

1 Using Batch Effects Interface Assessments

This document focuses on explaining the components of the Batch Effects Interface (BEI) involved with creating a job, loading data, and running assessments. This document will not address statistical issues or "how to spot" batch effects.

The URL for your install should be provided to you, but will likely be something like:

<http://your-server.your-company.com:9999/BatchEffectsInterface/>

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 at <https://github.com/MD-Anderson-Bioinformatics/MBatch/tree/master/pdf> for instructions on downloading test data.

RBN corrections uses replicates between the two datasets to combine the two data sets based on replicates between sets.

2 Starting a Job

See BEstack_02A_BEIUsingAssessments for more details about starting a job.

Use the "Start New Job" button and select "User Uploaded Data" for Step 1.a. From within the MATRIX_DATA.zip archive, upload brca_rnaseq2_matrix_data.tsv as the data matrix and brca_rnaseq2_batches.tsv as the batch matrix.

THE UNIVERSITY OF TEXAS
MDAnderson Cancer Center
Making Cancer History®

Batch Effects Interface External 2018-04-24-1417

[Home](#) Job Id:1524669586504 Job State:NEWJOB_START Job Message:Waiting for new job data setup. [Edit Details](#)

Step 1.a.: Specify Primary Dataset
User Uploaded Data

Use Data Uploaded
Matrix File Upload: brca_rnaseq2_matrix_data.tsv

Use Data Uploaded
Matrix File Upload: No file selected. ☒ **Primary Matrix Uploaded**
Batch File Upload: brca_rnaseq2_batches.tsv

Then for Step 1.b. also select User Uploaded Data, and use brca_agi4502_matrix_data.tsv for the Matrix File and brca_agi4502_batches.tsv for the Batch File.

☒ **Step 1.a.: Specify Primary Dataset Complete**

Step 1.b.: Specify Secondary Dataset
Some correction algorithms use two datasets. Use the Proceed button if you do not need a second dataset.
User Uploaded Data

Use Data Uploaded
Matrix File Upload: brca_agi4502_matrix_data.tsv

Use Data Uploaded
Matrix File Upload: No file selected. ☒ **Secondary Matrix Uploaded**
Batch File Upload: brca_agi4502_batches.tsv

Then select Proceed to MBatch Configuration.

THE UNIVERSITY OF TEXAS
MDAndersonCancerCenter
Making Cancer History®

Batch Effects Interface External 2018-04-24-1417

[Home](#) **Job Id:**1524669586504 **Job State:**NEWJOB_SECONDARY_DONE **Job Message:**Secondary Data Available. [Edit Details](#)

☒ **Step 1.a.: Specify Primary Dataset Complete**

☒ **Step 1.b.: Specify Secondary Dataset Complete**
[Proceed to MBatch Configuration](#)

3 Configuring and Running Assessments

See BEMStack_02A_BEIUsingAssessmentsExternal for more details about Configuring Assessments.

Below, we have selected Sample as the Sample Identifier, and selected BatchId as well as ShipDate as the assessment batch types. For Step 3, we have kept the defaults.

Step 2: Select Batches

[Reset Defaults](#)

Select Sample Identifier (Select Sample Id Batch Type):

Selected Types for Batch Analysis: BatchId,ShipDate

	Batch Type to Check	Batch Type Name	Batch Summary (Batch Name and Count)
<input checked="" type="checkbox"/>		BatchId	00047:99 00056:92 00061:61 00072:50 00074:39 00080:27 00085:92 00093:61 00096:58 00103:24 00109:89 00117:46 00120:29 00124:35 00136:36 00142:49 00147:30 00155:10 00167:13 00177:17 00185:11 00202:10 00216:13 00227:18 00234:10 00239:15 00255:15 00271:22 00288:15 00296:21 00305:14 00322:12 00334:12 00338:16 00360:10 00372:10 00379:34
<input type="checkbox"/>		PlateId	A00Z:90 A034:91 A056:47 A084:46 A089:22 A109:39 A10J:27 A10U:2 A115:92 A12D:60 A12P:58 A137:23 A13Q:87 A144:46 A14D:29 A14M:35 A157:36 A169:49 A16F:30 A17B:10 A180:13 A18M:17 A19E:1 A19W:11 A213:12 A21T:13 A22K:18 A22U:10 A239:15 A24H:16 A266:21 A26B:1 A27Q:15 A28M:21 A29R:14 A31O:12 A32P:12 A32Y:2 A33J:16 A352:10 A36F:10 A41B:34 A466:2
<input checked="" type="checkbox"/>		ShipDate	2010-06-16:90 2010-07-14:91 2010-08-11:47 2010-09-29:68 2010-11-01:39 2010-12-14:27 2011-01-12:94 2011-02-08:60 2011-02-22:58 2011-03-08:23 2011-03-23:87 2011-04-06:46 2011-04-19:29 2011-04-26:35 2011-05-17:36 2011-05-31:49 2011-06-14:30 2011-07-12:10 2011-08-31:13 2011-10-04:17 2011-10-25:1 2011-11-30:11 2012-01-24:12 2012-03-28:13 2012-05-30:18 2012-06-27:10 2012-07-30:15 2012-11-07:16 2013-01-30:22 2013-03-25:15 2013-04-17:21 2013-05-08:14 2013-07-31:13 2013-08-25:13 2013-10-03:2

