

Using MBatch Assessments: PCA_DualBatch_Structures

Tod Casasent

2019-10-10

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

PCA_Regular_Structures is a function used to perform batch effects assessments using the PCA-Plus algorithm, which adds centroids to the class PCA diagram. PCA-Plus is performed on each batch type available by default.

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 PCA-Plus plot, which is a PCA diagram with centroids marked and a DSC value. For information, see <http://bioinformatics.mdanderson.org/main/TCGABatchEffects:Overview>.

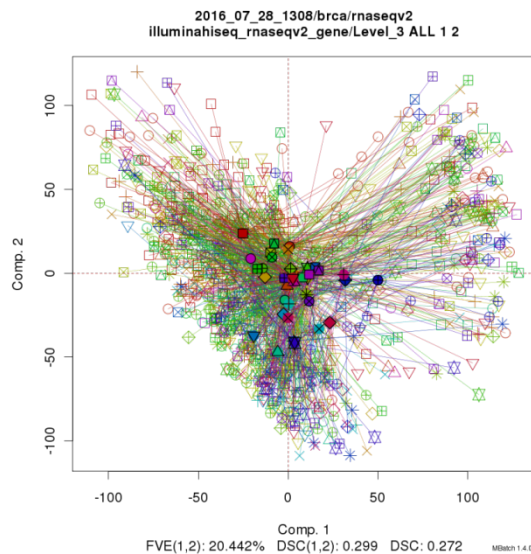


Figure 1: PCAPlus Example

4 Usage

`PCA_DualBatch_Structures(theData, theTitle, theOutputPath, theBatchTypeAndValuePairsToRemove, theBatchTypeAndValuePairsToKeep, theListForDoCentroidDualBatchType, theIsPcaTrendFunction=function(...) {FALSE}, theDoDSCFlag=TRUE, theDoDscPermsFileFlag=FALSE, theDoSampleLocatorFlag=TRUE, theListOfComponentsToPlot=c(1, 2, 1, 3, 1, 4, 2, 3, 2, 4, 3, 4), theDSCPermutations=2000, theDSCThreads=1, theMinBatchSize=2, theJavaParameters="-Xms1200m", theSeed=NULL, theMaxGeneCount=20000)`

5 Arguments

5.1 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.

5.2 theTitle

A string title to use in PNG files.

5.3 theOutputPath

String giving directory in which to place output PNG files.

5.4 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.

5.5 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.

5.6 theListForDoCentroidDualBatchType

A vector of strings. Each pair of strings should be batch types and are used to generate dual-batch plots.

5.7 theIsPcaTrendFunction

This function is used by the PCA Centroids algorithm to determine if trending lines should be added. If the function returns true, a line is drawn between the centroids, using the string sorted order. This means, for date or time fields, the batch ids should be formatted such that when sorted as strings, the batch ids are in chronological order from oldest to newest. The function receives arguments theBatchTypeName (a string identifying the batch type) and theListOfBatchIds (a list of strings, which are the batch ids). Default to returning false.

In this sample function, it checks if the batch type is ShipDate. `isTrendBatch<-function(theBatchTypeName, theListOfBatchIds) { return(is.element(theBatchTypeName, c("ShipDate"))) }`

5.8 theDoDSCFlag

A flag indicating whether or not to perform the DSC calculation for PCA-Plus plots. TRUE means perform DSC calculation.

5.9 theDoDscPermsFileFlag

A flag indicating whether or not to create the ANY_DSCPerms.txt and ANY_CompX_CompY_DSCPerms.txt files, which contains contain DSC, Db, Dw, a list of DSC, list of Dw, and list of Db for each permutation. File format consists of six lines per permutation, each line of the format "...". The labels provided are DSC, Db, Dw, DSCList, DwList and DbList.

5.10 theDoSampleLocatorFlag

A flag indicating whether or not to create the ANY_Scores.txt file. File format tab delimited with a header line of "Component Number" followed by sample ids. The first column of subsequent rows has PCA N with N replaced with a numeric. Other cells contain the PCA scores.

5.11 theListOfComponentsToPlot

A vector of integers. Each pair of integers is used to create a PCA plot of those two components. Defaults to `c(1, 2, 1, 3, 1, 4, 2, 3, 2, 4, 3, 4)`

5.12 theDSCPermutations

The number of permutations to perform for DSC calculations. Defaults to 2000

5.13 theDSCThreads

The number of threads (or CPUs or cores) to use to perform DSC calculations. Defaults to 1. Generally limit this to at most the number available minus 1, particularly on HPC systems.

5.14 theMinBatchSize

Minimum batch to allow in PCA calculations. Defaults to 2.

5.15 theJavaParameters

Command line arguments to Java for DSC calculations—for example if memory needs to be more than “-Xms1200m”.

5.16 theSeed

If non-zero, a seed to use when calling sample to select genes used in PCA plots. Default 0.

5.17 theMaxGeneCount

If non-zero, maximum number of genes allowed for PCA plots. If count is greater than this, sample is used to select a subset of genes. Default 20,000.

6 Example Call

If calling this function more than once, either change the output directory or call the clearDSCOverviewFiles function, to clear temporary RData files.

The following code is adapted from the tests/PCA_DualBatch_Structures.R 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.

```
{
  library(MBatch)

  # set the paths
  theGeneFile="/bea_testing/MATRIX_DATA/matrix_data-Tumor.tsv"
  theBatchFile="/bea_testing/MATRIX_DATA/batches-Tumor.tsv"
  theOutputDir="/bea_testing/output/PCA_DualBatch_Structures"
  theRandomSeed=314

  # make sure the output dir exists and is empty
  unlink(theOutputDir, recursive=TRUE)
  dir.create(theOutputDir, showWarnings=FALSE, recursive=TRUE)
  # load data
  myData <- mbatchLoadFiles(theGeneFile, theBatchFile)
  myData@mData <- mbatchTrimData(myData@mData, 100000)
  # trend function to handle time data
  isTrendBatch<-function(theBatchTypeName, theListOfBatchIds)
  {
    return(is.element(theBatchTypeName, c("ShipDate")))
  }
  ###
  # do two plots PlateId versus TSS and BatchId versus TSS
  PCA_DualBatch_Structures(myData,
    "Example Title for Test Data", theOutputDir, list(), list(),
    theIsPcaTrendFunction=isTrendBatch,
    theListForDoCentroidDualBatchType=c("PlateId", "TSS", "BatchId", "TSS"),
```

```

    theDoDSCFlag=TRUE, theDoDscPermsFileFlag=TRUE, theDoSampleLocatorFlag=TRUE,
    theListOfComponentsToPlot=c(1, 2, 1, 3), theDSCPermutations=100,
    theDSCThreads=5, theMinBatchSize=0,
    theJavaParameters="-Xms1000m", theSeed=0, theMaxGeneCount=0)
print(dir(theOutputDir, recursive=TRUE))
}

