

Using MBatch Corrections: EBNPlus__TrainAndValidateFromVector__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

EBNPlus__TrainAndValidateFromVector__Structures creates and returns a training and validation set and corrected data, based on the training and validation set and provided vectors of samples to use for training and validation, instead of using replicates and percentages.

3 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.

4 Usage

EBNPlus__TrainAndValidateFromVector__Structures(theDataMatrix1, theDataMatrix2, theBatchId1, theBatchId2, theEBNP_PsuedoReplicates1Train, theEBNP_PsuedoReplicates2Train, theEBNP_PsuedoReplicates1Validation, theEBNP_PsuedoReplicates2Validation, theEBNP_BatchWithZero, theEBNP_FixDataSet, theEBNP_CorrectForZero, theEBNP_ParametricPriorsFlag, theEBNP_TestRatio=0, theSeed=NULL, theTestSeed=NULL, theEBNP_PriorPlotsFile = NULL, theEBNP_MinSampleNum = 3)

5 Arguments

5.1 theDataMatrix1

A matrix for data set 1 containing numeric values with columns being sample ids and rows being feature (like gene or probe) ids.

5.2 theDataMatrix2

A matrix for data set 2 containing numeric values with columns being sample ids and rows being feature (like gene or probe) ids.

5.3 theBatchId1

The group name for data set 1 (for example, RNASeqV2)

5.4 theBatchId2

The group name for data set 2 (for example, Agilent4502)

5.5 theEBNP_PsuedoReplicates1Train

A vector of sample ids to use as replicates for data set 1 for training.

5.6 theEBNP_PsuedoReplicates2Train

A vector of sample ids to use as replicates for data set 2 for training.

5.7 theEBNP_PsuedoReplicates1Validation

A vector of sample ids to use as replicates for data set 1 for validation. Use NULL if no validation is desired. This means corrections will be done.

5.8 theEBNP_PsuedoReplicates2Validation

A vector of sample ids to use as replicates for data set 2 for validation. Use NULL if no validation is desired. This means corrections will be done.

5.9 theEBNP_BatchWithZero

Batch With Zero indicates which data set contains zero values. This is a string “1”, “2”, or “both”. These values will be changed in a future release.

5.10 theEBNP_FixDataSet

Fix Data Set indicates which data set should be set as invariate during the corrections. Value are numeric, 1 or 2 for the appropriate data set, or as.numeric(NA) for neither.

5.11 theEBNP_CorrectForZero

Correct For Zero indicates whether or not data sets marked as “Batch With Zero” should be modified from zero (TRUE or FALSE)

5.12 theEBNP_ParametricPriorsFlag

Use parametric adjustments for corrections (TRUE or FALSE)

5.13 theSeed

NULL or a seed to use. Defaults to NULL. This is passed to the standard R `set.seed` function.

5.14 theTestSeed

NULL or a seed to use. Defaults to NULL. This is passed to the standard R `set.seed` function. This is used only when creating Test Set.

5.15 theEBNP_PriorPlotsFile

Defaults to NULL and does not create PNG. Non-NULL should be the full path and filename for where to create the prior plots PNG.

5.16 theEBNP_MinSampleNum

Defaults to 3. Any row (gene) with less than this number of samples is dropped.

6 Example Call

The following code is adapted from the `tests/EB_withNonParametricPriors.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)

theDataFile1="/bea_testing/MATRIX_DATA/brca_rnaseq2_matrix_data.tsv"
theDataFile2="/bea_testing/MATRIX_DATA/brca_agi4502_matrix_data.tsv"
theOutputDir="/bea_testing/output/EBNPlus_TrainAndValidateReplicates_Structures"
theBatchId1="RNASeqV2"
theBatchId2="Agilent4502"
theRandomSeed=314

# trim genes to get just gene symbols from standardized data
trimGenes <- function(theGenes)
{
  foo <- as.vector(unlist(
    sapply(theGenes, function(theGene)
    {
      # keep the same if it starts with ?
      if (TRUE==grepl("^[?]+", theGene))
      {
        return(theGene)
      }
      else
      {
        # split on the | and take the first argument
        # this makes no change if no pipe
        return(strsplit(theGene, "|", fixed=TRUE)[[1]][1])
      }
    }
  ))
}
```

```

    }
  })
})
foo
}

# remove duplicates from columns (samples)
removeDuplicatesFromColumns <- function(theMatrix)
{
  indexOfDuplicats <- which(duplicated(colnames(theMatrix)))
  if (length(indexOfDuplicats) > 0)
  {
    # minus sign uses inverse of indexes
    theMatrix <- theMatrix[ , -indexOfDuplicats]
  }
  return(theMatrix)
}