Step 3: Select Filtering Options

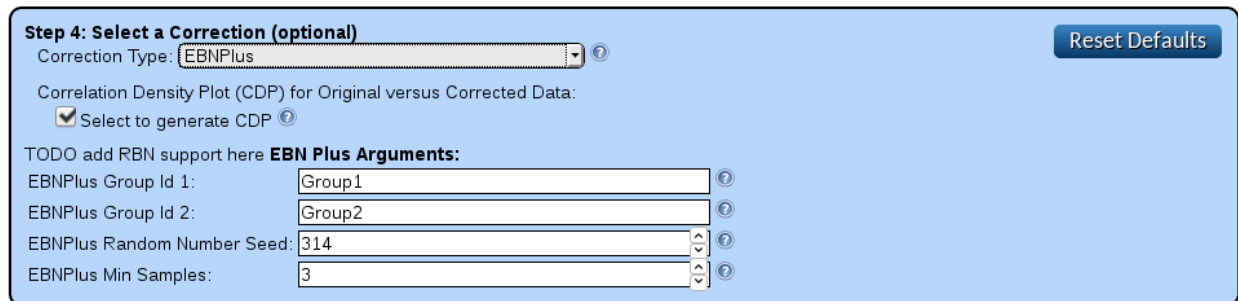
[Reset Defaults](#)

Auto-Filter to Maximum Number of Values:
Max Number of Values:

Exclude User Specified Batches:
Batch Type for Exclusion:

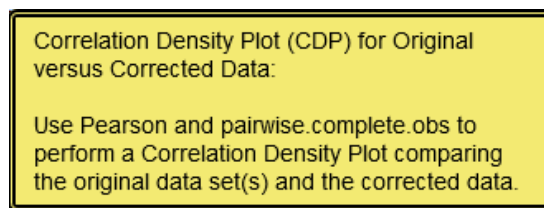
3.1 Step 4 (RBN with Replicates)

In Step 4, we begin by selecting RBN as the optional Correction Type.



The screenshot shows the 'Step 4: Select a Correction (optional)' panel. The 'Correction Type' dropdown is set to 'EBNPlus'. The 'Correlation Density Plot (CDP) for Original versus Corrected Data:' section has a checked checkbox for 'Select to generate CDP'. Below this, the 'EBN Plus Arguments:' section includes four input fields: 'EBNPlus Group Id 1' (Group1), 'EBNPlus Group Id 2' (Group2), 'EBNPlus Random Number Seed' (314), and 'EBNPlus Min Samples' (3). A 'Reset Defaults' button is in the top right corner.

Generating a Correlation Density Plot is selected automatically. (In a future version, this option will appear for any assessment, rather than just for certain correction options.)



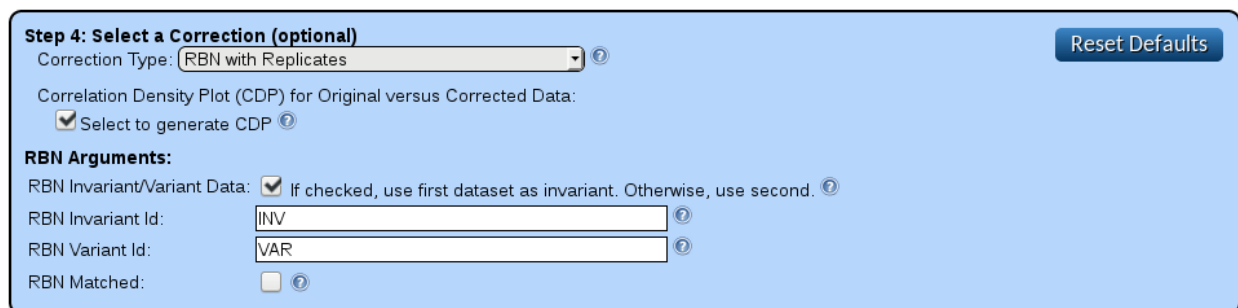
Correlation Density Plot (CDP) for Original versus Corrected Data:

Use Pearson and pairwise.complete.obs to perform a Correlation Density Plot comparing the original data set(s) and the corrected data.