```

```
## All sorting in this package requires using a Sys.setlocale("LC_COLLATE","C"). MBatch Version: 2019-09-04-1100
```

```

## 2019 10 10 11:15:45.423 DEBUG megazone23 Changing LC_COLLATE to C for duration of run
## 2019 10 10 11:15:45.424 INFO megazone23 \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/
## 2019 10 10 11:15:45.424 INFO megazone23 Starting mbatchLoadFiles
## 2019 10 10 11:15:45.424 INFO megazone23 MBatch Version: 2019-09-04-1100
## 2019 10 10 11:15:45.424 INFO megazone23 read batch file= /bea_testing/MATRIX_DATA/batches-Tumor.tsv
## 2019 10 10 11:15:45.436 INFO megazone23 read gene file= /bea_testing/MATRIX_DATA/matrix_data-Tumor.tsv
## 2019 10 10 11:15:50.902 INFO megazone23 filter samples in batches using gene samples
## 2019 10 10 11:15:50.903 INFO megazone23 sort batches by gene file samples
## 2019 10 10 11:15:50.978 INFO megazone23 Finishing mbatchLoadFiles
## 2019 10 10 11:15:50.978 INFO megazone23 ~~~~~
## 2019 10 10 11:15:50.978 DEBUG megazone23 Changing LC_COLLATE to C for duration of run
## 2019 10 10 11:15:50.979 INFO megazone23 \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/
## 2019 10 10 11:15:50.979 INFO megazone23 mbatchTrimData Starting
## 2019 10 10 11:15:50.979 INFO megazone23 MBatch Version: 2019-09-04-1100
## 2019 10 10 11:15:58.661 INFO megazone23 mbatchTrimData theMaxSize= 1e+05
## 2019 10 10 11:15:58.661 INFO megazone23 mbatchTrimData ncol(theMatrix)= 80
## 2019 10 10 11:15:58.661 INFO megazone23 mbatchTrimData nrow(theMatrix)= 1250
## 2019 10 10 11:15:58.661 INFO megazone23 mbatchTrimData Finishing
## 2019 10 10 11:15:58.662 INFO megazone23 ~~~~~
## 2019 10 10 11:15:58.663 DEBUG megazone23 Changing LC_COLLATE to C for duration of run
## 2019 10 10 11:15:58.663 INFO megazone23 \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/
## 2019 10 10 11:15:58.663 INFO megazone23 mbatchFilterData Starting
## 2019 10 10 11:15:58.663 INFO megazone23 MBatch Version: 2019-09-04-1100
## 2019 10 10 11:15:58.664 DEBUG megazone23 rows pre filter 1250
## 2019 10 10 11:15:58.904 DEBUG megazone23 rows post filter 1250
## 2019 10 10 11:15:58.904 DEBUG megazone23 mbatchFilterData Prefilter, gene data had 1250 while post
## 2019 10 10 11:15:58.905 DEBUG megazone23 mbatchFilterData Prefilter, batch data had 80 while post
## 2019 10 10 11:15:58.905 INFO megazone23 mbatchFilterData Finishing
## 2019 10 10 11:15:58.905 INFO megazone23 ~~~~~
## 2019 10 10 11:15:58.908 DEBUG megazone23 Calling .jinit for PCA
## 2019 10 10 11:15:58.908 DEBUG megazone23 classpath=, /tmp/RtmpGYrlWy/temp_libpath519538b7afb3/MBatch
## 2019 10 10 11:15:59.052 DEBUG megazone23 Changing LC_COLLATE to C for duration of run
## 2019 10 10 11:15:59.053 DEBUG megazone23 createBatchEffectsOutput_pca_dualBatch - start loop
## 2019 10 10 11:15:59.055 DEBUG megazone23 doSamplePcaCall - pre-filter matrix size= 1250,80
## 2019 10 10 11:15:59.056 DEBUG megazone23 filterBasedOnGeneLimit theMatrixGeneData nrow= 1250
## 2019 10 10 11:15:59.056 DEBUG megazone23 filterBasedOnGeneLimit theMatrixGeneData ncol= 80
## 2019 10 10 11:15:59.056 DEBUG megazone23 doSamplePcaCall - pre-filter matrix size= 1250,80
## 2019 10 10 11:15:59.057 DEBUG megazone23 doSamplePcaCall - no-NA matrix length= 1247
## 2019 10 10 11:15:59.097 DEBUG megazone23 doSamplePcaCall - max(theCompPairList)= 3
## 2019 10 10 11:15:59.098 DEBUG megazone23 doSamplePcaCall - pca scores size= 80,80
## 2019 10 10 11:15:59.099 DEBUG megazone23 checkCreateDir: /bea_testing/output/PCA_DualBatch_Structure
## 2019 10 10 11:15:59.099 DEBUG megazone23 checkCreateDir: /bea_testing/output/PCA_DualBatch_Structure
## 2019 10 10 11:15:59.100 DEBUG megazone23 pvalueDSC start
## 2019 10 10 11:15:59.100 DEBUG megazone23 pvalueDSC component A < 1
## 2019 10 10 11:15:59.100 DEBUG megazone23 pvalueDSC call pvalueDSCwithExcerpt

```

```

## 2019 10 10 11:15:59.101 DEBUG megazone23 pvalueDSCwithExcerpt start
## 2019 10 10 11:15:59.101 DEBUG megazone23 nrow(thePcaDataExcerpt)= 80
## 2019 10 10 11:15:59.101 DEBUG megazone23 ncol(thePcaDataExcerpt)= 80
## 2019 10 10 11:15:59.102 DEBUG megazone23 length(theBatchIdsForSamples)= 80
## 2019 10 10 11:15:59.112 DEBUG megazone23 getDSCwithExcerpt before java
## 2019 10 10 11:15:59.204 DEBUG megazone23 getDSCwithExcerpt after java
## 2019 10 10 11:15:59.206 DEBUG megazone23 pvalueDSCwithExcerpt after getDSCwithExcerpt
## 2019 10 10 11:15:59.207 DEBUG megazone23 pvalueDSCwithExcerpt length(unique(theBatchIdsForSamples))=
## 2019 10 10 11:15:59.207 DEBUG megazone23 pvalueDSCwithExcerpt 2222 length(permResults)= 1
## 2019 10 10 11:15:59.207 DEBUG megazone23 pvalueDSCwithExcerpt PCA-PVALUE-DSC
## 2019 10 10 11:15:59.216 DEBUG megazone23 openAndWriteDscPermsFile -- start
## 2019 10 10 11:15:59.217 DEBUG megazone23 openAndWriteDscPermsFile -- redo name
## 2019 10 10 11:15:59.217 DEBUG megazone23 openAndWriteDscPermsFile -- check file
## 2019 10 10 11:15:59.218 DEBUG megazone23 openAndWriteDscPermsFile -- no file, write
## 2019 10 10 11:15:59.218 DEBUG megazone23 openAndWriteDscPermsFile -- nothing to write
## 2019 10 10 11:15:59.218 DEBUG megazone23 write writePcaDataFilesForDataset
## 2019 10 10 11:15:59.218 DEBUG megazone23 checkCreateDir: /bea_testing/output/PCA_DualBatch_Structure
## 2019 10 10 11:15:59.219 DEBUG megazone23 writePCAValuesTSV
## 2019 10 10 11:15:59.219 DEBUG megazone23 writePCAValuesTSV outputFile
## 2019 10 10 11:15:59.219 DEBUG megazone23 writePCAValuesTSV components
## 2019 10 10 11:15:59.220 DEBUG megazone23 writePCAValuesTSV myDataFrame
## 2019 10 10 11:15:59.220 DEBUG megazone23 writePCAValuesTSV - values
## 2019 10 10 11:15:59.220 DEBUG megazone23 writePCAValuesTSV - matrix
## 2019 10 10 11:15:59.221 DEBUG megazone23 writePCAValuesTSV - write
## 2019 10 10 11:15:59.223 DEBUG megazone23 writePCAValuesTSV finished /bea_testing/output/PCA_DualBat
## 2019 10 10 11:15:59.223 DEBUG megazone23 writePCAAnnotations
## 2019 10 10 11:15:59.224 DEBUG megazone23 writePCAAnnotations - header and all data
## 2019 10 10 11:15:59.519 DEBUG megazone23 doInternalPca before filter - Number of rownames for pca sco
## 2019 10 10 11:15:59.520 DEBUG megazone23 doInternalPca before filter - length of batch ids for sample
## 2019 10 10 11:15:59.520 DEBUG megazone23 doInternalPca after filter - Number of rownames for pca sco
## 2019 10 10 11:15:59.520 DEBUG megazone23 doInternalPca after filter - length of batch ids for sample
## 2019 10 10 11:15:59.521 DEBUG megazone23 doInternalPca - pvalueDSC
## 2019 10 10 11:15:59.521 DEBUG megazone23 pvalueDSC start
## 2019 10 10 11:15:59.521 DEBUG megazone23 pvalueDSC transpose
## 2019 10 10 11:15:59.521 DEBUG megazone23 pvalueDSC call pvalueDSCwithExcerpt
## 2019 10 10 11:15:59.522 DEBUG megazone23 pvalueDSCwithExcerpt start
## 2019 10 10 11:15:59.522 DEBUG megazone23 nrow(thePcaDataExcerpt)= 2
## 2019 10 10 11:15:59.522 DEBUG megazone23 ncol(thePcaDataExcerpt)= 80
## 2019 10 10 11:15:59.524 DEBUG megazone23 length(theBatchIdsForSamples)= 80
## 2019 10 10 11:15:59.524 DEBUG megazone23 getDSCwithExcerpt before java
## 2019 10 10 11:15:59.531 DEBUG megazone23 getDSCwithExcerpt after java
## 2019 10 10 11:15:59.533 DEBUG megazone23 pvalueDSCwithExcerpt after getDSCwithExcerpt
## 2019 10 10 11:15:59.533 DEBUG megazone23 pvalueDSCwithExcerpt length(unique(theBatchIdsForSamples))=
## 2019 10 10 11:15:59.533 DEBUG megazone23 pvalueDSCwithExcerpt 2222 length(permResults)= 1
## 2019 10 10 11:15:59.533 DEBUG megazone23 pvalueDSCwithExcerpt PCA-PVALUE-DSC
## 2019 10 10 11:15:59.534 DEBUG megazone23 openAndWriteDscPermsFile -- start
## 2019 10 10 11:15:59.534 DEBUG megazone23 openAndWriteDscPermsFile -- redo name
## 2019 10 10 11:15:59.535 DEBUG megazone23 openAndWriteDscPermsFile -- check file
## 2019 10 10 11:15:59.535 DEBUG megazone23 openAndWriteDscPermsFile -- no file, write
## 2019 10 10 11:15:59.535 DEBUG megazone23 openAndWriteDscPermsFile -- nothing to write
## 2019 10 10 11:15:59.536 DEBUG megazone23 writePCAAnnotations
## 2019 10 10 11:15:59.536 DEBUG megazone23 writePCAAnnotations - data PC1,PC2
## 2019 10 10 11:15:59.561 DEBUG megazone23 writePcaAnalysis_Image - Number of rownames for pca score (
## 2019 10 10 11:15:59.562 DEBUG megazone23 writePcaAnalysis_Image - length of batch ids for samples 80

