batch file= /builds/BatchEffects_clean/BatchEffectsPack
## 2023 10 06 12:31:28.075 INFO qcprludev10 read gene file= /builds/BatchEffects_clean/BatchEffectsPack
## 2023 10 06 12:31:30.596 INFO qcprludev10 filter samples in batches using gene samples
## 2023 10 06 12:31:30.598 INFO qcprludev10 sort batches by gene file samples
## 2023 10 06 12:31:30.650 INFO qcprludev10 Finishing mbatchLoadFiles
## 2023 10 06 12:31:30.651 INFO qcprludev10 ~~~~~
## 2023 10 06 12:31:30.651 DEBUG qcprludev10 Changing LC_COLLATE to C for duration of run
## 2023 10 06 12:31:30.652 INFO qcprludev10 \ / \ / \ / \ / \ / \ / \ / \ / \ /
## 2023 10 06 12:31:30.652 INFO qcprludev10 mbatchTrimData Starting
## 2023 10 06 12:31:30.653 INFO qcprludev10 MBatch Version: BEA_VERSION_TIMESTAMP
## 2023 10 06 12:31:38.443 INFO qcprludev10 mbatchTrimData theMaxSize= 1e+05
## 2023 10 06 12:31:38.444 INFO qcprludev10 mbatchTrimData ncol(theMatrix)= 80
## 2023 10 06 12:31:38.444 INFO qcprludev10 mbatchTrimData nrow(theMatrix)= 1250
## 2023 10 06 12:31:38.445 INFO qcprludev10 mbatchTrimData Finishing
## 2023 10 06 12:31:38.445 INFO qcprludev10 ~~~~~
## 2023 10 06 12:31:38.447 DEBUG qcprludev10 Changing LC_COLLATE to C for duration of run
## 2023 10 06 12:31:38.447 INFO qcprludev10 \ / \ / \ / \ / \ / \ / \ / \ / \ /
## 2023 10 06 12:31:38.448 INFO qcprludev10 mbatchFilterData Starting
## 2023 10 06 12:31:38.448 INFO qcprludev10 MBatch Version: BEA_VERSION_TIMESTAMP
## 2023 10 06 12:31:38.449 DEBUG qcprludev10 rows pre filter 1250
## 2023 10 06 12:31:38.667 DEBUG qcprludev10 rows post filter 1250
## 2023 10 06 12:31:38.667 DEBUG qcprludev10 mbatchFilterData Prefilter, gene data had 1250 while pos
## 2023 10 06 12:31:38.669 DEBUG qcprludev10 mbatchFilterData Prefilter, batch data had 80 while post
## 2023 10 06 12:31:38.669 INFO qcprludev10 mbatchFilterData Finishing
## 2023 10 06 12:31:38.670 INFO qcprludev10 ~~~~~
## 2023 10 06 12:31:38.670 INFO qcprludev10 createBatchEffectsOutput_SupervisedClustering_pairs - batch
## 2023 10 06 12:31:38.671 INFO qcprludev10 createBatchEffectsOutput_SupervisedClustering_pairs - batch

```

```

## 2023 10 06 12:31:38.672 INFO qcprludev10 createBatchEffectsOutput_SupervisedClustering_pairs - pre b
## 2023 10 06 12:31:38.673 INFO qcprludev10 makeBiasClust - starting
## 2023 10 06 12:31:38.775 INFO qcprludev10 makeBiasClust - quantile dat dim = 1250,80
## 2023 10 06 12:31:38.777 INFO qcprludev10 makeBiasClust - quantile U.data is.data.frame = FALSE
## 2023 10 06 12:31:38.778 INFO qcprludev10 makeBiasClust - quantile U.data is.array = TRUE
## 2023 10 06 12:31:38.778 INFO qcprludev10 makeBiasClust - quantile U.data is.list = FALSE
## 2023 10 06 12:31:38.779 INFO qcprludev10 makeBiasClust - quantile U.data nrow = 312
## 