MBatch 04-03
Using MBatch Assessments: HierarchicalClustering_Structures
Tod Casasent
2017-11-02-0900

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.docx for instructions on downloading test data.

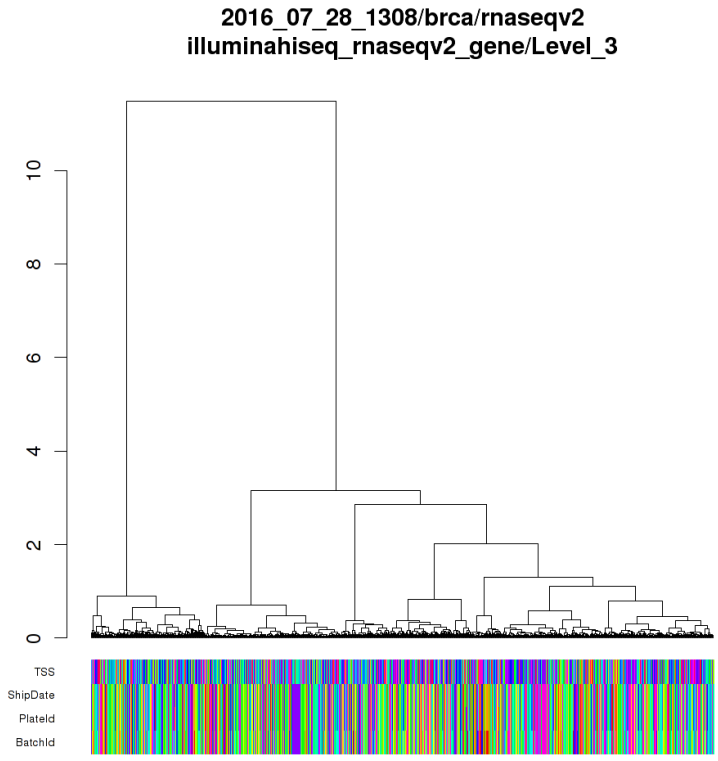
2 Algorithm

HierarchicalClustering_Structures is a function used to perform batch effects assessments using the hierarchical clustering algorithm. Hierarchical clustering is performed with each batch type available within the covariate bars.

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 covariate bar with the batch information with dendrograms for the clustering at the top. The covariate rows are batch types, while the columns are samples.



4 Usage

```
HierarchicalClustering_Structures(theData,
        theTitle,
        theOutputPath,
        theBatchTypeAndValuePairsToRemove=list(),
        theBatchTypeAndValuePairsToKeep=list())
```

5 Arguments

theData An instance of BEA_DATA BEA_DATA-class. This is the MBatch Data Object (of class BEA_DATA) described in MBatch_03_UserData.docx, and returned from mbatchLoadFiles or mbatchLoadStructures.

theTitle Object of class "character". Title to use in PNG files.

theOutputPath Object of class "character". Directory in which to place output PNG files and related data files used by the Batch Effects Website.

theBatchTypeAndValuePairsToRemove Object of class "list". A list of vectors containing the batch type (or * for all types) and the value to remove. list() or NULL indicates do nothing. This type of list is described in MBatch_04-00_ParametersBatchTypesValues.docx.

theBatchTypeAndValuePairsToKeep Object of class "list". A list of vectors containing the batch type (or * for all types) and a vector of the value(s) to keep. list() or NULL indicates do nothing. This type of list is described in MBatch_04-00_ParametersBatchTypesValues.docx.

6 Example Call

The following code performs Supervised Clustering and is taken from the tests/SupervisedClustering_Batches_Structures.R file. Data used is from the testing data as per the MBatch_01_InstallLinux.docx document.

```
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/HierarchicalClustering_Structures"

# 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
  HierarchicalClustering_Structures(theData=myData,
                                    theTitle="Test PCA",
                                    theOutputPath=theOutputDir)
```