```

```

## 2019 10 10 11:15:59.669 DEBUG megazone23 mbatchStandardLegend - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:15:59.674 DEBUG megazone23 mbatchStandardLegend - .jinit complete
## 2019 10 10 11:15:59.674 DEBUG megazone23 mbatchStandardLegend - theTitle Rays: PlateId (80)
## 2019 10 10 11:15:59.674 DEBUG megazone23 mbatchStandardLegend - theVersion MBatch 1.5.2
## 2019 10 10 11:15:59.675 DEBUG megazone23 mbatchStandardLegend - theFilenamePath /bea_testing/output
## 2019 10 10 11:15:59.676 DEBUG megazone23 mbatchStandardLegend - theLegendNames A29J (80)
## 2019 10 10 11:15:59.676 DEBUG megazone23 mbatchStandardLegend - theLegendNames 1
## 2019 10 10 11:15:59.676 DEBUG megazone23 mbatchStandardLegend - theLegendColors 1
## 2019 10 10 11:15:59.676 DEBUG megazone23 mbatchStandardLegend - theLegendSymbols 0
## 2019 10 10 11:15:59.676 DEBUG megazone23 mbatchStandardLegend - myColors #b30000
## 2019 10 10 11:15:59.677 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:00.211 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:00.216 DEBUG megazone23 mbatchStandardLegend - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:16:00.229 DEBUG megazone23 mbatchStandardLegend - .jinit complete
## 2019 10 10 11:16:00.230 DEBUG megazone23 mbatchStandardLegend - theTitle Points: TSS (80)
## 2019 10 10 11:16:00.230 DEBUG megazone23 mbatchStandardLegend - theVersion MBatch 1.5.2
## 2019 10 10 11:16:00.231 DEBUG megazone23 mbatchStandardLegend - theFilenamePath /bea_testing/output
## 2019 10 10 11:16:00.231 DEBUG megazone23 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)
## 2019 10 10 11:16:00.231 DEBUG megazone23 mbatchStandardLegend - theLegendNames 5
## 2019 10 10 11:16:00.231 DEBUG megazone23 mbatchStandardLegend - theLegendColors 5
## 2019 10 10 11:16:00.232 DEBUG megazone23 mbatchStandardLegend - theLegendSymbols 5
## 2019 10 10 11:16:00.232 DEBUG megazone23 mbatchStandardLegend - myColors #b30000,#8fb300,#00b347,#0
## 2019 10 10 11:16:00.232 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:00.404 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:00.489 DEBUG megazone23 mbatchStandardLegend - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:16:00.499 DEBUG megazone23 mbatchStandardLegend - .jinit complete
## 2019 10 10 11:16:00.500 DEBUG megazone23 mbatchStandardLegend - theTitle Dispersion Metrics
## 2019 10 10 11:16:00.500 DEBUG megazone23 mbatchStandardLegend - theVersion MBatch 1.5.2
## 2019 10 10 11:16:00.512 DEBUG megazone23 mbatchStandardLegend - theFilenamePath /bea_testing/output
## 2019 10 10 11:16:00.513 DEBUG megazone23 mbatchStandardLegend - theLegendNames First PCA Component:
## 2019 10 10 11:16:00.513 DEBUG megazone23 mbatchStandardLegend - theLegendNames 14
## 2019 10 10 11:16:00.513 DEBUG megazone23 mbatchStandardLegend - theLegendColors 0
## 2019 10 10 11:16:00.514 DEBUG megazone23 mbatchStandardLegend - theLegendSymbols 0
## 2019 10 10 11:16:00.514 DEBUG megazone23 mbatchStandardLegend - myColors
## 2019 10 10 11:16:00.515 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:00.726 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:00.728 DEBUG megazone23 mbatchStandardCombineLegends - Calling .jinit /tmp/RtmpGYr
## 2019 10 10 11:16:00.734 DEBUG megazone23 mbatchStandardCombineLegends - .jinit complete
## 2019 10 10 11:16:00.734 DEBUG megazone23 mbatchStandardCombineLegends - theTitle Example Title for '
## 2019 10 10 11:16:00.734 DEBUG megazone23 mbatchStandardCombineLegends - theFilenamePath /bea_testing
## 2019 10 10 11:16:00.734 DEBUG megazone23 mbatchStandardCombineLegends - theListOfFiles /bea_testing
## 2019 10 10 11:16:00.738 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:01.265 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:01.268 DEBUG megazone23 doInternalPca before filter - Number of rownames for pca sco
## 2019 10 10 11:16:01.269 DEBUG megazone23 doInternalPca before filter - length of batch ids for sample
## 2019 10 10 11:16:01.269 DEBUG megazone23 doInternalPca after filter - Number of rownames for pca sco
## 2019 10 10 11:16:01.269 DEBUG megazone23 doInternalPca after filter - length of batch ids for sample
## 2019 10 10 11:16:01.269 DEBUG megazone23 doInternalPca - pvalueDSC
## 2019 10 10 11:16:01.269 DEBUG megazone23 pvalueDSC start
## 2019 10 10 11:16:01.270 DEBUG megazone23 pvalueDSC transpose