2023 10 06 12:31:38.779 INFO qcprludev10 makeBiasClust - quantile U.data ncol = 80
## 2023 10 06 12:31:38.780 INFO qcprludev10 makeBiasClust - quantile U.data length = 24960
## 2023 10 06 12:31:38.780 INFO qcprludev10 makeBiasClust - quantile U.data dim = 312,80
## 2023 10 06 12:31:38.780 INFO qcprludev10 makeBiasClust - quantile U.data is.null = FALSE
## 2023 10 06 12:31:38.781 INFO qcprludev10 makeBiasClust - biasedDend <- biasedClusterFunction
## 2023 10 06 12:31:38.785 INFO qcprludev10 makeBiasClust new.dis size - 80-80
## 2023 10 06 12:31:38.786 INFO qcprludev10 makeBiasClust orig - 80-80
## 2023 10 06 12:31:38.786 INFO qcprludev10 makeBiasClust is.na - 80-80
## 2023 10 06 12:31:38.787 INFO qcprludev10 makeBiasClust is.infinite - 80-80
## 2023 10 06 12:31:38.788 DEBUG qcprludev10 hierclust_calc
## 2023 10 06 12:31:38.789 DEBUG qcprludev10 Changing LC_COLLATE to C for duration of run
## 2023 10 06 12:31:38.790 DEBUG qcprludev10 calculating HC
## 2023 10 06 12:31:38.800 INFO qcprludev10 writeHCDataTSVs start
## 2023 10 06 12:31:38.800 INFO qcprludev10 writeHCDataTSVs ugend
## 2023 10 06 12:31:38.801 INFO qcprludev10 writeHCDataTSVs HCData= /BEA/BatchEffectsPackage_data/testi
## 2023 10 06 12:31:38.802 INFO qcprludev10 writeHCDataTSVs HCOrder= /BEA/BatchEffectsPackage_data/test
## 2023 10 06 12:31:38.803 INFO qcprludev10 writeHCDataTSVs rdataFile= /BEA/BatchEffectsPackage_data/te
## 2023 10 06 12:31:38.804 INFO qcprludev10 writeHCDataTSVs done
## 2023 10 06 12:31:38.804 INFO qcprludev10 writeHCDataTSVs start
## 2023 10 06 12:31:38.805 INFO qcprludev10 writeHCDataTSVs ugend
## 2023 10 06 12:31:38.805 INFO qcprludev10 writeHCDataTSVs HCData= /BEA/BatchEffectsPackage_data/testi
## 2023 10 06 12:31:38.807 INFO qcprludev10 writeHCDataTSVs HCOrder= /BEA/BatchEffectsPackage_data/test
## 2023 10 06 12:31:38.808 INFO qcprludev10 writeHCDataTSVs rdataFile= /BEA/BatchEffectsPackage_data/te
## 2023 10 06 12:31:38.810 INFO qcprludev10 writeHCDataTSVs done

## 2023 10 06 12:31:39.075 INFO qcprludev10 createBatchEffectsOutput_SupervisedClustering_pairs - post l
## 2023 10 06 12:31:39.076 DEBUG qcprludev10 mbatchStandardLegend - theTitle PlateId
## 2023 10 06 12:31:39.077 DEBUG qcprludev10 mbatchStandardLegend - theVersion MBatch 2.0.3
## 2023 10 06 12:31:39.077 DEBUG qcprludev10 mbatchStandardLegend - theFilenamePath /BEA/BatchEffectsP
## 2023 10 06 12:31:39.078 DEBUG qcprludev10 mbatchStandardLegend - theLegendNames A29J (80)
## 2023 10 06 12:31:39.078 DEBUG qcprludev10 mbatchStandardLegend - theLegendNames length 1
## 2023 10 06 12:31:39.079 DEBUG qcprludev10 mbatchStandardLegend - theLegendColors #B30000
## 2023 10 06 12:31:39.079 DEBUG qcprludev10 mbatchStandardLegend - theLegendColors length 1
## 2023 10 06 