# remove duplicates from rows (genes/probes)
removeDuplicatesFromRows <- function(theMatrix)
{
  indexOfDuplicats <- which(duplicated(rownames(theMatrix)))
  if (length(indexOfDuplicats) > 0)
  {
    # minus sign uses inverse of indexes
    theMatrix <- theMatrix[-indexOfDuplicats, ]
  }
  return(theMatrix)
}

printMatrix <- function(theMatrix)
{
  print(is.matrix(theMatrix))
  print(dim(theMatrix))
  rowMax <- dim(theMatrix)[1]
  colMax <- dim(theMatrix)[2]
  rowMax <- min(rowMax, 4)
  colMax <- min(colMax, 4)
  print(theMatrix[1:rowMax, 1:colMax])
}

if ((!dir.exists(theDataFile1)) && (!dir.exists(theDataFile2)))
{
  warnLevel <- getOption("warn")
  on.exit(options(warn=warnLevel))
  # warnings are errors
  options(warn=3)
  # if there is a warning, show the calls leading up to it
  options(showWarnCalls=TRUE)
  # if there is an error, show the calls leading up to it
  options(showErrorCalls=TRUE)
  #

```

```

unlink(theOutputDir, recursive=TRUE)
dir.create(theOutputDir, showWarnings=FALSE, recursive=TRUE)
# read the files in. This can be done however you want
print("read the files")
theDataMatrix1 <- readAsGenericMatrix(theDataFile1)
theDataMatrix2 <- readAsGenericMatrix(theDataFile2)
# this is the reduce genes to just gene symbols, handling those from standardized data
print("reduce to gene symbols")
rownames(theDataMatrix1) <- trimGenes(rownames(theDataMatrix1))
rownames(theDataMatrix2) <- trimGenes(rownames(theDataMatrix2))
# remove any duplicates (this is a requirement for EBNplus)
print("remove duplicates")
theDataMatrix1 <- removeDuplicatesFromColumns(removeDuplicatesFromRows(theDataMatrix1))
theDataMatrix2 <- removeDuplicatesFromColumns(removeDuplicatesFromRows(theDataMatrix2))
# pseudoreplicate vectors
theEBNP_PsuedoReplicates1Train <- c("TCGA-A1-AOSH-01A-11R-A084-07", "TCGA-A1-AOSJ-01A-11R-A084-07",
theEBNP_PsuedoReplicates2Train <- c("TCGA-A1-AOSH-01A-11R-A084-07", "TCGA-A1-AOSJ-01A-11R-A084-07",
theEBNP_PsuedoReplicates1Validation <- c("TCGA-A1-AOSD-01A-11R-A115-07", "TCGA-A1-AOSK-01A-12R-A084-07",
theEBNP_PsuedoReplicates2Validation <- c("TCGA-A1-AOSD-01A-11R-A115-07", "TCGA-A1-AOSK-01A-12R-A084-07",
# call function
print("EBNplus_TrainAndValidateReplicates_Structures")
resultsList <- EBNplus_TrainAndValidateFromVector_Structures(theDataMatrix1, theDataMatrix2,
  theBatchId1, theBatchId2,
  theEBNP_PsuedoReplicates1Train,
  theEBNP_PsuedoReplicates2Train,
  theEBNP_PsuedoReplicates1Validation,
  theEBNP_PsuedoReplicates2Validation,
  theEBNP_BatchWithZero="1",
  theEBNP_FixDataSet=as.numeric(NA),
  theEBNP_CorrectForZero=TRUE,
  theEBNP_ParametricPriorsFlag=TRUE,
  theEBNP_TestRatio=0.3,
  theSeed=theRandomSeed,
  theTestSeed=theRandomSeed,
  theEBNP_PriorPlotsFile=file.path(theOutputDir, "priorplots.PNG"))
print("TestSet1")
printMatrix(resultsList$TestSet1)
print("TestSet2")
printMatrix(resultsList$TestSet2)
print("TrainingSet1")
printMatrix(resultsList$TrainingSet1)
print("TrainingSet2")
printMatrix(resultsList$TrainingSet2)
print("TrainingResults")
printMatrix(resultsList$TrainingResults)
print("ValidationSet1")
printMatrix(resultsList$ValidationSet1)
print("ValidationSet2")
printMatrix(resultsList$ValidationSet2)
print("ValidationResults")
printMatrix(resultsList$ValidationResults)
print("CorrectedResults")
printMatrix(resultsList$CorrectedResults)

```