```

```

## 2019 10 10 11:16:01.270 DEBUG megazone23 pvalueDSC call pvalueDSCwithExcerpt
## 2019 10 10 11:16:01.270 DEBUG megazone23 pvalueDSCwithExcerpt start
## 2019 10 10 11:16:01.271 DEBUG megazone23 nrow(thePcaDataExcerpt)= 2
## 2019 10 10 11:16:01.271 DEBUG megazone23 ncol(thePcaDataExcerpt)= 80
## 2019 10 10 11:16:01.271 DEBUG megazone23 length(theBatchIdsForSamples)= 80
## 2019 10 10 11:16:01.276 DEBUG megazone23 getDSCwithExcerpt before java
## 2019 10 10 11:16:01.278 DEBUG megazone23 getDSCwithExcerpt after java
## 2019 10 10 11:16:01.279 DEBUG megazone23 pvalueDSCwithExcerpt after getDSCwithExcerpt
## 2019 10 10 11:16:01.279 DEBUG megazone23 pvalueDSCwithExcerpt length(unique(theBatchIdsForSamples))=
## 2019 10 10 11:16:01.279 DEBUG megazone23 pvalueDSCwithExcerpt 2222 length(permResults)= 1
## 2019 10 10 11:16:01.280 DEBUG megazone23 pvalueDSCwithExcerpt PCA-PVALUE-DSC
## 2019 10 10 11:16:01.280 DEBUG megazone23 openAndWriteDscPermsFile -- start
## 2019 10 10 11:16:01.280 DEBUG megazone23 openAndWriteDscPermsFile -- redo name
## 2019 10 10 11:16:01.281 DEBUG megazone23 openAndWriteDscPermsFile -- check file
## 2019 10 10 11:16:01.281 DEBUG megazone23 openAndWriteDscPermsFile -- no file, write
## 2019 10 10 11:16:01.281 DEBUG megazone23 openAndWriteDscPermsFile -- nothing to write
## 2019 10 10 11:16:01.282 DEBUG megazone23 writePCAAnnotations
## 2019 10 10 11:16:01.283 DEBUG megazone23 writePCAAnnotations - data PC1,PC3
## 2019 10 10 11:16:01.295 DEBUG megazone23 writePcaAnalysis_Image - Number of rownames for pca score (
## 2019 10 10 11:16:01.295 DEBUG megazone23 writePcaAnalysis_Image - length of batch ids for samples 80

## 2019 10 10 11:16:01.437 DEBUG megazone23 mbatchStandardLegend - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:16:01.444 DEBUG megazone23 mbatchStandardLegend - .jinit complete
## 2019 10 10 11:16:01.445 DEBUG megazone23 mbatchStandardLegend - theTitle Rays: PlateId (80)
## 2019 10 10 11:16:01.445 DEBUG megazone23 mbatchStandardLegend - theVersion MBatch 1.5.2
## 2019 10 10 11:16:01.446 DEBUG megazone23 mbatchStandardLegend - theFilenamePath /bea_testing/output,
## 2019 10 10 11:16:01.446 DEBUG megazone23 mbatchStandardLegend - theLegendNames A29J (80)
## 2019 10 10 11:16:01.446 DEBUG megazone23 mbatchStandardLegend - theLegendNames 1
## 2019 10 10 11:16:01.446 DEBUG megazone23 mbatchStandardLegend - theLegendColors 1
## 2019 10 10 11:16:01.446 DEBUG megazone23 mbatchStandardLegend - theLegendSymbols 0
## 2019 10 10 11:16:01.447 DEBUG megazone23 mbatchStandardLegend - myColors #b30000
## 2019 10 10 11:16:01.447 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:01.464 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:01.466 DEBUG megazone23 mbatchStandardLegend - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:16:01.471 DEBUG megazone23 mbatchStandardLegend - .jinit complete
## 2019 10 10 11:16:01.472 DEBUG megazone23 mbatchStandardLegend - theTitle Points: TSS (80)
## 2019 10 10 11:16:01.472 DEBUG megazone23 mbatchStandardLegend - theVersion MBatch 1.5.2
## 2019 10 10 11:16:01.473 DEBUG megazone23 mbatchStandardLegend - theFilenamePath /bea_testing/output,
## 2019 10 10 11:16:01.473 DEBUG megazone23 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)
## 2019 10 10 11:16:01.473 DEBUG megazone23 mbatchStandardLegend - theLegendNames 5
## 2019 10 10 11:16:01.473 DEBUG megazone23 mbatchStandardLegend - theLegendColors 5
## 2019 10 10 11:16:01.474 DEBUG megazone23 mbatchStandardLegend - theLegendSymbols 5
## 2019 10 10 11:16:01.474 DEBUG megazone23 mbatchStandardLegend - myColors #b30000,#8fb300,#00b347,#0
## 2019 10 10 11:16:01.474 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:01.545 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:01.548 DEBUG megazone23 mbatchStandardLegend - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:16:01.553 DEBUG megazone23 mbatchStandardLegend - .jinit complete
## 2019 10 10 11:16:01.559 DEBUG megazone23 mbatchStandardLegend - theTitle Dispersion Metrics
## 2019 10 10 11:16:01.559 DEBUG megazone23 mbatchStandardLegend - theVersion MBatch 1.5.2
## 2019 10 10 11:16:01.560 DEBUG megazone23 mbatchStandardLegend - theFilenamePath /bea_testing/output,

```



```

## 2019 10 10 11:16:01.560 DEBUG megazone23 mbatchStandardLegend - theLegendNames First PCA Component:
## 2019 10 10 11:16:01.561 DEBUG megazone23 mbatchStandardLegend - theLegendNames 14
## 2019 10 10 11:16:01.561 DEBUG megazone23 mbatchStandardLegend - theLegendColors 0
## 2019 10 10 11:16:01.561 DEBUG megazone23 mbatchStandardLegend - theLegendSymbols 0
## 2019 10 10 11:16:01.561 DEBUG megazone23 mbatchStandardLegend - myColors
## 2019 10 10 11:16:01.562 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:01.717 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:01.719 DEBUG megazone23 mbatchStandardCombineLegends - Calling .jinit /tmp/RtmpGYr
## 2019 10 10 11:16:01.726 DEBUG megazone23 mbatchStandardCombineLegends - .jinit complete
## 2019 10 10 11:16:01.726 DEBUG megazone23 mbatchStandardCombineLegends - theTitle Example Title for '
## 2019 10 10 11:16:01.726 DEBUG megazone23 mbatchStandardCombineLegends - theFilenamePath /bea_testing
## 2019 10 10 11:16:01.727 DEBUG megazone23 mbatchStandardCombineLegends - theListOfFiles /bea_testing
## 2019 10 10 11:16:01.727 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:02.054 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:02.055 DEBUG megazone23 doSamplePcaCall - pre-filter matrix size= 1250,80
## 2019 10 10 11:16:02.055 DEBUG megazone23 filterBasedOnGeneLimit theMatrixGeneData nrow= 1250
## 2019 10 10 11:16:02.055 DEBUG megazone23 filterBasedOnGeneLimit theMatrixGeneData ncol= 80
## 2019 10 10 11:16:02.056 DEBUG megazone23 doSamplePcaCall - pre-filter matrix size= 1250,80
## 2019 10 10 11:16:02.056 DEBUG megazone23 doSamplePcaCall - no-NA matrix length= 1247
## 2019 10 10 11:16:02.102 DEBUG megazone23 doSamplePcaCall - max(theCompPairList)= 3
## 2019 10 10 11:16:02.102 DEBUG megazone23 doSamplePcaCall - pca scores size= 80,80
## 2019 10 10 11:16:02.103 DEBUG megazone23 checkCreateDir: /bea_testing/output/PCA_DualBatch_Structur
## 2019 10 10 11:16:02.103 DEBUG megazone23 checkCreateDir: /bea_testing/output/PCA_DualBatch_Structur
## 2019 10 10 11:16:02.104 DEBUG megazone23 pvalueDSC start
## 2019 10 10 11:16:02.104 DEBUG megazone23 pvalueDSC component A < 1
## 2019 10 10 11:16:02.104 DEBUG megazone23 pvalueDSC call pvalueDSCwithExcerpt
## 2019 10 10 11:16:02.104 DEBUG megazone23 pvalueDSCwithExcerpt start
## 2019 10 10 11:16:02.105 DEBUG megazone23 nrow(thePcaDataExcerpt)= 80
## 2019 10 10 11:16:02.105 DEBUG megazone23 ncol(thePcaDataExcerpt)= 80
## 2019 10 10 11:16:02.105 DEBUG megazone23 length(theBatchIdsForSamples)= 80
## 2019 10 10 11:16:02.106 DEBUG megazone23 getDSCwithExcerpt before java
## 2019 10 10 11:16:02.118 DEBUG megazone23 getDSCwithExcerpt after java
## 2019 10 10 11:16:02.119 DEBUG megazone23 pvalueDSCwithExcerpt after getDSCwithExcerpt
## 2019 10 10 11:16:02.120 DEBUG megazone23 pvalueDSCwithExcerpt length(unique(theBatchIdsForSamples))=
## 2019 10 10 11:16:02.120 DEBUG megazone23 pvalueDSCwithExcerpt 2222 length(permResults)= 1
## 2019 10 10 11:16:02.121 DEBUG megazone23 pvalueDSCwithExcerpt PCA-PVALUE-DSC
## 2019 10 10 11:16:02.129 DEBUG megazone23 openAndWriteDscPermsFile -- start
## 2019 10 10 11:16:02.130 DEBUG megazone23 openAndWriteDscPermsFile -- redo name
## 2019 10 10 11:16:02.130 DEBUG megazone23 openAndWriteDscPermsFile -- check file
## 2019 10 10 11:16:02.130 DEBUG megazone23 openAndWriteDscPermsFile -- no file, write
## 2019 10 10 11:16:02.131 DEBUG megazone23 openAndWriteDscPermsFile -- nothing to write
## 2019 10 10 11:16:02.131 DEBUG megazone23 write writePcaDataFilesForDataset
## 2019 10 10 11:16:02.131 DEBUG megazone23 checkCreateDir: /bea_testing/output/PCA_DualBatch_Structur
## 2019 10 10 11:16:02.131 DEBUG megazone23 writePCAValuesTSV
## 2019 10 10 11:16:02.132 DEBUG megazone23 writePCAValuesTSV outputFile
## 2019 10 10 11:16:02.132 DEBUG megazone23 writePCAValuesTSV components
## 2019 10 10 11:16:02.132 DEBUG megazone23 writePCAValuesTSV myDataFrame
## 2019 10 10 11:16:02.132 DEBUG megazone23 writePCAValuesTSV - values
## 2019 10 10 11:16:02.133 DEBUG megazone23 writePCAValuesTSV - matrix
## 2019 10 10 11:16:02.133 DEBUG megazone23 writePCAValuesTSV - write
## 2019 10 10 11:16:02.135 DEBUG megazone23 writePCAValuesTSV finished /bea_testing/output/PCA_DualBat
## 2019 10 10 11:16:02.135 DEBUG megazone23 writePCAAnnotations
## 2019 10 10 11:16:02.135 DEBUG megazone23 writePCAAnnotations - header and all data
## 2019 10 10 11:16:02.380 DEBUG megazone23 doInternalPca before filter - Number of rownames for pca sc