12:31:39.080 DEBUG qcprludev10 mbatchStandardLegend - theLegendSymbols
## 2023 10 06 12:31:39.080 DEBUG qcprludev10 mbatchStandardLegend - theLegendSymbols length 0
## 2023 10 06 12:31:39.081 DEBUG qcprludev10 mbatchStandardLegend - myColors #b30000
## 2023 10 06 12:31:39.081 DEBUG qcprludev10 mbatchStandardLegend - theTitle UTF-8 = PlateId
## 2023 10 06 12:31:39.081 DEBUG qcprludev10 mbatchStandardLegend - theVersion UTF-8 = MBatch 2.0.3
## 2023 10 06 12:31:39.082 DEBUG qcprludev10 mbatchStandardLegend - theFilenamePath UTF-8 = /BEA/Batch
## 2023 10 06 12:31:39.082 DEBUG qcprludev10 mbatchStandardLegend before Python
## 2023 10 06 12:31:39.083 DEBUG qcprludev10 mbatchStandardLegend - getGlobalMBatchEnv() = /BEA/gendev
## 2023 10 06 12:31:39.083 DEBUG qcprludev10 mbatchStandardLegend - import(mbatch.legend.legend)
## 2023 10 06 12:31:39.084 DEBUG qcprludev10 mbatchStandardLegend - after import
## 2023 10 06 12:31:39.084 DEBUG qcprludev10 mbatchStandardLegend - after color list
## 2023 10 06 12:31:39.084 DEBUG qcprludev10 mbatchStandardLegend - after symbol list
## 2023 10 06 12:31:39.085 DEBUG qcprludev10 mbatchStandardLegend - colorList = #b30000
## 2023 10 06 12:31:39.085 DEBUG qcprludev10 mbatchStandardLegend - symbolList =

```

```

## 2023 10 06 12:31:39.086 DEBUG qcprludev10 mbatchStandardLegend - legendNameList = A29J (80)
## 2023 10 06 12:31:39.086 DEBUG qcprludev10 mbatchStandardLegend - myTitle = PlateId MBatch 2.0.3
## 2023 10 06 12:31:39.087 DEBUG qcprludev10 mbatchStandardLegend - theFilenamePath = /BEA/BatchEffectsP
## 2023 10 06 12:31:39.125 DEBUG qcprludev10 mbatchStandardLegend after Python
## 2023 10 06 12:31:39.125 INFO qcprludev10 createBatchEffectsOutput_SupervisedClustering_pairs - after
## 2023 10 06 12:31:39.127 DEBUG qcprludev10 mbatchStandardLegend - theTitle TSS
## 2023 10 06 12:31:39.127 DEBUG qcprludev10 mbatchStandardLegend - theVersion MBatch 2.0.3
## 2023 10 06 12:31:39.127 DEBUG qcprludev10 mbatchStandardLegend - theFilenamePath /BEA/BatchEffectsP
## 2023 10 06 12:31:39.128 DEBUG qcprludev10 mbatchStandardLegend - theLegendNames OR - University of M
## (72), OU - Roswell Park (1), P6 - Translational Genomics
## Research Institute (2), PA - University of Minnesota
## (1), PK - University Health
## Network (4)
## 2023 10 06 12:31:39.128 DEBUG qcprludev10 mbatchStandardLegend - theLegendNames length 5
## 2023 10 06 12:31:39.129 DEBUG qcprludev10 mbatchStandardLegend - theLegendColors #B30000, #8FB300, #
## 2023 10 06 12:31:39.129 DEBUG qcprludev10 mbatchStandardLegend - theLegendColors length 5
## 2023 10 06 12:31:39.129 DEBUG qcprludev10 mbatchStandardLegend - theLegendSymbols
## 2023 10 06 12:31:39.130 DEBUG qcprludev10 mbatchStandardLegend - theLegendSymbols length 0
## 2023 