```
}
}
```

```
## [1] "read the files"
## [1] "reduce to gene symbols"
## [1] "remove duplicates"
## [1] "EBNPlus_TrainAndValidateReplicates_Structures"
## 2019 10 10 11:20:01.899 DEBUG megazone23 EBNPlus_TrainAndValidateFromVector_Structures - start
## 2019 10 10 11:20:01.899 DEBUG megazone23 removeDuplicatessFromRows
## 2019 10 10 11:20:01.900 DEBUG megazone23 removeDuplicatessFromRows
## 2019 10 10 11:20:01.900 DEBUG megazone23 removeDuplicatessFromColumns
## 2019 10 10 11:20:01.900 DEBUG megazone23 removeDuplicatessFromColumns
## 2019 10 10 11:20:01.901 DEBUG megazone23 EBNPlus theEBNP_PriorPlotsFile= /bea_testing/output/EBNPlus
## 2019 10 10 11:20:01.901 DEBUG megazone23 dim(theDataMatrix1)=10000 1215
## 2019 10 10 11:20:01.901 DEBUG megazone23 dim(theDataMatrix2)=10000 600
## 2019 10 10 11:20:01.901 DEBUG megazone23 remove unknown genes, that start with ?
## 2019 10 10 11:20:01.981 DEBUG megazone23 dim(theDataMatrix1)=10000 1215
## 2019 10 10 11:20:01.981 DEBUG megazone23 dim(theDataMatrix2)=10000 600
## 2019 10 10 11:20:01.982 DEBUG megazone23 make EBNplus
## 2019 10 10 11:20:01.985 DEBUG megazone23 before callNextMethod
## 2019 10 10 11:20:01.986 DEBUG megazone23 dim(mData1)=10000 1215
## 2019 10 10 11:20:01.986 DEBUG megazone23 dim(mData2)=10000 600
## 2019 10 10 11:20:01.986 DEBUG megazone23 after callNextMethod
## 2019 10 10 11:20:01.986 DEBUG megazone23 dim(mData1)=10000 1215
## 2019 10 10 11:20:01.987 DEBUG megazone23 dim(mData2)=10000 600
## 2019 10 10 11:20:01.987 DEBUG megazone23 after EBNplus
## 2019 10 10 11:20:01.987 DEBUG megazone23 dim(ebObj@mData1)=10000 1215
## 2019 10 10 11:20:01.987 DEBUG megazone23 dim(ebObj@mData2)=10000 600
## 2019 10 10 11:20:01.989 DEBUG megazone23 getBiComOrder
## 2019 10 10 11:20:01.990 DEBUG megazone23 makeCommonRows
## 2019 10 10 11:20:02.045 DEBUG megazone23 dim(mData1)=7785 1215
## 2019 10 10 11:20:02.045 DEBUG megazone23 dim(mData2)=7785 600
## 2019 10 10 11:20:02.046 DEBUG megazone23 makeCommonCols
## 2019 10 10 11:20:02.072 DEBUG megazone23 dim(mData1)=7785 586
## 2019 10 10 11:20:02.072 DEBUG megazone23 dim(mData2)=7785 586
## 2019 10 10 11:20:02.072 DEBUG megazone23 asSameOrder
## 2019 10 10 11:20:02.092 DEBUG megazone23 asSameOrder before
## 2019 10 10 11:20:02.092 DEBUG megazone23 asSameOrder after if 1
## 2019 10 10 11:20:02.112 DEBUG megazone23 asSameOrder after if 2
## 2019 10 10 11:20:02.113 DEBUG megazone23 row.names(mat1)
## 2019 10 10 11:20:02.113 DEBUG megazone23 rownames(mat1)
## 2019 10 10 11:20:02.113 DEBUG megazone23 row.names(mat2)
## 2019 10 10 11:20:02.114 DEBUG megazone23 rownames(mat2)
## 2019 10 10 11:20:02.114 DEBUG megazone23 m.i
## 2019 10 10 11:20:02.133 DEBUG megazone23 asSameOrder after if 3
## 2019 10 10 11:20:02.133 DEBUG megazone23 dim(ebObj@mat1Com)=7785 586
## 2019 10 10 11:20:02.134 DEBUG megazone23 dim(ebObj@mat2Com)=7785 586
## 2019 10 10 11:20:02.134 DEBUG megazone23 get Validation and Training sets from vectors
## 2019 10 10 11:20:02.157 DEBUG megazone23 dim(ebObj@mat1Validation)=10000 175
## 2019 10 10 11:20:02.157 DEBUG megazone23 dim(ebObj@mat2Validation)=10000 175
## 2019 10 10 11:20:02.157 DEBUG megazone23 dim(ebObj@mat1Train)=10000 288
## 2019 10 10 11:20:02.158 DEBUG megazone23 dim(ebObj@mat2Train)=10000 288
## 2019 10 10 11:20:02.158 DEBUG megazone23 before Test Set
## 2019 10 10 11:20:02.187 DEBUG megazone23 after Test Set
```