```

```

## 2019 10 10 11:16:02.381 DEBUG megazone23 doInternalPca before filter - length of batch ids for samples
## 2019 10 10 11:16:02.381 DEBUG megazone23 doInternalPca after filter - Number of rownames for pca score
## 2019 10 10 11:16:02.381 DEBUG megazone23 doInternalPca after filter - length of batch ids for samples
## 2019 10 10 11:16:02.381 DEBUG megazone23 doInternalPca - pvalueDSC
## 2019 10 10 11:16:02.382 DEBUG megazone23 pvalueDSC start
## 2019 10 10 11:16:02.382 DEBUG megazone23 pvalueDSC transpose
## 2019 10 10 11:16:02.382 DEBUG megazone23 pvalueDSC call pvalueDSCwithExcerpt
## 2019 10 10 11:16:02.382 DEBUG megazone23 pvalueDSCwithExcerpt start
## 2019 10 10 11:16:02.382 DEBUG megazone23 nrow(thePcaDataExcerpt)= 2
## 2019 10 10 11:16:02.383 DEBUG megazone23 ncol(thePcaDataExcerpt)= 80
## 2019 10 10 11:16:02.383 DEBUG megazone23 length(theBatchIdsForSamples)= 80
## 2019 10 10 11:16:02.383 DEBUG megazone23 getDSCwithExcerpt before java
## 2019 10 10 11:16:02.389 DEBUG megazone23 getDSCwithExcerpt after java
## 2019 10 10 11:16:02.389 DEBUG megazone23 pvalueDSCwithExcerpt after getDSCwithExcerpt
## 2019 10 10 11:16:02.390 DEBUG megazone23 pvalueDSCwithExcerpt length(unique(theBatchIdsForSamples))=
## 2019 10 10 11:16:02.390 DEBUG megazone23 pvalueDSCwithExcerpt 2222 length(permResults)= 1
## 2019 10 10 11:16:02.390 DEBUG megazone23 pvalueDSCwithExcerpt PCA-PVALUE-DSC
## 2019 10 10 11:16:02.392 DEBUG megazone23 openAndWriteDscPermsFile -- start
## 2019 10 10 11:16:02.392 DEBUG megazone23 openAndWriteDscPermsFile -- redo name
## 2019 10 10 11:16:02.392 DEBUG megazone23 openAndWriteDscPermsFile -- check file
## 2019 10 10 11:16:02.393 DEBUG megazone23 openAndWriteDscPermsFile -- no file, write
## 2019 10 10 11:16:02.393 DEBUG megazone23 openAndWriteDscPermsFile -- nothing to write
## 2019 10 10 11:16:02.393 DEBUG megazone23 writePCAAnnotations
## 2019 10 10 11:16:02.394 DEBUG megazone23 writePCAAnnotations - data PC1,PC2
## 2019 10 10 11:16:02.396 DEBUG megazone23 writePcaAnalysis_Image - Number of rownames for pca score (
## 2019 10 10 11:16:02.396 DEBUG megazone23 writePcaAnalysis_Image - length of batch ids for samples 80

## 2019 10 10 11:16:02.495 DEBUG megazone23 mbatchStandardLegend - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:16:02.501 DEBUG megazone23 mbatchStandardLegend - .jinit complete
## 2019 10 10 11:16:02.501 DEBUG megazone23 mbatchStandardLegend - theTitle Rays: BatchId (80)
## 2019 10 10 11:16:02.502 DEBUG megazone23 mbatchStandardLegend - theVersion MBatch 1.5.2
## 2019 10 10 11:16:02.502 DEBUG megazone23 mbatchStandardLegend - theFilePath /bea_testing/output
## 2019 10 10 11:16:02.502 DEBUG megazone23 mbatchStandardLegend - theLegendNames 00304 (80)
## 2019 10 10 11:16:02.503 DEBUG megazone23 mbatchStandardLegend - theLegendNames 1
## 2019 10 10 11:16:02.503 DEBUG megazone23 mbatchStandardLegend - theLegendColors 1
## 2019 10 10 11:16:02.503 DEBUG megazone23 mbatchStandardLegend - theLegendSymbols 0
## 2019 10 10 11:16:02.503 DEBUG megazone23 mbatchStandardLegend - myColors #b30000
## 2019 10 10 11:16:02.504 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:02.537 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:02.541 DEBUG megazone23 mbatchStandardLegend - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:16:02.547 DEBUG megazone23 mbatchStandardLegend - .jinit complete
## 2019 10 10 11:16:02.547 DEBUG megazone23 mbatchStandardLegend - theTitle Points: TSS (80)
## 2019 10 10 11:16:02.547 DEBUG megazone23 mbatchStandardLegend - theVersion MBatch 1.5.2
## 2019 10 10 11:16:02.549 DEBUG megazone23 mbatchStandardLegend - theFilePath /bea_testing/output
## 2019 10 10 11:16:02.549 DEBUG megazone23 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)
## 2019 10 10 11:16:02.549 DEBUG megazone23 mbatchStandardLegend - theLegendNames 5
## 2019 10 10 11:16:02.549 DEBUG megazone23 mbatchStandardLegend - theLegendColors 5
## 2019 10 10 11:16:02.549 DEBUG megazone23 mbatchStandardLegend - theLegendSymbols 5
## 2019 10 10 11:16:02.550 DEBUG megazone23 mbatchStandardLegend - myColors #b30000,#8fb300,#00b347,#00
## 2019 10 10 11:16:02.550 DEBUG megazone23 mbatchStandardLegend before java