10 06 12:31:39.130 DEBUG qcprludev10 mbatchStandardLegend - myColors #b30000,#8fb300,#00b347,#
## 2023 10 06 12:31:39.130 DEBUG qcprludev10 mbatchStandardLegend - theTitle UTF-8 = TSS
## 2023 10 06 12:31:39.131 DEBUG qcprludev10 mbatchStandardLegend - theVersion UTF-8 = MBatch 2.0.3
## 2023 10 06 12:31:39.131 DEBUG qcprludev10 mbatchStandardLegend - theFilenamePath UTF-8 = /BEA/BatchE
## 2023 10 06 12:31:39.132 DEBUG qcprludev10 mbatchStandardLegend before Python
## 2023 10 06 12:31:39.132 DEBUG qcprludev10 mbatchStandardLegend - getGlobalMBatchEnv() = /BEA/gendev
## 2023 10 06 12:31:39.132 DEBUG qcprludev10 mbatchStandardLegend - import(mbatch.legend.legend)
## 2023 10 06 12:31:39.133 DEBUG qcprludev10 mbatchStandardLegend - after import
## 2023 10 06 12:31:39.133 DEBUG qcprludev10 mbatchStandardLegend - after color list
## 2023 10 06 12:31:39.134 DEBUG qcprludev10 mbatchStandardLegend - after symbol list
## 2023 10 06 12:31:39.134 DEBUG qcprludev10 mbatchStandardLegend - colorList = #b30000, mbatchStandar
## 2023 10 06 12:31:39.134 DEBUG qcprludev10 mbatchStandardLegend - symbolList =
## 2023 10 06 12:31:39.135 DEBUG qcprludev10 mbatchStandardLegend - legendNameList = OR - University o
## (72), mbatchStandardLegend - legendNameList = OU - Roswell Park (1), mbatchStandardLegend - legendNa
## Research Institute (2), mbatchStandardLegend - legendNameList = PA - University of Minnesota
## (1), mbatchStandardLegend - legendNameList = PK - University Health
## Network (4)
## 2023 10 06 12:31:39.135 DEBUG qcprludev10 mbatchStandardLegend - myTitle = TSS MBatch 2.0.3
## 2023 10 06 12:31:39.135 DEBUG qcprludev10 mbatchStandardLegend - theFilenamePath = /BEA/BatchEffectsP
## 2023 10 06 12:31:39.257 DEBUG qcprludev10 mbatchStandardLegend after Python
## 2023 10 06 12:31:39.258 INFO qcprludev10 createBatchEffectsOutput_SupervisedClustering_pairs - after
## 2023 10 06 12:31:39.258 INFO qcprludev10 createBatchEffectsOutput_SupervisedClustering_pairs - batch
## 2023 10 06 12:31:39.259 INFO qcprludev10 createBatchEffectsOutput_SupervisedClustering_pairs - batch
## 2023 10 06 12:31:39.260 INFO qcprludev10 createBatchEffectsOutput_SupervisedClustering_pairs - pre b
## 2023 10 06 12:31:39.260 INFO qcprludev10 makeBiasClust - starting
## 2023 10 06 12:31:39.362 INFO qcprludev10 makeBiasClust - quantile dat dim = 1250,80
## 2023 10 06 12:31:39.363 INFO qcprludev10 makeBiasClust - quantile U.data is.data.frame = FALSE
## 2023 10 06 12:31:39.364 INFO qcprludev10 makeBiasClust - quantile U.data is.array = TRUE
## 2023 10 06 12:31:39.364 INFO qcprludev10 makeBiasClust - quantile U.data is.list = FALSE
## 2023 10 06 12:31:39.365 INFO qcprludev10 makeBiasClust - quantile U.data nrow = 312
## 2023 10 06 12:31:39.365 INFO qcprludev10 