

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

```
MP_Overall(theBeaData, thePath = NULL, theWriteToFile = 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.

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_Overall.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_Overall")
  theRandomSeed=314
  theBatchType="TSS"

  # 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)
```

```
# call
outputFile <- MP_Overall(theBeaData=myData,
                        thePath=theOutputDir,
                        theWriteToFile=TRUE)
correctedMatrix <- readAsGenericMatrix(outputFile)
print(correctedMatrix[1:4, 1:4])
}

## 2020 11 18 16:19:40.363 DEBUG ab7c64738d52 Changing LC_COLLATE to C for duration of run
## 2020 11 18 16:19:40.364 INFO ab7c64738d52 \ / \ / \ / \ / \ / \ / \ / \ / \ /
## 2020 11 18 16:19:40.364 INFO ab7c64738d52 Starting mbatchLoadFiles
## 2020 11 18 16:19:40.364 INFO ab7c64738d52 MBatch Version: BEA_VERSION_TIMESTAMP
## 2020 11 18 16:19:40.364 INFO ab7c64738d52 read batch file= /builds/BatchEffects_clean/Bat
## 2020 11 18 16:19:40.365 INFO ab7c64738d52 read gene file= /builds/BatchEffects_clean/Bato
## 2020 11 18 16:19:42.854 INFO ab7c64738d52 filter samples in batches using gene samples
## 2020 11 18 16:19:42.855 INFO ab7c64738d52 sort batches by gene file samples
## 2020 11 18 16:19:42.895 INFO ab7c64738d52 Finishing mbatchLoadFiles
## 2020 11 18 16:19:42.895 INFO ab7c64738d52 ~~~~~
## 2020 11 18 16:19:42.895 DEBUG ab7c64738d52 Changing LC_COLLATE to C for duration of run
## 2020 11 18 16:19:42.896 INFO ab7c64738d52 \ / \ / \ / \ / \ / \ / \ / \ / \ /
## 2020 11 18 16:19:42.896 INFO ab7c64738d52 mbatchTrimData Starting
## 2020 11 18 16:19:42.896 INFO ab7c64738d52 MBatch Version: BEA_VERSION_TIMESTAMP
## 2020 11 18 16:19:50.358 INFO ab7c64738d52 mbatchTrimData theMaxSize= 1e+05
## 2020 11 18 16:19:50.359 INFO ab7c64738d52 mbatchTrimData ncol(theMatrix)= 80
## 2020 11 18 16:19:50.359 INFO ab7c64738d52 mbatchTrimData nrow(theMatrix)= 1250
## 2020 11 18 16:19:50.359 INFO ab7c64738d52 mbatchTrimData Finishing
## 2020 11 18 16:19:50.359 INFO ab7c64738d52 ~~~~~
## 2020 11 18 16:19:50.360 INFO ab7c64738d52 MP_Internal - starting
## 2020 11 18 16:19:50.360 DEBUG ab7c64738d52 checkCreateDir: /builds/BatchEffects_clean/Ba
## 2020 11 18 16:19:50.604 DEBUG ab7c64738d52 starting BeaMP
## 2020 11 18 16:19:50.605 DEBUG ab7c64738d52 starting MP
## 2020 11 18 16:19:50.605 DEBUG ab7c64738d52 MP overall
## 2020 11 18 16:19:50.965 DEBUG ab7c64738d52 finishing BeaMP
## 2020 11 18 16:19:50.965 TIMING ab7c64738d52 0.3580000000000004 0.363 MPOverall /bu
## 2020 11 18 16:19:50.966 DEBUG ab7c64738d52 Write to file /builds/BatchEffects_clean/Bato
## 2020 11 18 16:19:51.068 DEBUG ab7c64738d52 Finished write to file /builds/BatchEffects_c
## 2020 11 18 16:19:51.069 INFO ab7c64738d52 MP_Internal - completed
## TCGA-OR-A5J1-01A-11D-A29J-05
## ABR-cg06968724-17-1012579 -0.3515316
## ABR-cg23568341-17-1011974 -0.3788064
## ABR-cg24479027-17-1012576 -0.3478635
## ACOT7-cg16034168-1-6336711 0.4332530
## TCGA-OR-A5J2-01A-11D-A29J-05
## ABR-cg06968724-17-1012579 0.4391718
## ABR-cg23568341-17-1011974 0.4171611
## ABR-cg24479027-17-1012576 0.4449490
```

```

## ACOT7-cg16034168-1-6336711      0.3414584
##                                TCGA-OR-A5J3-01A-11D-A29J-05
## ABR-cg06968724-17-1012579      0.973778288
## ABR-cg23568341-17-1011974      0.878712547
## ABR-cg24479027-17-1012576      0.988616482
## ACOT7-cg16034168-1-6336711      0.009950896
##                                TCGA-OR-A5J4-01A-11D-A29J-05
## ABR-cg06968724-17-1012579      0.5607078
## ABR-cg23568341-17-1011974      0.5198881
## ABR-cg24479027-17-1012576      0.5579590
## ACOT7-cg16034168-1-6336711      0.4131724

```

Example File Output

The above code creates the following output file. File is named using the following naming convention: ANY__Corrections-MPOverall.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.