```

## 2019 10 10 11:20:02.288 DEBUG megazone23 train
## 2019 10 10 11:20:02.308 DEBUG megazone23 inside train
## 2019 10 10 11:20:02.308 DEBUG megazone23 train theEBNP_PriorPlotsFile= /bea_testing/output/EBNPlus_T
## 2019 10 10 11:20:02.315 DEBUG megazone23 colnames(Object@mat1Train)=TCGA-A1-A0SJ-01A-11R-A084-07.RNA
## 2019 10 10 11:20:02.315 DEBUG megazone23 colnames(Object@mat2Train)=TCGA-A1-A0SJ-01A-11R-A084-07.Agi
## 2019 10 10 11:20:02.316 DEBUG megazone23 inside train, call getData4EB
## 2019 10 10 11:20:02.499 DEBUG megazone23 data4EB <- as.matrix
## 2019 10 10 11:20:02.499 DEBUG megazone23 dim(data4EB)=7785 404
## 2019 10 10 11:20:02.499 DEBUG megazone23 cbinds and rbinds
## 2019 10 10 11:20:02.499 DEBUG megazone23 Object@DF1batch= RNASeqV2
## 2019 10 10 11:20:02.500 DEBUG megazone23 Object@DF2batch= Agilent4502
## 2019 10 10 11:20:02.501 DEBUG megazone23 row.names(sampBatch)
## 2019 10 10 11:20:02.501 DEBUG megazone23 TCGA-A1-A0SJ-01A-11R-A084-07.RNASeqV2 TCGA-A2-A04N-01A-11R-
## 2019 10 10 11:20:02.501 DEBUG megazone23 colnames(sampBatch)
## 2019 10 10 11:20:02.502 DEBUG megazone23 sample batch
## 2019 10 10 11:20:02.502 DEBUG megazone23 inside train, after getData4EB
## 2019 10 10 11:20:02.502 DEBUG megazone23 call EB MBatch
## 2019 10 10 11:20:02.575 DEBUG megazone23 TDC HERE trainEB mbatch
## 2019 10 10 11:20:02.576 DEBUG megazone23 EBNplus args Obj
## 2019 10 10 11:20:02.576 DEBUG megazone23 EBNplus args par.prior
## 2019 10 10 11:20:02.576 DEBUG megazone23 EBNplus args minSampleNum
## 2019 10 10 11:20:02.576 DEBUG megazone23 EBNplus theEBNP_PriorPlotsFile= /bea_testing/output/EBNPlus
## 2019 10 10 11:20:02.576 DEBUG megazone23 dat
## [1] 7785 404
## 2019 10 10 11:20:02.577 DEBUG megazone23 saminfo
## [1] 404 2
## 2019 10 10 11:20:02.577 DEBUG megazone23 check column and row names
## 2019 10 10 11:20:02.839 DEBUG megazone23 in design.mat_plus
## 2019 10 10 11:20:02.840 DEBUG megazone23 in build.design_plus
## 2019 10 10 11:20:02.840 DEBUG megazone23 after build.design_plus loop
## 2019 10 10 11:20:02.841 DEBUG megazone23 design
## 2019 10 10 11:20:02.841 DEBUG megazone23 list.batch_plus(saminfo)
## 2019 10 10 11:20:02.841 DEBUG megazone23 in list.batch_plus
## 2019 10 10 11:20:02.842 DEBUG megazone23 list.batch_plus tmp1
## 2019 10 10 11:20:02.842 DEBUG megazone23 list.batch_plus uniTmp[i]
## 2019 10 10 11:20:02.842 DEBUG megazone23 list.batch_plus uniTmp[i]
## 2019 10 10 11:20:02.842 DEBUG megazone23 list.batch_plus batches
## 2019 10 10 11:20:02.843 DEBUG megazone23 after list.batch_plus(saminfo)
## 2019 10 10 11:20:02.844 DEBUG megazone23 n.batches
## 2019 10 10 11:20:02.844 DEBUG megazone23 n.array
## 2019 10 10 11:20:05.029 DEBUG megazone23 missbatch matrix
## 2019 10 10 11:20:05.041 DEBUG megazone23 There are 63 genes that were removed because of whole bat
## 2019 10 10 11:20:05.526 DEBUG megazone23 B.hat
## 2019 10 10 11:20:05.527 DEBUG megazone23 grand.mean
## 2019 10 10 11:20:05.766 DEBUG megazone23 var.pooled
## 2019 10 10 11:20:05.771 DEBUG megazone23 stand.mean 1
## 2019 10 10 11:20:05.817 DEBUG megazone23 stand.mean 2
## 2019 10 10 11:20:05.846 DEBUG megazone23 s.data
## 2019 10 10 11:20:06.348 DEBUG megazone23 gamma.hat
## 2019 10 10 11:20:06.665 DEBUG megazone23 plot priors
## 2019 10 10 11:20:06.666 DEBUG megazone23 plotPrior priorPlotsFile= /bea_testing/output/EBNPlus_Train
## 2019 10 10 11:20:07.128 DEBUG megazone23 parametric adjustments
## 2019 10 10 11:20:07.463 DEBUG megazone23 parametric adjustments

```