6.1 Command Line Output

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.

```
> # set the paths
> theGeneFile <- "/bea_testing/MATRIX_DATA/matrix_data-Tumor.tsv"
> theBatchFile <- "/bea_testing/MATRIX_DATA/batches-Tumor.tsv"
> theOutputDir <- "/bea_testing/output/HierarchicalClustering_Structures"
>
> # 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)
2017 10 13 12:44:13.187 DEBUG machinename Changing LC_COLLATE to C for duration of run
2017 10 13 12:44:13.214 INFO machinename \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/
2017 10 13 12:44:13.215 INFO machinename Starting mbatchLoadFiles
2017 10 13 12:44:13.216 INFO machinename MBatch Version: 2017-09-19-1530
2017 10 13 12:44:13.217 INFO machinename read batch file= /bea_testing/MATRIX_DATA/batches-
Tumor.tsv
2017 10 13 12:44:13.338 INFO machinename read gene file= /bea_testing/MATRIX_DATA/matrix_data-
Tumor.tsv
Read 100000 records
2017 10 13 12:44:26.017 INFO machinename filter samples in batches using gene samples
2017 10 13 12:44:26.019 INFO machinename sort batches by gene file samples
2017 10 13 12:44:26.142 INFO machinename Finishing mbatchLoadFiles
2017 10 13 12:44:26.143 INFO machinename ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
>   myData@mData <- mbatchTrimData(myData@mData, 100000)
2017 10 13 12:44:26.144 DEBUG machinename Changing LC_COLLATE to C for duration of run
2017 10 13 12:44:26.145 INFO machinename \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/
2017 10 13 12:44:26.145 INFO machinename mbatchTrimData Starting
2017 10 13 12:44:26.146 INFO machinename MBatch Version: 2017-09-19-1530
2017 10 13 12:44:34.148 INFO machinename mbatchTrimData Finishing
2017 10 13 12:44:34.149 INFO machinename ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
>
>   # here, we take most defaults
>   HierarchicalClustering_Structures(theData=myData,
+                                     theTitle="Test PCA",
+                                     theOutputPath=theOutputDir)
2017 10 13 12:44:34.153 DEBUG machinename Changing LC_COLLATE to C for duration of run
2017 10 13 12:44:34.153 INFO machinename \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/
2017 10 13 12:44:34.153 INFO machinename mbatchFilterData Starting
2017 10 13 12:44:34.154 INFO machinename MBatch Version: 2017-09-19-1530
2017 10 13 12:44:34.154 DEBUG machinename rows pre filter  1250
2017 10 13 12:44:34.421 DEBUG machinename rows post filter  1250
2017 10 13 12:44:34.422 DEBUG machinename mbatchFilterData Prefilter, gene data had  1250  while
post filter  1250
```

```
2017 10 13 12:44:34.423 DEBUG machinename mbatchFilterData Prefilter, batch data had  80  while
post filter  80
2017 10 13 12:44:34.424 INFO machinename mbatchFilterData Finishing
2017 10 13 12:44:34.424 INFO machinename ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
2017 10 13 12:44:34.425 DEBUG machinename hierclust_outputGraphics
2017 10 13 12:44:34.425 DEBUG machinename Changing LC_COLLATE to C for duration of run
2017 10 13 12:44:34.426 DEBUG machinename hierclust_calc
2017 10 13 12:44:34.427 DEBUG machinename Changing LC_COLLATE to C for duration of run
2017 10 13 12:44:34.429 DEBUG machinename calculating HC
The "ward" method has been renamed to "ward.D"; note new "ward.D2"
2017 10 13 12:44:34.466 DEBUG machinename checkCreateDir:
/bea_testing/output/HierarchicalClustering_Structures/
2017 10 13 12:44:34.469 DEBUG machinename hierclust_draw
2017 10 13 12:44:34.470 DEBUG machinename Changing LC_COLLATE to C for duration of run
2017 10 13 12:44:34.489 DEBUG machinename writeClusteringImage
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Diagram.png
2017 10 13 12:44:34.661 DEBUG machinename writeIndividualLegend
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-BatchId.png
2017 10 13 12:44:34.664 DEBUG machinename mbatchStandardLegend - Calling .jinit
/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/LegendJava.jar:/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/jcommon-1.0.17.jar:/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/jfreechart-1.0.14.jar
2017 10 13 12:44:36.091 DEBUG machinename mbatchStandardLegend - .jinit complete
2017 10 13 12:44:36.091 DEBUG machinename mbatchStandardLegend - theTitle  BatchId
2017 10 13 12:44:36.092 DEBUG machinename mbatchStandardLegend - theVersion  MBatch 1.4.16
2017 10 13 12:44:36.093 DEBUG machinename mbatchStandardLegend - theFilenamePath
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-BatchId.png
2017 10 13 12:44:36.093 DEBUG machinename mbatchStandardLegend - theLegendNames  00304 (80)
2017 10 13 12:44:36.094 DEBUG machinename mbatchStandardLegend - theLegendNames  1
2017 10 13 12:44:36.095 DEBUG machinename mbatchStandardLegend - theLegendColors  1
2017 10 13 12:44:36.095 DEBUG machinename mbatchStandardLegend - theLegendSymbols  0
2017 10 13 12:44:36.096 DEBUG machinename mbatchStandardLegend - myColors  #0066ff
2017 10 13 12:44:36.097 DEBUG machinename mbatchStandardLegend before java
LegendJava 2013_05_03_0823
writeLegendWithSymbols theTitle = BatchId
writeLegendWithSymbols theVersion = MBatch 1.4.16
writeLegendWithSymbols theFilenamePath =
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-BatchId.png
Colors is non-null
writeLegendWithSymbols write
writeLegendWithSymbols done
2017 10 13 12:44:44.734 DEBUG machinename mbatchStandardLegend after java
2017 10 13 12:44:44.752 DEBUG machinename writeIndividualLegend
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-PlateId.png
2017 10 13 12:44:44.782 DEBUG machinename mbatchStandardLegend - Calling .jinit
/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/LegendJava.jar:/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/jcommon-1.0.17.jar:/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/jfreechart-1.0.14.jar
2017 10 13 12:44:44.867 DEBUG machinename mbatchStandardLegend - .jinit complete
```