makeBiasClust - quantile U.data ncol = 80
## 2023 10 06 12:31:39.366 INFO qcprludev10 makeBiasClust - quantile U.data length = 24960
## 2023 10 06 12:31:39.366 INFO qcprludev10 makeBiasClust - quantile U.data dim = 312,80
## 2023 10 06 12:31:39.366 INFO qcprludev10 makeBiasClust - quantile U.data is.null = FALSE
## 2023 10 06 12:31:39.367 INFO qcprludev10 makeBiasClust - biasedDend <- biasedClusterFunction
## 2023 10 06 12:31:39.370 INFO qcprludev10 makeBiasClust new.dis size - 80-80

```

```

## 2023 10 06 12:31:39.371 INFO qcprludev10 makeBiasClust orig - 80-80
## 2023 10 06 12:31:39.371 INFO qcprludev10 makeBiasClust is.na - 80-80
## 2023 10 06 12:31:39.372 INFO qcprludev10 makeBiasClust is.infinite - 80-80
## 2023 10 06 12:31:39.373 DEBUG qcprludev10 hierclust_calc
## 2023 10 06 12:31:39.373 DEBUG qcprludev10 Changing LC_COLLATE to C for duration of run
## 2023 10 06 12:31:39.374 DEBUG qcprludev10 calculating HC
## 2023 10 06 12:31:39.384 INFO qcprludev10 writeHCDataTSVs start
## 2023 10 06 12:31:39.384 INFO qcprludev10 writeHCDataTSVs udend
## 2023 10 06 12:31:39.385 INFO qcprludev10 writeHCDataTSVs HCData= /BEA/BatchEffectsPackage_data/testi
## 2023 10 06 12:31:39.386 INFO qcprludev10 writeHCDataTSVs HCOOrder= /BEA/BatchEffectsPackage_data/test
## 2023 10 06 12:31:39.387 INFO qcprludev10 writeHCDataTSVs rdataFile= /BEA/BatchEffectsPackage_data/te
## 2023 10 06 12:31:39.388 INFO qcprludev10 writeHCDataTSVs done
## 2023 10 06 12:31:39.389 INFO qcprludev10 writeHCDataTSVs start
## 2023 10 06 12:31:39.389 INFO qcprludev10 writeHCDataTSVs udend
## 2023 10 06 12:31:39.389 INFO qcprludev10 writeHCDataTSVs HCData= /BEA/BatchEffectsPackage_data/testi
## 2023 10 06 12:31:39.391 INFO qcprludev10 writeHCDataTSVs HCOOrder= /BEA/BatchEffectsPackage_data/test
## 2023 10 06 12:31:39.392 INFO qcprludev10 writeHCDataTSVs rdataFile= /BEA/BatchEffectsPackage_data/te
## 2023 10 06 12:31:39.394 INFO qcprludev10 writeHCDataTSVs done

## 2023 10 06 12:31:39.642 INFO qcprludev10 createBatchEffectsOutput_SupervisedClustering_pairs - post l
## 2023 10 06 12:31:39.643 DEBUG qcprludev10 mbatchStandardLegend - theTitle BatchId
## 2023 10 06 12:31:39.643 DEBUG qcprludev10 mbatchStandardLegend - theVersion MBatch 2.0.3
## 2023 10 06 12:31:39.644 DEBUG qcprludev10 mbatchStandardLegend - theFilenamePath /BEA/BatchEffectsP
## 2023 10 06 12:31:39.644 DEBUG qcprludev10 mbatchStandardLegend - theLegendNames 00304 (80)
## 2023 10 06 12:31:39.645 DEBUG qcprludev10 mbatchStandardLegend - theLegendNames length 1
## 2023 10 06 12:31:39.645 DEBUG qcprludev10 mbatchStandardLegend - theLegendColors #B30000
## 2023 10 06 12:31:39.646 DEBUG qcprludev10 mbatchStandardLegend - theLegendColors length 1
## 2023 10 06 12:31:39.646 DEBUG qcprludev10 