```

## 2019 10 10 11:20:07.640 DEBUG megazone23 in getBayesData
## 2019 10 10 11:20:07.641 DEBUG megazone23 dim(s.data)=7722 404
## 2019 10 10 11:20:07.641 DEBUG megazone23 dim(stand.mean)=7722 404
## 2019 10 10 11:20:07.908 DEBUG megazone23 dim(bayesdata)=7722 404
## 2019 10 10 11:20:07.922 DEBUG megazone23 Second dim(bayesdata)=7722 404
## 2019 10 10 11:20:07.922 DEBUG megazone23 keep the dim of Original.dat, keep the gene with all NA
## 2019 10 10 11:20:07.932 DEBUG megazone23 resultsDat
## 2019 10 10 11:20:07.933 DEBUG megazone23 after EB MBatch
## 2019 10 10 11:20:07.933 DEBUG megazone23 inside TEST branch MBatch
## 2019 10 10 11:20:07.937 DEBUG megazone23 test
## 2019 10 10 11:20:07.937 DEBUG megazone23 dim(mat1)=10000 86
## 2019 10 10 11:20:07.938 DEBUG megazone23 dim(mat1)=
## 2019 10 10 11:20:07.938 DEBUG megazone23 dim(mat2)=10000 86
## 2019 10 10 11:20:07.938 DEBUG megazone23 dim(mat2)=
## 2019 10 10 11:20:07.995 DEBUG megazone23 dim(dat)=7785 172
## 2019 10 10 11:20:07.995 DEBUG megazone23 dim(dat)=
## 2019 10 10 11:20:08.001 DEBUG megazone23 after parlist
## 2019 10 10 11:20:08.001 DEBUG megazone23 dim(dat)=7785 172
## 2019 10 10 11:20:08.002 DEBUG megazone23 valid.genes
## 2019 10 10 11:20:08.003 DEBUG megazone23 nonValidGenes
## 2019 10 10 11:20:08.006 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2019 10 10 11:20:08.006 DEBUG megazone23 after Non-valid
## 2019 10 10 11:20:08.007 DEBUG megazone23 dim(dat)=7722 172
## 2019 10 10 11:20:08.023 DEBUG megazone23 after getStandData
## 2019 10 10 11:20:08.023 DEBUG megazone23 dim(stand.data)=7722 172
## 2019 10 10 11:20:08.024 DEBUG megazone23 batches
## 2019 10 10 11:20:08.024 DEBUG megazone23 after batch.design
## 2019 10 10 11:20:08.024 DEBUG megazone23 dim(batch.design)=172 2
## 2019 10 10 11:20:08.024 DEBUG megazone23 n.batches
## 2019 10 10 11:20:08.024 DEBUG megazone23 in getBayesData
## 2019 10 10 11:20:08.025 DEBUG megazone23 dim(s.data)=7722 172
## 2019 10 10 11:20:08.025 DEBUG megazone23 dim(stand.mean)=7722 172
## 2019 10 10 11:20:08.046 DEBUG megazone23 dim(bayesdata)=7722 172
## 2019 10 10 11:20:08.052 DEBUG megazone23 Second dim(bayesdata)=7722 172
## 2019 10 10 11:20:08.053 DEBUG megazone23 after bayesData
## 2019 10 10 11:20:08.053 DEBUG megazone23 dim(bayesData)=7722 172
## 2019 10 10 11:20:08.053 DEBUG megazone23 EBadj datNonValid
## 2019 10 10 11:20:08.058 DEBUG megazone23 EBadj EBadj after non valid
## 2019 10 10 11:20:08.058 DEBUG megazone23 Check the adjusted test set
## 2019 10 10 11:20:08.059 DEBUG megazone23 after train
## 2019 10 10 11:20:08.059 DEBUG megazone23 EBadj
## 2019 10 10 11:20:08.059 DEBUG megazone23 EBadj train
## 2019 10 10 11:20:08.059 DEBUG megazone23 train
## 2019 10 10 11:20:08.059 DEBUG megazone23 dim(mat1)=10000 202
## 2019 10 10 11:20:08.060 DEBUG megazone23 dim(mat1)=
## 2019 10 10 11:20:08.060 DEBUG megazone23 dim(mat2)=10000 202
## 2019 10 10 11:20:08.060 DEBUG megazone23 dim(mat2)=
## 2019 10 10 11:20:08.174 DEBUG megazone23 dim(dat)=7785 404
## 2019 10 10 11:20:08.174 DEBUG megazone23 dim(dat)=
## 2019 10 10 11:20:08.186 DEBUG megazone23 after parlist
## 2019 10 10 11:20:08.186 DEBUG megazone23 dim(dat)=7785 404
## 2019 10 10 11:20:08.186 DEBUG megazone23 valid.genes
## 2019 10 10 11:20:08.187 DEBUG megazone23 nonValidGenes
## 2019 10 10 11:20:08.194 DEBUG megazone23 Non-valid genes were removed before adjustment

```