```

```

## 2019 10 10 11:16:02.634 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:02.636 DEBUG megazone23 mbatchStandardLegend - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:16:02.643 DEBUG megazone23 mbatchStandardLegend - .jinit complete
## 2019 10 10 11:16:02.644 DEBUG megazone23 mbatchStandardLegend - theTitle Dispersion Metrics
## 2019 10 10 11:16:02.644 DEBUG megazone23 mbatchStandardLegend - theVersion MBatch 1.5.2
## 2019 10 10 11:16:02.645 DEBUG megazone23 mbatchStandardLegend - theFilenamePath /bea_testing/output
## 2019 10 10 11:16:02.645 DEBUG megazone23 mbatchStandardLegend - theLegendNames First PCA Component:
## 2019 10 10 11:16:02.646 DEBUG megazone23 mbatchStandardLegend - theLegendNames 14
## 2019 10 10 11:16:02.646 DEBUG megazone23 mbatchStandardLegend - theLegendColors 0
## 2019 10 10 11:16:02.646 DEBUG megazone23 mbatchStandardLegend - theLegendSymbols 0
## 2019 10 10 11:16:02.646 DEBUG megazone23 mbatchStandardLegend - myColors
## 2019 10 10 11:16:02.646 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:02.803 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:02.805 DEBUG megazone23 mbatchStandardCombineLegends - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:16:02.811 DEBUG megazone23 mbatchStandardCombineLegends - .jinit complete
## 2019 10 10 11:16:02.811 DEBUG megazone23 mbatchStandardCombineLegends - theTitle Example Title for '
## 2019 10 10 11:16:02.811 DEBUG megazone23 mbatchStandardCombineLegends - theFilenamePath /bea_testing/output
## 2019 10 10 11:16:02.811 DEBUG megazone23 mbatchStandardCombineLegends - theListOfFiles /bea_testing/output
## 2019 10 10 11:16:02.812 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:03.107 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:03.108 DEBUG megazone23 doInternalPca before filter - Number of rownames for pca score
## 2019 10 10 11:16:03.108 DEBUG megazone23 doInternalPca before filter - length of batch ids for samples
## 2019 10 10 11:16:03.108 DEBUG megazone23 doInternalPca after filter - Number of rownames for pca score
## 2019 10 10 11:16:03.108 DEBUG megazone23 doInternalPca after filter - length of batch ids for samples
## 2019 10 10 11:16:03.109 DEBUG megazone23 doInternalPca - pvalueDSC
## 2019 10 10 11:16:03.109 DEBUG megazone23 pvalueDSC start
## 2019 10 10 11:16:03.109 DEBUG megazone23 pvalueDSC transpose
## 2019 10 10 11:16:03.109 DEBUG megazone23 pvalueDSC call pvalueDSCwithExcerpt
## 2019 10 10 11:16:03.110 DEBUG megazone23 pvalueDSCwithExcerpt start
## 2019 10 10 11:16:03.110 DEBUG megazone23 nrow(thePcaDataExcerpt)= 2
## 2019 10 10 11:16:03.110 DEBUG megazone23 ncol(thePcaDataExcerpt)= 80
## 2019 10 10 11:16:03.110 DEBUG megazone23 length(theBatchIdsForSamples)= 80
## 2019 10 10 11:16:03.111 DEBUG megazone23 getDSCwithExcerpt before java
## 2019 10 10 11:16:03.116 DEBUG megazone23 getDSCwithExcerpt after java
## 2019 10 10 11:16:03.117 DEBUG megazone23 pvalueDSCwithExcerpt after getDSCwithExcerpt
## 2019 10 10 11:16:03.118 DEBUG megazone23 pvalueDSCwithExcerpt length(unique(theBatchIdsForSamples))=
## 2019 10 10 11:16:03.118 DEBUG megazone23 pvalueDSCwithExcerpt 2222 length(permResults)= 1
## 2019 10 10 11:16:03.118 DEBUG megazone23 pvalueDSCwithExcerpt PCA-PVALUE-DSC
## 2019 10 10 11:16:03.119 DEBUG megazone23 openAndWriteDscPermsFile -- start
## 2019 10 10 11:16:03.119 DEBUG megazone23 openAndWriteDscPermsFile -- redo name
## 2019 10 10 11:16:03.121 DEBUG megazone23 openAndWriteDscPermsFile -- check file
## 2019 10 10 11:16:03.121 DEBUG megazone23 openAndWriteDscPermsFile -- no file, write
## 2019 10 10 11:16:03.121 DEBUG megazone23 openAndWriteDscPermsFile -- nothing to write
## 2019 10 10 11:16:03.122 DEBUG megazone23 writePCAAnnotations
## 2019 10 10 11:16:03.122 DEBUG megazone23 writePCAAnnotations - data PC1,PC3
## 2019 10 10 11:16:03.124 DEBUG megazone23 writePcaAnalysis_Image - Number of rownames for pca score (
## 2019 10 10 11:16:03.125 DEBUG megazone23 writePcaAnalysis_Image - length of batch ids for samples 80

## 2019 10 10 11:16:03.216 DEBUG megazone23 mbatchStandardLegend - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:16:03.223 DEBUG megazone23 mbatchStandardLegend - .jinit complete
## 2019 10 10 11:16:03.224 DEBUG megazone23 mbatchStandardLegend - theTitle Rays: BatchId (80)
## 2019 10 10 11:16:03.225 DEBUG megazone23 mbatchStandardLegend - theVersion MBatch 1.5.2
## 2019 10 10 11:16:03.227 DEBUG megazone23 mbatchStandardLegend - theFilenamePath /bea_testing/output
## 2019 10 10 11:16:03.227 DEBUG megazone23 mbatchStandardLegend - theLegendNames 00304 (80)