mbatchStandardLegend - theLegendSymbols
## 2023 10 06 12:31:39.646 DEBUG qcprludev10 mbatchStandardLegend - theLegendSymbols length 0
## 2023 10 06 12:31:39.647 DEBUG qcprludev10 mbatchStandardLegend - myColors #b30000
## 2023 10 06 12:31:39.647 DEBUG qcprludev10 mbatchStandardLegend - theTitle UTF-8 = BatchId
## 2023 10 06 12:31:39.648 DEBUG qcprludev10 mbatchStandardLegend - theVersion UTF-8 = MBatch 2.0.3
## 2023 10 06 12:31:39.648 DEBUG qcprludev10 mbatchStandardLegend - theFilenamePath UTF-8 = /BEA/Batch
## 2023 10 06 12:31:39.649 DEBUG qcprludev10 mbatchStandardLegend before Python
## 2023 10 06 12:31:39.649 DEBUG qcprludev10 mbatchStandardLegend - getGlobalMBatchEnv() = /BEA/gendev
## 2023 10 06 12:31:39.649 DEBUG qcprludev10 mbatchStandardLegend - import(mbatch.legend.legend)
## 2023 10 06 12:31:39.650 DEBUG qcprludev10 mbatchStandardLegend - after import
## 2023 10 06 12:31:39.650 DEBUG qcprludev10 mbatchStandardLegend - after color list
## 2023 10 06 12:31:39.651 DEBUG qcprludev10 mbatchStandardLegend - after symbol list
## 2023 10 06 12:31:39.651 DEBUG qcprludev10 mbatchStandardLegend - colorList = #b30000
## 2023 10 06 12:31:39.652 DEBUG qcprludev10 mbatchStandardLegend - symbolList =
## 2023 10 06 12:31:39.652 DEBUG qcprludev10 mbatchStandardLegend - legendNameList = 00304 (80)
## 2023 10 06 12:31:39.652 DEBUG qcprludev10 mbatchStandardLegend - myTitle = BatchId MBatch 2.0.3
## 2023 10 06 12:31:39.653 DEBUG qcprludev10 mbatchStandardLegend - theFilenamePath = /BEA/BatchEffect
## 2023 10 06 12:31:39.689 DEBUG qcprludev10 mbatchStandardLegend after Python
## 2023 10 06 12:31:39.690 INFO qcprludev10 createBatchEffectsOutput_SupervisedClustering_pairs - after
## 2023 10 06 12:31:39.691 DEBUG qcprludev10 mbatchStandardLegend - theTitle TSS
## 2023 10 06 12:31:39.692 DEBUG qcprludev10 mbatchStandardLegend - theVersion MBatch 2.0.3
## 2023 10 06 12:31:39.692 DEBUG qcprludev10 mbatchStandardLegend - theFilenamePath /BEA/BatchEffectsP
## 2023 10 06 12:31:39.693 DEBUG qcprludev10 mbatchStandardLegend - theLegendNames OR - University of M
## (72), OU - Roswell Park (1), P6 - Translational Genomics
## Research Institute (2), PA - University of Minnesota
## (1), PK - University Health
## Network (4)

```

```

## 2023 10 06 12:31:39.693 DEBUG qcprludev10 mbatchStandardLegend - theLegendNames length 5
## 2023 10 06 12:31:39.694 DEBUG qcprludev10 mbatchStandardLegend - theLegendColors #B30000, #8FB300, #
## 2023 10 06 12:31:39.694 DEBUG qcprludev10 mbatchStandardLegend - theLegendColors length 5
## 2023 10 06 12:31:39.694 DEBUG qcprludev10 mbatchStandardLegend - theLegendSymbols
## 2023 10 06 12:31:39.695 DEBUG qcprludev10 mbatchStandardLegend - theLegendSymbols length 0
## 2023 10 06 12:31:39.695 DEBUG