2017 10 13 12:44:44.868 DEBUG machinename mbatchStandardLegend - theTitle PlateId
2017 10 13 12:44:44.869 DEBUG machinename mbatchStandardLegend - theVersion MBatch 1.4.16
2017 10 13 12:44:44.869 DEBUG machinename mbatchStandardLegend - theFilenamePath
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-PlateId.png
2017 10 13 12:44:44.869 DEBUG machinename mbatchStandardLegend - theLegendNames A29J (80)
2017 10 13 12:44:44.870 DEBUG machinename mbatchStandardLegend - theLegendNames 1
2017 10 13 12:44:44.870 DEBUG machinename mbatchStandardLegend - theLegendColors 1
2017 10 13 12:44:44.870 DEBUG machinename mbatchStandardLegend - theLegendSymbols 0
2017 10 13 12:44:44.871 DEBUG machinename mbatchStandardLegend - myColors #0066ff
2017 10 13 12:44:44.871 DEBUG machinename mbatchStandardLegend before java
LegendJava 2013_05_03_0823
writeLegendWithSymbols theTitle = PlateId
writeLegendWithSymbols theVersion = MBatch 1.4.16
writeLegendWithSymbols theFilenamePath =
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-PlateId.png
Colors is non-null
writeLegendWithSymbols write
writeLegendWithSymbols done
2017 10 13 12:44:44.889 DEBUG machinename mbatchStandardLegend after java
2017 10 13 12:44:44.890 DEBUG machinename writeIndividualLegend
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-ShipDate.png
2017 10 13 12:44:44.893 DEBUG machinename mbatchStandardLegend - Calling .jinit
/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/LegendJava.jar:/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/jcommon-1.0.17.jar:/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/jfreechart-1.0.14.jar
2017 10 13 12:44:44.901 DEBUG machinename mbatchStandardLegend - .jinit complete
2017 10 13 12:44:44.902 DEBUG machinename mbatchStandardLegend - theTitle ShipDate
2017 10 13 12:44:44.902 DEBUG machinename mbatchStandardLegend - theVersion MBatch 1.4.16
2017 10 13 12:44:44.903 DEBUG machinename mbatchStandardLegend - theFilenamePath
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-ShipDate.png
2017 10 13 12:44:44.903 DEBUG machinename mbatchStandardLegend - theLegendNames 2013-05-08
(80)
2017 10 13 12:44:44.904 DEBUG machinename mbatchStandardLegend - theLegendNames 1
2017 10 13 12:44:44.904 DEBUG machinename mbatchStandardLegend - theLegendColors 1
2017 10 13 12:44:44.905 DEBUG machinename mbatchStandardLegend - theLegendSymbols 0
2017 10 13 12:44:44.905 DEBUG machinename mbatchStandardLegend - myColors #0066ff
2017 10 13 12:44:44.906 DEBUG machinename mbatchStandardLegend before java
LegendJava 2013_05_03_0823
writeLegendWithSymbols theTitle = ShipDate
writeLegendWithSymbols theVersion = MBatch 1.4.16
writeLegendWithSymbols theFilenamePath =
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-ShipDate.png
Colors is non-null
writeLegendWithSymbols write
writeLegendWithSymbols done
2017 10 13 12:44:44.922 DEBUG machinename mbatchStandardLegend after java
2017 10 13 12:44:44.923 DEBUG machinename writeIndividualLegend
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-TSS.png
2017 10 13 12:44:44.926 DEBUG machinename mbatchStandardLegend - Calling .jinit
/home/linux/R/x86_64-pc-linux-gnu-

```
library/3.4/MBatch/LegendJava/LegendJava.jar:/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/jcommon-1.0.17.jar:/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/jfreechart-1.0.14.jar
2017 10 13 12:44:44.934 DEBUG machinename mbatchStandardLegend - .jinit complete
2017 10 13 12:44:44.935 DEBUG machinename mbatchStandardLegend - theTitle  TSS
2017 10 13 12:44:44.935 DEBUG machinename mbatchStandardLegend - theVersion  MBatch 1.4.16
2017 10 13 12:44:44.936 DEBUG machinename mbatchStandardLegend - theFilenamePath
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-TSS.png
2017 10 13 12:44:44.936 DEBUG machinename mbatchStandardLegend - theLegendNames  OR -
University of Michigan (72), OU - Roswell Park (1), P6 - Translational Genomics Research Institute (2),
PA - University of Minnesota (1), PK - University Health Network (4)
2017 10 13 12:44:44.936 DEBUG machinename mbatchStandardLegend - theLegendNames  5
2017 10 13 12:44:44.937 DEBUG machinename mbatchStandardLegend - theLegendColors  5
2017 10 13 12:44:44.937 DEBUG machinename mbatchStandardLegend - theLegendSymbols  0
2017 10 13 12:44:44.937 DEBUG machinename mbatchStandardLegend - myColors
#0066ff,#00ff66,#ccff00,#ff0000,#cc00ff
2017 10 13 12:44:44.938 DEBUG machinename mbatchStandardLegend before java
LegendJava 2013_05_03_0823
writeLegendWithSymbols theTitle = TSS
writeLegendWithSymbols theVersion = MBatch 1.4.16
writeLegendWithSymbols theFilenamePath =
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-TSS.png
Colors is non-null
writeLegendWithSymbols write
writeLegendWithSymbols done
2017 10 13 12:44:44.997 DEBUG machinename mbatchStandardLegend after java
2017 10 13 12:44:44.998 DEBUG machinename writeCombinedLegendHC
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-ALL.png
2017 10 13 12:44:45.000 DEBUG machinename mbatchStandardCombineLegends - Calling .jinit
/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/LegendJava.jar:/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/jcommon-1.0.17.jar:/home/linux/R/x86_64-pc-linux-gnu-
library/3.4/MBatch/LegendJava/jfreechart-1.0.14.jar
2017 10 13 12:44:45.010 DEBUG machinename mbatchStandardCombineLegends - .jinit complete
2017 10 13 12:44:45.010 DEBUG machinename mbatchStandardCombineLegends - theTitle  Test PCA
2017 10 13 12:44:45.011 DEBUG machinename mbatchStandardCombineLegends - theFilenamePath
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-ALL.png
2017 10 13 12:44:45.046 DEBUG machinename mbatchStandardCombineLegends - theListOfFiles
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-BatchId.png,
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-PlateId.png,
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-ShipDate.png,
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-TSS.png
2017 10 13 12:44:45.047 DEBUG machinename mbatchStandardLegend before java
LegendJava 2013_05_03_0823
combineLegends theTitle = Test PCA
combineLegends theFilenamePath =
/bea_testing/output/HierarchicalClustering_Structures//HierarchicalClustering_Legend-ALL.png
combineLegends write
combineLegends done
2017 10 13 12:44:45.322 DEBUG machinename mbatchStandardLegend after java
NULL
```