Tooltip Text for Correlation Density Plot

The Group Ids must be alphanumeric values without spaces.

Here, we have kept all the default settings. Invariant and Variant ids are alphanumeric values without spaces but with underscores allowed.



The screenshot shows the 'Step 4: Select a Correction (optional)' panel with 'RBN with Replicates' selected in the 'Correction Type' dropdown. The 'Correlation Density Plot (CDP) for Original versus Corrected Data:' section has a checked checkbox for 'Select to generate CDP'. The 'RBN Arguments:' section includes: 'RBN Invariant/Variant Data:' with a checked checkbox and a note 'If checked, use first dataset as invariant. Otherwise, use second.'; 'RBN Invariant Id:' (INV); 'RBN Variant Id:' (VAR); and 'RBN Matched:' with an unchecked checkbox. A 'Reset Defaults' button is in the top right corner.

We do not use these for this run, instead we use the other RBN option.

3.1.1 Do MBatch Run

See BEstack_02A_BEIUsingAssessments for more details about running and monitoring a run. We press Do MBatch Run from the configuration page. After a bit, the run will fail.

THE UNIVERSITY OF TEXAS
MDAnderson Cancer Center
Making Cancer History®

Batch Effects Interface External 2018-04-25-1433

Home Job Id:1528136174339 Job State:MBATCHRUN_END_FAILURE Job Message:MBatch Run Failed Edit Details

MBatch Failed.
Download MBatch Results

Log File Tail (last 100 lines):

```
2018 06 05 18:03:38.624 DEBUG 6917b28273e6 readAsDataFrame - length(myCols) 6
2018 06 05 18:03:38.626 DEBUG 6917b28273e6 readAsDataFrame - length(myRows) 0
2018 06 05 18:03:38.628 DEBUG 6917b28273e6 readAsDataFrame - myCols Sample, Type, Batch
2018 06 05 18:03:38.630 DEBUG 6917b28273e6 readAsDataFrame - myRows
2018 06 05 18:03:38.707 DEBUG 6917b28273e6 readAsDataFrame - thePar -Xmx2000m
2018 06 05 18:03:38.709 DEBUG 6917b28273e6 readAsDataFrame - theFile /BEI/OUTPUT/152813
2018 06 05 18:03:38.711 DEBUG 6917b28273e6 readAsDataFrame - Calling .jinit /usr/local/lib/R/sit
2018 06 05 18:03:39.518 DEBUG 6917b28273e6 readAsDataFrame - .jinit complete
2018 06 05 18:03:39.520 DEBUG 6917b28273e6 readAsDataFrame before java
2018 06 05 18:03:39.548 DEBUG 6917b28273e6 readAsDataFrame after java
```

3.1.2 Failed Job

After the job has finished, use the Download MBatch Results button to get the results, which include a log file.

THE UNIVERSITY OF TEXAS
MDAnderson Cancer Center
Making Cancer History®

Batch Effects Interface External 2018-04-25-1433

Home Job Id:1528136174339 Job State:MBATCHRUN_END_FAILURE Job Message:MBatch Run Failed Edit Details

MBatch Failed.
Download MBatch Results

Log File Tail (last 100 lines):

```
2018 06 05 18:03:38.624 DEBUG 6917b28273e6 readAsDataFrame - length(myCols) 6
2018 06 05 18:03:38.626 DEBUG 6917b28273e6 readAsDataFrame - length(myRows) 0
2018 06 05 18:03:38.628 DEBUG 6917b28273e6 readAsDataFrame - myCols Sample, Type, Batch
2018 06 05 18:03:38.630 DEBUG 6917b28273e6 readAsDataFrame - myRows
2018 06 05 18:03:38.707 DEBUG 6917b28273e6 readAsDataFrame - thePar -Xmx2000m
2018 06 05 18:03:38.709 DEBUG 6917b28273e6 readAsDataFrame - theFile /BEI/OUTPUT/152813
2018 06 05 18:03:38.711 DEBUG 6917b28273e6 readAsDataFrame - Calling .jinit /usr/local/lib/R/sit
2018 06 05 18:03:39.518 DEBUG 6917b28273e6 readAsDataFrame - .jinit complete
2018 06 05 18:03:39.520 DEBUG 6917b28273e6 readAsDataFrame before java
2018 06 05 18:03:39.548 DEBUG 6917b28273e6 readAsDataFrame after java
```

Opening 1528136174339_Results.zip

You have chosen to open:
1528136174339_Results.zip
which is: Compressed (zipped) Folder
from: http://qcprpudev6.mdanderson.edu:9999

What should Firefox do with this file?