```

```

## 2019 10 10 11:16:03.228 DEBUG megazone23 mbatchStandardLegend - theLegendNames 1
## 2019 10 10 11:16:03.228 DEBUG megazone23 mbatchStandardLegend - theLegendColors 1
## 2019 10 10 11:16:03.228 DEBUG megazone23 mbatchStandardLegend - theLegendSymbols 0
## 2019 10 10 11:16:03.228 DEBUG megazone23 mbatchStandardLegend - myColors #b30000
## 2019 10 10 11:16:03.229 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:03.269 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:03.271 DEBUG megazone23 mbatchStandardLegend - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:16:03.284 DEBUG megazone23 mbatchStandardLegend - .jinit complete
## 2019 10 10 11:16:03.285 DEBUG megazone23 mbatchStandardLegend - theTitle Points: TSS (80)
## 2019 10 10 11:16:03.285 DEBUG megazone23 mbatchStandardLegend - theVersion MBatch 1.5.2
## 2019 10 10 11:16:03.289 DEBUG megazone23 mbatchStandardLegend - theFilenamePath /bea_testing/output
## 2019 10 10 11:16:03.290 DEBUG megazone23 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)
## 2019 10 10 11:16:03.290 DEBUG megazone23 mbatchStandardLegend - theLegendNames 5
## 2019 10 10 11:16:03.290 DEBUG megazone23 mbatchStandardLegend - theLegendColors 5
## 2019 10 10 11:16:03.290 DEBUG megazone23 mbatchStandardLegend - theLegendSymbols 5
## 2019 10 10 11:16:03.329 DEBUG megazone23 mbatchStandardLegend - myColors #b30000,#8fb300,#00b347,#0
## 2019 10 10 11:16:03.330 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:03.425 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:03.427 DEBUG megazone23 mbatchStandardLegend - Calling .jinit /tmp/RtmpGYrlWy/temp
## 2019 10 10 11:16:03.434 DEBUG megazone23 mbatchStandardLegend - .jinit complete
## 2019 10 10 11:16:03.434 DEBUG megazone23 mbatchStandardLegend - theTitle Dispersion Metrics
## 2019 10 10 11:16:03.435 DEBUG megazone23 mbatchStandardLegend - theVersion MBatch 1.5.2
## 2019 10 10 11:16:03.435 DEBUG megazone23 mbatchStandardLegend - theFilenamePath /bea_testing/output
## 2019 10 10 11:16:03.436 DEBUG megazone23 mbatchStandardLegend - theLegendNames First PCA Component:
## 2019 10 10 11:16:03.436 DEBUG megazone23 mbatchStandardLegend - theLegendNames 14
## 2019 10 10 11:16:03.436 DEBUG megazone23 mbatchStandardLegend - theLegendColors 0
## 2019 10 10 11:16:03.437 DEBUG megazone23 mbatchStandardLegend - theLegendSymbols 0
## 2019 10 10 11:16:03.437 DEBUG megazone23 mbatchStandardLegend - myColors
## 2019 10 10 11:16:03.437 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:03.595 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:03.596 DEBUG megazone23 mbatchStandardCombineLegends - Calling .jinit /tmp/RtmpGYr
## 2019 10 10 11:16:03.601 DEBUG megazone23 mbatchStandardCombineLegends - .jinit complete
## 2019 10 10 11:16:03.601 DEBUG megazone23 mbatchStandardCombineLegends - theTitle Example Title for
## 2019 10 10 11:16:03.601 DEBUG megazone23 mbatchStandardCombineLegends - theFilenamePath /bea_testing
## 2019 10 10 11:16:03.601 DEBUG megazone23 mbatchStandardCombineLegends - theListOfFiles /bea_testing
## 2019 10 10 11:16:03.601 DEBUG megazone23 mbatchStandardLegend before java
## 2019 10 10 11:16:04.062 DEBUG megazone23 mbatchStandardLegend after java
## 2019 10 10 11:16:04.064 DEBUG megazone23 myCompListDscFile
## [1] "ALL__CompListDSC.RData"
## [2] "BatchIdwithTSS/DualBatch/ANY_Comp1_Comp2_DSC.txt"
## [3] "BatchIdwithTSS/DualBatch/ANY_Comp1_Comp2_DSC.txt__CompDSC.RData"
## [4] "BatchIdwithTSS/DualBatch/ANY_Comp1_Comp3_DSC.txt"
## [5] "BatchIdwithTSS/DualBatch/ANY_Comp1_Comp3_DSC.txt__CompDSC.RData"
## [6] "BatchIdwithTSS/DualBatch/ANY_DSC.txt"
## [7] "BatchIdwithTSS/DualBatch/ANY_DSC.txt__OverallDSC.RData"
## [8] "BatchIdwithTSS/DualBatch/ANY_FVE.txt"
## [9] "BatchIdwithTSS/DualBatch/ANY_Scores.txt"
## [10] "BatchIdwithTSS/DualBatch/ANY_Weights.txt"
## [11] "BatchIdwithTSS/DualBatch/PCA-Plus/BatchIdwithTSS_Comp1_Comp2_Diagram.png"
## [12] "BatchIdwithTSS/DualBatch/PCA-Plus/BatchIdwithTSS_Comp1_Comp2_Legend-ALL.png"

```

```
## [13] "BatchIdwithTSS/DualBatch/PCA-Plus/BatchIdwithTSS_Comp1_Comp2_Legend-DSC.png"
## [14] "BatchIdwithTSS/DualBatch/PCA-Plus/BatchIdwithTSS_Comp1_Comp2_Legend-Points.png"
## [15] "BatchIdwithTSS/DualBatch/PCA-Plus/BatchIdwithTSS_Comp1_Comp2_Legend-Rays.png"
## [16] "BatchIdwithTSS/DualBatch/PCA-Plus/BatchIdwithTSS_Comp1_Comp3_Diagram.png"
## [17] "BatchIdwithTSS/DualBatch/PCA-Plus/BatchIdwithTSS_Comp1_Comp3_Legend-ALL.png"
## [18] "BatchIdwithTSS/DualBatch/PCA-Plus/BatchIdwithTSS_Comp1_Comp3_Legend-DSC.png"
## [19] "BatchIdwithTSS/DualBatch/PCA-Plus/BatchIdwithTSS_Comp1_Comp3_Legend-Points.png"
## [20] "BatchIdwithTSS/DualBatch/PCA-Plus/BatchIdwithTSS_Comp1_Comp3_Legend-Rays.png"
## [21] "BatchIdwithTSS/DualBatch/PCAAnnotations.tsv"
## [22] "BatchIdwithTSS/DualBatch/PCAVvalues.tsv"
## [23] "DSCOverview.tsv"
## [24] "PlateIdwithTSS/DualBatch/ANY_Comp1_Comp2_DSC.txt"
## [25] "PlateIdwithTSS/DualBatch/ANY_Comp1_Comp2_DSC.txt__CompDSC.RData"
## [26] "PlateIdwithTSS/DualBatch/ANY_Comp1_Comp3_DSC.txt"
## [27] "PlateIdwithTSS/DualBatch/ANY_Comp1_Comp3_DSC.txt__CompDSC.RData"
## [28] "PlateIdwithTSS/DualBatch/ANY_DSC.txt"
## [29] "PlateIdwithTSS/DualBatch/ANY_DSC.txt__OverallDSC.RData"
## [30] "PlateIdwithTSS/DualBatch/ANY_FVE.txt"
## [31] "PlateIdwithTSS/DualBatch/ANY_Scores.txt"
## [32] "PlateIdwithTSS/DualBatch/ANY_Weights.txt"
## [33] "PlateIdwithTSS/DualBatch/PCA-Plus/PlateIdwithTSS_Comp1_Comp2_Diagram.png"
## [34] "PlateIdwithTSS/DualBatch/PCA-Plus/PlateIdwithTSS_Comp1_Comp2_Legend-ALL.png"
## [35] "PlateIdwithTSS/DualBatch/PCA-Plus/PlateIdwithTSS_Comp1_Comp2_Legend-DSC.png"
## [36] "PlateIdwithTSS/DualBatch/PCA-Plus/PlateIdwithTSS_Comp1_Comp2_Legend-Points.png"
## [37] "PlateIdwithTSS/DualBatch/PCA-Plus/PlateIdwithTSS_Comp1_Comp2_Legend-Rays.png"
## [38] "PlateIdwithTSS/DualBatch/PCA-Plus/PlateIdwithTSS_Comp1_Comp3_Diagram.png"
## [39] "PlateIdwithTSS/DualBatch/PCA-Plus/PlateIdwithTSS_Comp1_Comp3_Legend-ALL.png"
## [40] "PlateIdwithTSS/DualBatch/PCA-Plus/PlateIdwithTSS_Comp1_Comp3_Legend-DSC.png"
## [41] "PlateIdwithTSS/DualBatch/PCA-Plus/PlateIdwithTSS_Comp1_Comp3_Legend-Points.png"
## [42] "PlateIdwithTSS/DualBatch/PCA-Plus/PlateIdwithTSS_Comp1_Comp3_Legend-Rays.png"
## [43] "PlateIdwithTSS/DualBatch/PCAAnnotations.tsv"
## [44] "PlateIdwithTSS/DualBatch/PCAVvalues.tsv"
```

7 Example File Output

The above code creates the following subdirectories and files. The RData and DSCOverview.tsv files are used internally. The subdirectories correspond to the Batch Types Pairs on which the PCA-Plus assessment was performed.

```
/bea_testing/output/PCA_DualBatch_Structures$ ls -l
total 16
-rw-r--r-- 1 linux linux 77 Jun 14 09:34 ALL__CompListDSC.RData
drwxr-xr-x 3 linux linux 4096 Jun 14 09:34 BatchIdwithTSS
-rw-r--r-- 1 linux linux 348 Jun 14 09:34 DSCOverview.tsv
drwxr-xr-x 3 linux linux 4096 Jun 14 09:34 PlateIdwithTSS
```

Looking at the “BatchIdwithTSS” subdirectory, it contains another directory called DualBatch (a different function provides “ManyToMany” comparisons). The DualBatch directory contains the following files and a PCA-Plus sub-directory.