```

## 2019 10 10 11:20:08.194 DEBUG megazone23 after Non-valid
## 2019 10 10 11:20:08.194 DEBUG megazone23 dim(dat)=7722 404
## 2019 10 10 11:20:08.260 DEBUG megazone23 after getStandData
## 2019 10 10 11:20:08.261 DEBUG megazone23 dim(stand.data)=7722 404
## 2019 10 10 11:20:08.261 DEBUG megazone23 batches
## 2019 10 10 11:20:08.261 DEBUG megazone23 after batch.design
## 2019 10 10 11:20:08.262 DEBUG megazone23 dim(batch.design)=404 2
## 2019 10 10 11:20:08.262 DEBUG megazone23 n.batches
## 2019 10 10 11:20:08.262 DEBUG megazone23 in getBayesData
## 2019 10 10 11:20:08.262 DEBUG megazone23 dim(s.data)=7722 404
## 2019 10 10 11:20:08.262 DEBUG megazone23 dim(stand.mean)=7722 404
## 2019 10 10 11:20:08.343 DEBUG megazone23 dim(bayesdata)=7722 404
## 2019 10 10 11:20:08.374 DEBUG megazone23 Second dim(bayesdata)=7722 404
## 2019 10 10 11:20:08.374 DEBUG megazone23 after bayesData
## 2019 10 10 11:20:08.375 DEBUG megazone23 dim(bayesData)=7722 404
## 2019 10 10 11:20:08.375 DEBUG megazone23 EBadj datNonValid
## 2019 10 10 11:20:08.391 DEBUG megazone23 EBadj EBadj after non valid
## 2019 10 10 11:20:08.391 DEBUG megazone23 objafterEB@trainEB
## 2019 10 10 11:20:08.392 DEBUG megazone23 EBadj validation
## 2019 10 10 11:20:08.392 DEBUG megazone23 validation
## 2019 10 10 11:20:08.392 DEBUG megazone23 dim(mat1)=10000 175
## 2019 10 10 11:20:08.393 DEBUG megazone23 dim(mat1)=
## 2019 10 10 11:20:08.393 DEBUG megazone23 dim(mat2)=10000 175
## 2019 10 10 11:20:08.393 DEBUG megazone23 dim(mat2)=
## 2019 10 10 11:20:08.514 DEBUG megazone23 dim(dat)=7785 350
## 2019 10 10 11:20:08.515 DEBUG megazone23 dim(dat)=
## 2019 10 10 11:20:08.525 DEBUG megazone23 after parlist
## 2019 10 10 11:20:08.525 DEBUG megazone23 dim(dat)=7785 350
## 2019 10 10 11:20:08.525 DEBUG megazone23 valid.genes
## 2019 10 10 11:20:08.526 DEBUG megazone23 nonValidGenes
## 2019 10 10 11:20:08.532 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2019 10 10 11:20:08.532 DEBUG megazone23 after Non-valid
## 2019 10 10 11:20:08.533 DEBUG megazone23 dim(dat)=7722 350
## 2019 10 10 11:20:08.575 DEBUG megazone23 after getStandData
## 2019 10 10 11:20:08.576 DEBUG megazone23 dim(stand.data)=7722 350
## 2019 10 10 11:20:08.576 DEBUG megazone23 batches
## 2019 10 10 11:20:08.576 DEBUG megazone23 after batch.design
## 2019 10 10 11:20:08.576 DEBUG megazone23 dim(batch.design)=350 2
## 2019 10 10 11:20:08.577 DEBUG megazone23 n.batches
## 2019 10 10 11:20:08.577 DEBUG megazone23 in getBayesData
## 2019 10 10 11:20:08.577 DEBUG megazone23 dim(s.data)=7722 350
## 2019 10 10 11:20:08.577 DEBUG megazone23 dim(stand.mean)=7722 350
## 2019 10 10 11:20:08.635 DEBUG megazone23 dim(bayesdata)=7722 350
## 2019 10 10 11:20:08.658 DEBUG megazone23 Second dim(bayesdata)=7722 350
## 2019 10 10 11:20:08.659 DEBUG megazone23 after bayesData
## 2019 10 10 11:20:08.659 DEBUG megazone23 dim(bayesData)=7722 350
## 2019 10 10 11:20:08.659 DEBUG megazone23 EBadj datNonValid
## 2019 10 10 11:20:08.677 DEBUG megazone23 EBadj EBadj after non valid
## 2019 10 10 11:20:08.677 DEBUG megazone23 objafterEB@validationEB
## 2019 10 10 11:20:08.678 DEBUG megazone23 EBadj train for test results
## 2019 10 10 11:20:08.678 DEBUG megazone23 test
## 2019 10 10 11:20:08.678 DEBUG megazone23 dim(mat1)=10000 86
## 2019 10 10 11:20:08.678 DEBUG megazone23 dim(mat1)=
## 2019 10 10 11:20:08.679 DEBUG megazone23 dim(mat2)=10000 86

```