qcprludev10 mbatchStandardLegend - myColors #b30000,#8fb300,#00b347,#
## 2023 10 06 12:31:39.696 DEBUG qcprludev10 mbatchStandardLegend - theTitle UTF-8 = TSS
## 2023 10 06 12:31:39.696 DEBUG qcprludev10 mbatchStandardLegend - theVersion UTF-8 = MBatch 2.0.3
## 2023 10 06 12:31:39.697 DEBUG qcprludev10 mbatchStandardLegend - theFilenamePath UTF-8 = /BEA/BatchEffect
## 2023 10 06 12:31:39.697 DEBUG qcprludev10 mbatchStandardLegend before Python
## 2023 10 06 12:31:39.697 DEBUG qcprludev10 mbatchStandardLegend - getGlobalMBatchEnv() = /BEA/gendev
## 2023 10 06 12:31:39.698 DEBUG qcprludev10 mbatchStandardLegend - import(mbatch.legend.legend)
## 2023 10 06 12:31:39.698 DEBUG qcprludev10 mbatchStandardLegend - after import
## 2023 10 06 12:31:39.699 DEBUG qcprludev10 mbatchStandardLegend - after color list
## 2023 10 06 12:31:39.699 DEBUG qcprludev10 mbatchStandardLegend - after symbol list
## 2023 10 06 12:31:39.700 DEBUG qcprludev10 mbatchStandardLegend - colorList = #b30000, mbatchStandard
## 2023 10 06 12:31:39.700 DEBUG qcprludev10 mbatchStandardLegend - symbolList =
## 2023 10 06 12:31:39.700 DEBUG qcprludev10 mbatchStandardLegend - legendNameList = OR - University of
## (72), mbatchStandardLegend - legendNameList = OU - Roswell Park (1), mbatchStandardLegend - legendName
## Research Institute (2), mbatchStandardLegend - legendNameList = PA - University of Minnesota
## (1), mbatchStandardLegend - legendNameList = PK - University Health
## Network (4)
## 2023 10 06 12:31:39.701 DEBUG qcprludev10 mbatchStandardLegend - myTitle = TSS MBatch 2.0.3
## 2023 10 06 12:31:39.701 DEBUG qcprludev10 mbatchStandardLegend - theFilenamePath = /BEA/BatchEffect
## 2023 10 06 12:31:39.820 DEBUG qcprludev10 mbatchStandardLegend after Python
## 2023 10 06 12:31:39.821 INFO qcprludev10 createBatchEffectsOutput_SupervisedClustering_pairs - after

```

7 Example File Output

The above code creates the following subdirectories and files. The subdirectories correspond to the Batch Type Pairs on which assessments were requested.

```

/output/SupervisedClustering_Pairs_Structures$ ls -l
total 8
drwxr-xr-x 2 linux linux 4096 Jun 14 12:56 BatchId-TSS
drwxr-xr-x 2 linux linux 4096 Jun 14 12:56 PlateId-TSS

```

Looking at the “BatchId-TSS” subdirectory, it contains the following diagram and legend files. This algorithm does not currently generate data usable with dynamic displays.

```

/output/SupervisedClustering_Pairs_Structures/BatchId-TSS$ ls -l
total 276
-rw-r--r-- 1 linux linux 261168 Jun 19 09:58 SupervisedClust_Diagram.png
-rw-r--r-- 1 linux linux 2701 Jun 19 09:58 SupervisedClust_Legend-BatchId.png
-rw-r--r-- 1 linux linux 12899 Jun 19 09:58 SupervisedClust_Legend-TSS.png

```

Here is the diagram generated from this code.