DSC txt files are described above in theDoDscPermsFileFlag.

ANY_Scores.txt is described above in theDoSampleLocatorFlag.

The below files are tab delimited.

ANY_FVE.txt has three columns: “Component Number”, “FVE (%)”, “Cumulative FVE (%)”.

ANY_Weights.txt has PCA components as the headers, and features (such as, genes or probes) at the start of subsequent rows.

PCAVvalues.tsv This has an initial header row, a row for FVE, and then rows giving sample ids and PCA values.

```
/bea_testing/output/PCA_DualBatch_Structures/BatchIdwithTSS/DualBatch$ ls -l
total 2120
-rw-r--r-- 1 linux linux      383 Jun 14 09:34 ANY_Comp1_Comp2_DSC.txt
-rw-r--r-- 1 linux linux      267 Jun 14 09:34 ANY_Comp1_Comp2_DSC.txt__CompDSC.RData
-rw-r--r-- 1 linux linux      383 Jun 14 09:34 ANY_Comp1_Comp3_DSC.txt
-rw-r--r-- 1 linux linux      267 Jun 14 09:34 ANY_Comp1_Comp3_DSC.txt__CompDSC.RData
-rw-r--r-- 1 linux linux     2266 Jun 14 09:34 ANY_DSC.txt
-rw-r--r-- 1 linux linux      923 Jun 14 09:34 ANY_DSC.txt__OverallDSC.RData
-rw-r--r-- 1 linux linux     1289 Jun 14 09:34 ANY_FVE.txt
-rw-r--r-- 1 linux linux    120163 Jun 14 09:34 ANY_Scores.txt
-rw-r--r-- 1 linux linux   2001324 Jun 14 09:34 ANY_Weights.txt
-rw-r--r-- 1 linux linux      639 Jun 14 09:34 PCAAnnotations.tsv
drwxr-xr-x 2 linux linux     4096 Jun 14 09:34 PCA-Plus
-rw-r--r-- 1 linux linux     6643 Jun 14 09:34 PCAValues.tsv
```

The PCA-Plus directory contains the different PCA-Plus PNG files and legends.

```
/bea_testing/output/PCA_DualBatch_Structures/BatchIdwithTSS/DualBatch/PCA-Plus$ ls -l
total 436
-rw-r--r-- 1 linux linux 120980 Jun 14 09:34 BatchIdwithTSS_Comp1_Comp2_Diagram.png
-rw-r--r-- 1 linux linux  51860 Jun 14 09:34 BatchIdwithTSS_Comp1_Comp2_Legend-ALL.png
-rw-r--r-- 1 linux linux  28370 Jun 14 09:34 BatchIdwithTSS_Comp1_Comp2_Legend-DSC.png
-rw-r--r-- 1 linux linux  14517 Jun 14 09:34 BatchIdwithTSS_Comp1_Comp2_Legend-Points.png
-rw-r--r-- 1 linux linux   3664 Jun 14 09:34 BatchIdwithTSS_Comp1_Comp2_Legend-Rays.png
-rw-r--r-- 1 linux linux 117616 Jun 14 09:34 BatchIdwithTSS_Comp1_Comp3_Diagram.png
-rw-r--r-- 1 linux linux  52066 Jun 14 09:34 BatchIdwithTSS_Comp1_Comp3_Legend-ALL.png
-rw-r--r-- 1 linux linux  28484 Jun 14 09:34 BatchIdwithTSS_Comp1_Comp3_Legend-DSC.png
-rw-r--r-- 1 linux linux  14517 Jun 14 09:34 BatchIdwithTSS_Comp1_Comp3_Legend-Points.png
-rw-r--r-- 1 linux linux   3664 Jun 14 09:34 BatchIdwithTSS_Comp1_Comp3_Legend-Rays.png
```

Here is a diagram generated from this code.

Example Title for Test Data BatchIdwithTSS 1 2

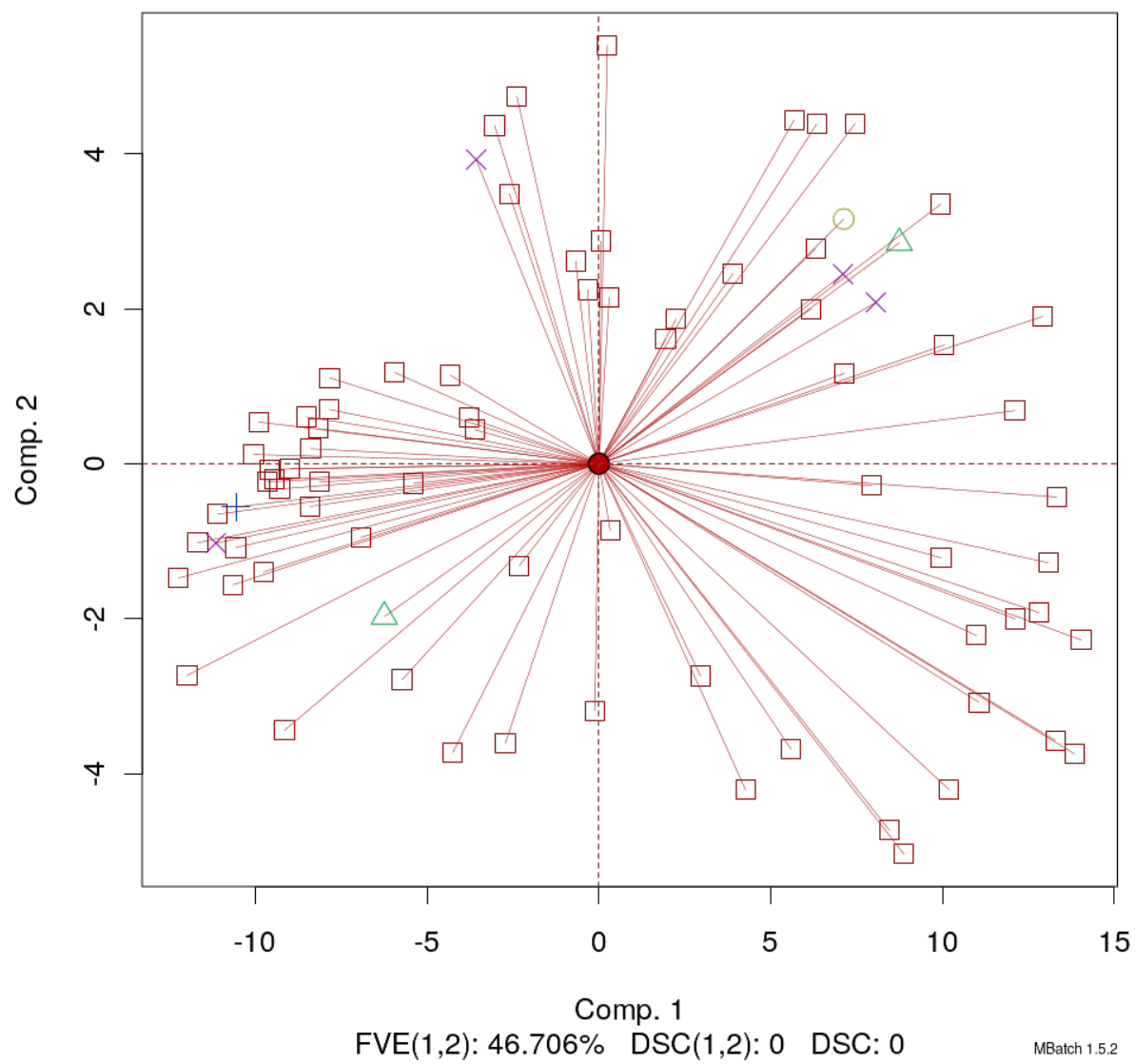


Figure 2: PCAPlus Output