```

## 2019 10 10 11:20:08.679 DEBUG megazone23 dim(mat2)=
## 2019 10 10 11:20:08.744 DEBUG megazone23 dim(dat)=7785 172
## 2019 10 10 11:20:08.745 DEBUG megazone23 dim(dat)=
## 2019 10 10 11:20:08.751 DEBUG megazone23 after parlist
## 2019 10 10 11:20:08.751 DEBUG megazone23 dim(dat)=7785 172
## 2019 10 10 11:20:08.752 DEBUG megazone23 valid.genes
## 2019 10 10 11:20:08.753 DEBUG megazone23 nonValidGenes
## 2019 10 10 11:20:08.756 DEBUG megazone23 Non-valid genes were removed before adjustment
## 2019 10 10 11:20:08.756 DEBUG megazone23 after Non-valid
## 2019 10 10 11:20:08.757 DEBUG megazone23 dim(dat)=7722 172
## 2019 10 10 11:20:08.774 DEBUG megazone23 after getStandData
## 2019 10 10 11:20:08.774 DEBUG megazone23 dim(stand.data)=7722 172
## 2019 10 10 11:20:08.774 DEBUG megazone23 batches
## 2019 10 10 11:20:08.775 DEBUG megazone23 after batch.design
## 2019 10 10 11:20:08.775 DEBUG megazone23 dim(batch.design)=172 2
## 2019 10 10 11:20:08.775 DEBUG megazone23 n.batches
## 2019 10 10 11:20:08.775 DEBUG megazone23 in getBayesData
## 2019 10 10 11:20:08.775 DEBUG megazone23 dim(s.data)=7722 172
## 2019 10 10 11:20:08.776 DEBUG megazone23 dim(stand.mean)=7722 172
## 2019 10 10 11:20:08.797 DEBUG megazone23 dim(bayesdata)=7722 172
## 2019 10 10 11:20:08.803 DEBUG megazone23 Second dim(bayesdata)=7722 172
## 2019 10 10 11:20:08.804 DEBUG megazone23 after bayesData
## 2019 10 10 11:20:08.804 DEBUG megazone23 dim(bayesData)=7722 172
## 2019 10 10 11:20:08.804 DEBUG megazone23 EBadj datNonValid
## 2019 10 10 11:20:08.809 DEBUG megazone23 EBadj EBadj after non valid
## 2019 10 10 11:20:08.809 DEBUG megazone23 testResults@testEB
## [1] "TestSet1"
## [1] TRUE
## [1] 10000      86
##          TCGA-A1-AOSH-01A-11R-A084-07 TCGA-A2-A04X-01A-21R-A034-07
## A1BG                      7.4960                      5.431
## A1CF                      -2.0160                      -2.016
## A2BP1                     -0.7169                      -2.016
## A2LD1                      7.0290                      5.904
##          TCGA-A2-AOCU-01A-12R-A034-07 TCGA-A2-AOCX-01A-21R-A00Z-07
## A1BG                      6.707                      6.4130
## A1CF                      -2.016                      -0.6197
## A2BP1                     -2.016                      -2.0160
## A2LD1                      5.635                      5.9360
## [1] "TestSet2"
## [1] TRUE
## [1] 10000      86
##          TCGA-A1-AOSH-01A-11R-A084-07 TCGA-A2-A04X-01A-21R-A034-07
## 15E1.2                   -1.9440000                   -0.6882500
## 2'-PDE                    -0.0295625                    0.3431250
## 7A5                      1.8550000                    2.1730000
## A1BG                      0.7486667                   -0.6243333
##          TCGA-A2-AOCU-01A-12R-A034-07 TCGA-A2-AOCX-01A-21R-A00Z-07
## 15E1.2                   -1.42200                   -0.36475000
## 2'-PDE                    0.43875                    0.21425000
## 7A5                      -0.96550                    0.34000000
## A1BG                      0.28800                   -0.02466667
## [1] "TrainingSet1"
## [1] TRUE

```

```

## [1] 10000 202
## TCGA-A1-AOSJ-01A-11R-A084-07 TCGA-A2-A04N-01A-11R-A115-07
## A1BG 7.820 7.521
## A1CF -2.016 -2.016
## A2BP1 -2.016 -2.016
## A2LD1 7.035 6.995
## TCGA-A2-A04R-01A-41R-A109-07 TCGA-A2-A04T-01A-21R-A034-07
## A1BG 7.7130 6.03800
## A1CF -0.7145 -2.01600
## A2BP1 -2.0160 -0.08512
## A2LD1 6.0800 6.25700
## [1] "TrainingSet2"
## [1] TRUE
## [1] 10000 202
## TCGA-A1-AOSJ-01A-11R-A084-07 TCGA-A2-A04N-01A-11R-A115-07
## 15E1.2 -2.0980000 -0.94900000
## 2'-PDE 0.7919375 -0.11818750
## 7A5 1.4080000 2.64550000
## A1BG 0.7500000 0.08566667
## TCGA-A2-A04R-01A-41R-A109-07 TCGA-A2-A04T-01A-21R-A034-07
## 15E1.2 -1.652250 -1.2670000
## 2'-PDE 0.635625 0.2290625
## 7A5 -0.392000 1.2170000
## A1BG 1.278000 -0.0885000
## [1] "TrainingResults"
## [1] TRUE
## [1] 7785 404
## TCGA-A1-AOSJ-01A-11R-A084-07.Agilent4502
## A1BG 4.3247402
## A2BP1 -0.3851765
## A2M 8.0978466
## A2ML1 1.8998344
## TCGA-A1-AOSJ-01A-11R-A084-07.RNASeqV2
## A1BG 4.2896562
## A2BP1 -2.1547612
## A2M 8.0183429
## A2ML1 -0.4772133
## TCGA-A2-A04N-01A-11R-A115-07.Agilent4502
## A1BG 3.4531507
## A2BP1 -0.2851838
## A2M 6.7163478
## A2ML1 1.9818843
## TCGA-A2-A04N-01A-11R-A115-07.RNASeqV2
## A1BG 4.0392535
## A2BP1 -2.1547612
## A2M 6.8664714
## A2ML1 -0.9104235
## [1] "ValidationSet1"
## [1] TRUE
## [1] 10000 175
## TCGA-A1-AOSD-01A-11R-A115-07 TCGA-A1-AOSK-01A-12R-A084-07
## A1BG 7.1552 6.440
## A1CF -2.0160 -2.016
## A2BP1 -0.7908 -2.016

```