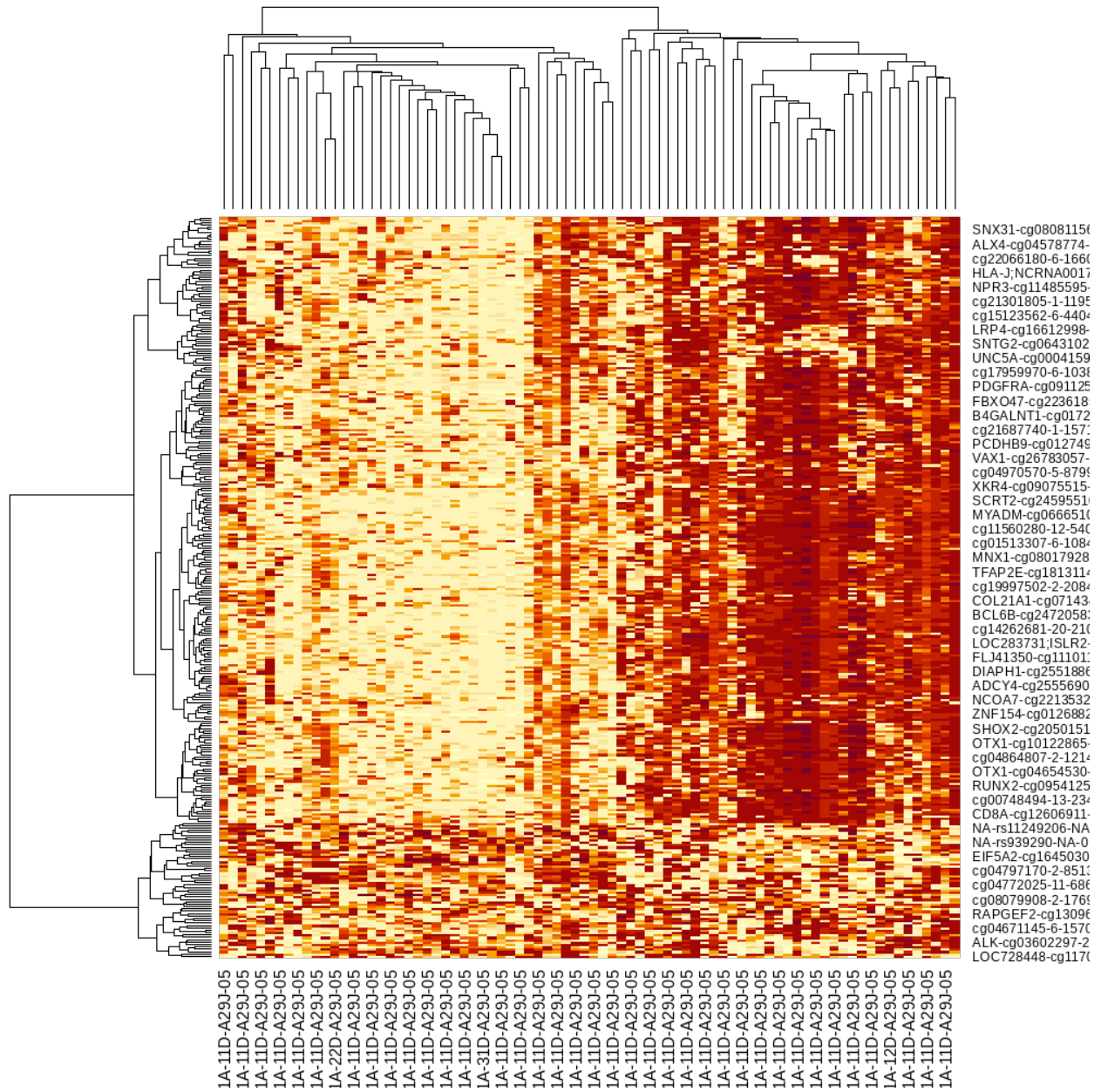


Figure 2: SupervisedClustering_Pairs_Structures Output