## Introduction

### 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.

## Algorithm

MP Overall performs a Median Polish Overall correction taking a BEA\_DATA object (with data matrix and batch dataframe) and returning either a corrected matrix or a string containing the path to where the data file was written.

## Output

The primary output method for MBatch is to view results in the Batch Effects Website. Correction algorithms generally do not create graphical output and instead create TSV output files.

# Usage

 $\label{eq:mp_bybatch} \mbox{MP\_ByBatch(theBeaData, theBatchType, thePath} = \mbox{NULL, theWriteToFile} = \mbox{FALSE)}$ 

# Arguments

#### theBeaData

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.

### the Batch Type

A string identifying the batch type to correct.

#### thePath

Output path for any files.

#### theWriteToFile

TRUE to write the corrected data to file and return the file pathname instead of the corrected matrix.

## **Example Call**

The following code is adapted from the tests/MP\_ByBatch.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.

```
{
  library(MBatch)

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

# set the paths
  theGeneFile=file.path(inputDir, "matrix_data-Tumor.tsv")
  theBatchFile=file.path(inputDir, "batches-Tumor.tsv")
  theOutputDir=file.path(outputDir, "MP_ByBatch")
  theRandomSeed=314
  theBatchType="TSS"

# make sure the output dir exists and is empty
  unlink(theOutputDir, recursive=TRUE)</pre>
```

```
dir.create(theOutputDir, showWarnings=FALSE, recursive=TRUE)
    myData <- mbatchLoadFiles(theGeneFile, theBatchFile)</pre>
    myData@mData <- mbatchTrimData(myData@mData, 100000)</pre>
     outputFile <- MP_ByBatch(theBeaData=myData,</pre>
                                                               theBatchType=theBatchType,
                                                               thePath=theOutputDir,
                                                               theWriteToFile=TRUE)
     correctedMatrix <- readAsGenericMatrix(outputFile)</pre>
    print(correctedMatrix[1:4, 1:4])
}
## 2020 11 18 16:19:51.157 DEBUG ab7c64738d52 Changing LC_COLLATE to C for duration of run
## 2020 11 18 16:19:51.157 INFO ab7c64738d52 \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/
## 2020 11 18 16:19:51.157 INFO ab7c64738d52 Starting mbatchLoadFiles
## 2020 11 18 16:19:51.158 INFO ab7c64738d52 MBatch Version: BEA_VERSION_TIMESTAMP
## 2020 11 18 16:19:51.158 INFO ab7c64738d52 read batch file= /builds/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/
## 2020 11 18 16:19:51.159 INFO ab7c64738d52 read gene file= /builds/BatchEffects_clean/Batch
## 2020 11 18 16:19:53.561 INFO ab7c64738d52 filter samples in batches using gene samples
## 2020 11 18 16:19:53.562 INFO ab7c64738d52 sort batches by gene file samples
## 2020 11 18 16:19:53.611 INFO ab7c64738d52 Finishing mbatchLoadFiles
## 2020 11 18 16:19:53.611 INFO ab7c64738d52
## 2020 11 18 16:19:53.611 DEBUG ab7c64738d52 Changing LC_COLLATE to C for duration of run
## 2020 11 18 16:19:53.611 INFO ab7c64738d52 \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/ \/
## 2020 11 18 16:19:53.612 INFO ab7c64738d52 mbatchTrimData Starting
## 2020 11 18 16:19:53.612 INFO ab7c64738d52 MBatch Version: BEA_VERSION_TIMESTAMP
## 2020 11 18 16:20:01.053 INFO ab7c64738d52 mbatchTrimData theMaxSize= 1e+05
## 2020 11 18 16:20:01.053 INFO ab7c64738d52 mbatchTrimData ncol(theMatrix)= 80
## 2020 11 18 16:20:01.053 INFO ab7c64738d52 mbatchTrimData nrow(theMatrix)= 1250
## 2020 11 18 16:20:01.053 INFO ab7c64738d52 mbatchTrimData Finishing
## 2020 11 18 16:20:01.054 INFO ab7c64738d52
## 2020 11 18 16:20:01.054 INFO ab7c64738d52 MP_Internal - starting
## 2020 11 18 16:20:01.054 DEBUG ab7c64738d52 checkCreateDir: /builds/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/BatchEffects_clean/
## 2020 11 18 16:20:01.295 DEBUG ab7c64738d52 starting BeaMP
## 2020 11 18 16:20:01.296 DEBUG ab7c64738d52 starting MP
## 2020 11 18 16:20:01.297 DEBUG ab7c64738d52 MP batch
## 2020 11 18 16:20:01.297 DEBUG ab7c64738d52 convertDataFrameToSi start
## 2020 11 18 16:20:01.297 DEBUG ab7c64738d52 convertDataFrameToSi asmatrixWithIssues
## 2020 11 18 16:20:01.298 DEBUG ab7c64738d52 convertDataFrameToSi rownames
## 2020 11 18 16:20:01.298 DEBUG ab7c64738d52 convertDataFrameToSi colnames
## 2020 11 18 16:20:01.298 DEBUG ab7c64738d52 convertDataFrameToSi done
## 2020 11 18 16:20:02.355 DEBUG ab7c64738d52 finishing BeaMP
## 2020 11 18 16:20:02.355 TIMING ab7c64738d52
                                                                                                                1.054
                                                                                                                                    1.06
                                                                                                                                                       MPByBatch
                                                                                                                                                                                  /builds/BatchE
## 2020 11 18 16:20:02.356 DEBUG ab7c64738d52 Write to file /builds/BatchEffects clean/Bat
## 2020 11 18 16:20:02.458 DEBUG ab7c64738d52 Finished write to file /builds/BatchEffects_
```

```
## 2020 11 18 16:20:02.459 INFO ab7c64738d52 MP_Internal - completed
                              TCGA-OR-A5J1-01A-11D-A29J-05
##
## ABR-cg06968724-17-1012579
                                                 -0.3519384
## ABR-cg23568341-17-1011974
                                                 -0.3830757
## ABR-cg24479027-17-1012576
                                                 -0.3482736
## ACOT7-cg16034168-1-6336711
                                                  0.4385596
##
                              TCGA-OR-A5J2-01A-11D-A29J-05
## ABR-cg06968724-17-1012579
                                                  0.4393411
## ABR-cg23568341-17-1011974
                                                  0.4134679
## ABR-cg24479027-17-1012576
                                                  0.4451149
## ACOT7-cg16034168-1-6336711
                                                  0.3473411
##
                              TCGA-OR-A5J3-01A-11D-A29J-05
## ABR-cg06968724-17-1012579
                                                 0.97342791
## ABR-cg23568341-17-1011974
                                                 0.87449968
## ABR-cg24479027-17-1012576
                                                 0.98826271
## ACOT7-cg16034168-1-6336711
                                                 0.01531393
##
                              TCGA-OR-A5J4-01A-11D-A29J-05
## ABR-cg06968724-17-1012579
                                                  0.5599392
## ABR-cg23568341-17-1011974
                                                  0.5152570
## ABR-cg24479027-17-1012576
                                                  0.5571871
## ACOT7-cg16034168-1-6336711
                                                  0.4181173
```

## Example File Output

The above code creates the following output file. File is named using the following naming convention: ANY\_Corrections-MPByBatch.tsv The TSV file with the corrected dataset is written by the MBatch package. The end of the output shows a snippet from the corrected matrix.