6.2 Example File Output

The above code creates the following output files. Files are named using the following naming convention:

HierarchicalClustering_Diagram.png

HierarchicalClustering_Legend-ALL.png

HierarchicalClustering_Legend-<BatchType>.png

The diagram file contains a Hierarchical Clustering plot for all batch types (the columns from the batches.tsv file). The legends give the list of batches within the given batch type, and the colors used for each batch within the covariate bars. The "ALL" Legend has all batch types.

The two TSV files are used internally for display of dynamic Hierarchical Clustering diagrams on the Batch Effect Website.

```
linux@machinename:/bea_testing/output/HierarchicalClustering_Structures$ ls -l
total 76
-rw-r--r-- 1 linux linux  1952 Oct 13 12:44 HCData.tsv
-rw-r--r-- 1 linux linux  2560 Oct 13 12:44 HCOrder.tsv
-rw-r--r-- 1 linux linux 15883 Oct 13 12:44 HierarchicalClustering_Diagram.png
-rw-r--r-- 1 linux linux 23574 Oct 13 12:44 HierarchicalClustering_Legend-ALL.png
-rw-r--r-- 1 linux linux  2697 Oct 13 12:44 HierarchicalClustering_Legend-BatchId.png
-rw-r--r-- 1 linux linux  2840 Oct 13 12:44 HierarchicalClustering_Legend-PlateId.png
-rw-r--r-- 1 linux linux  3191 Oct 13 12:44 HierarchicalClustering_Legend-ShipDate.png
-rw-r--r-- 1 linux linux 12976 Oct 13 12:44 HierarchicalClustering_Legend-TSS.png
```