```

## A2LD1                6.0680                8.659
##      TCGA-A1-AOSM-01A-11R-A084-07 TCGA-A2-A04P-01A-31R-A034-07
## A1BG                8.265                6.7681
## A1CF                -2.016                -2.0160
## A2BP1               -0.829                -2.0160
## A2LD1                6.686                6.4740
## [1] "ValidationSet2"
## [1] TRUE
## [1] 10000    175
##      TCGA-A1-AOSD-01A-11R-A115-07 TCGA-A1-AOSK-01A-12R-A084-07
## 15E1.2                -1.4477500                -1.2850000
## 2'-PDE                 0.0153750                -0.1911250
## 7A5                   2.4105000                -1.3005000
## A1BG                   0.9493333                -0.5471667
##      TCGA-A1-AOSM-01A-11R-A084-07 TCGA-A2-A04P-01A-31R-A034-07
## 15E1.2                -2.1540000                -1.5990000
## 2'-PDE                 0.3263125                -0.3471250
## 7A5                   2.1600000                1.7275000
## A1BG                   0.5976667                -0.1568333
## [1] "ValidationResults"
## [1] TRUE
## [1] 7785    350
##      TCGA-A1-AOSD-01A-11R-A115-07.Agilent4502
## A1BG                   4.586261
## A2BP1                  2.363897
## A2M                    6.632396
## A2ML1                  1.648391
##      TCGA-A1-AOSD-01A-11R-A115-07.RNASeqV2
## A1BG                   3.732908
## A2BP1                  -1.327814
## A2M                    8.122790
## A2ML1                  1.673291
##      TCGA-A1-AOSK-01A-12R-A084-07.Agilent4502
## A1BG                   2.622888
## A2BP1                  -1.583156
## A2M                    5.507855
## A2ML1                  1.337396
##      TCGA-A1-AOSK-01A-12R-A084-07.RNASeqV2
## A1BG                   3.133951
## A2BP1                  -2.154761
## A2M                    5.653508
## A2ML1                  2.874396
## [1] "CorrectedResults"
## [1] FALSE
## NULL
## NULL

```

7 Example Output

A list of matrices with the following names().

7.1 TestSet1

a matrix of the test set from data set 1

7.2 TestSet2

a matrix of the test set from data set 2

7.3 TrainingSet1

a matrix of the training set from data set 1

7.4 TrainingSet2

a matrix of the training set from data set 2

7.5 TrainingResults

a matrix of the corrected training data

7.6 ValidationSet1

a matrix of the validation set from data set 1

7.7 ValidationSet2

a matrix of the validation set from data set 2

7.8 ValidationResults

a matrix of the corrected validation data

7.9 CorrectedResults

If Validation Ratio is 0, the validation set values will be 0 x 0 matrices, but this list member will contain a matrix of corrected values, using all replicates.

This also generates the prior plots output file like this one.

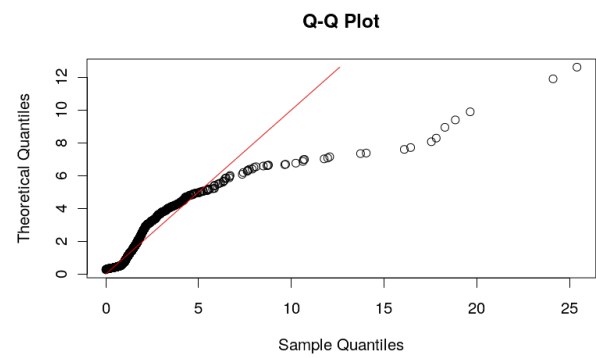
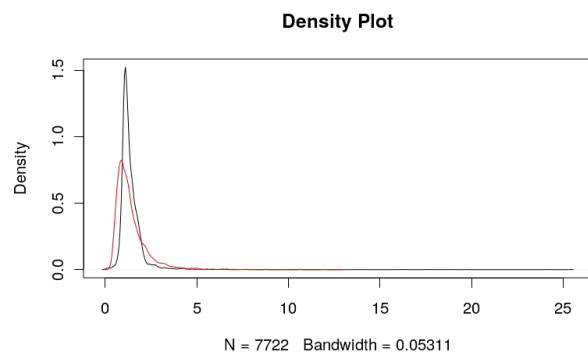
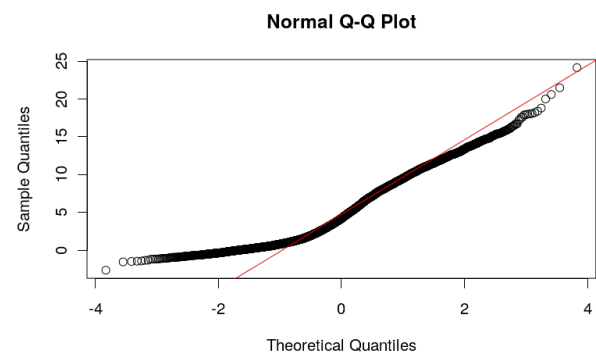
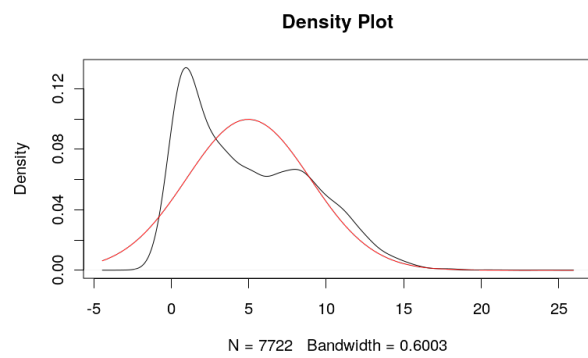


Figure 1: Prior Plots File