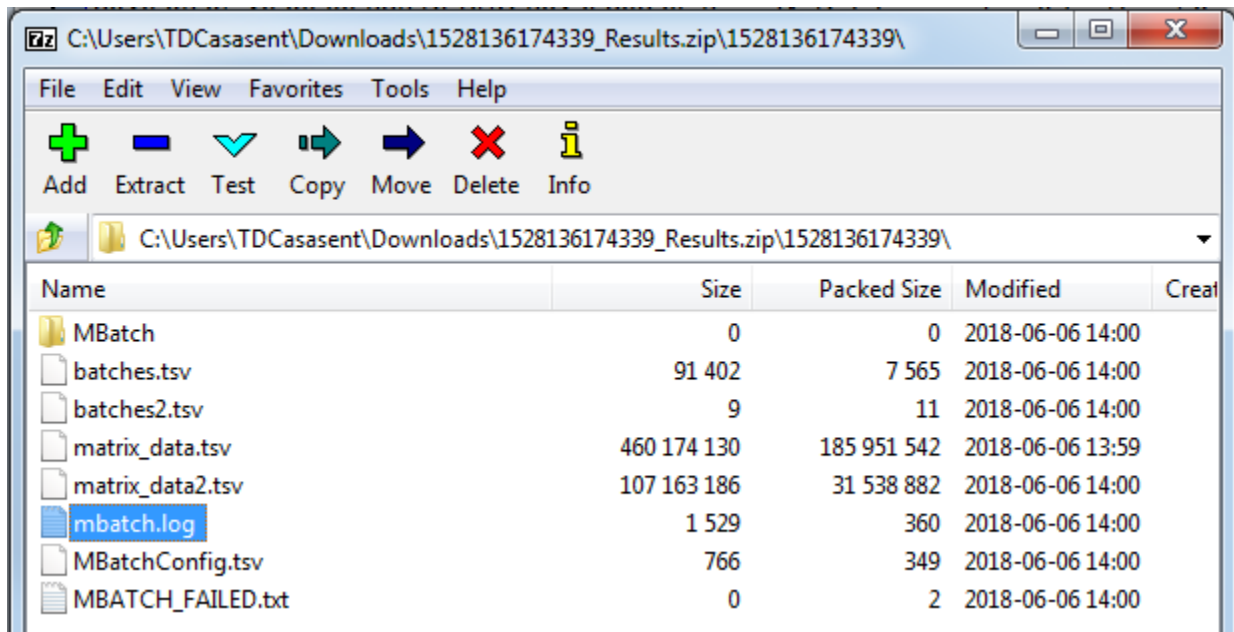
☐ Open with Windows Explorer (default)

☒ Save File

☐ Do this automatically for files like this from now on.

OK Cancel

Find the mbatch.log file inside the one directory (with the job id) in the ZIP file.



Extract and view the log file. It will look much like the tail from the web page.

```
2018 06 05 18:03:37.535 DEBUG 6917b28273e6 readAsDataFrame - thePar -Xmx2000m
2018 06 05 18:03:37.572 DEBUG 6917b28273e6 readAsDataFrame - theFile
/BEI/OUTPUT/1528136174339/batches.tsv
2018 06 05 18:03:37.572 DEBUG 6917b28273e6 readAsDataFrame - Calling .jinit
/usr/local/lib/R/site-library/MBatch/ReadRJava/ReadRJava.jar
2018 06 05 18:03:38.355 DEBUG 6917b28273e6 readAsDataFrame - .jinit complete
2018 06 05 18:03:38.357 DEBUG 6917b28273e6 readAsDataFrame before java
2018 06 05 18:03:38.616 DEBUG 6917b28273e6 readAsDataFrame after java
2018 06 05 18:03:38.622 DEBUG 6917b28273e6 readAsDataFrame - length(myData) 7290
2018 06 05 18:03:38.624 DEBUG 6917b28273e6 readAsDataFrame - length(myCols) 6
2018 06 05 18:03:38.626 DEBUG 6917b28273e6 readAsDataFrame - length(myRows) 0
2018 06 05 18:03:38.628 DEBUG 6917b28273e6 readAsDataFrame - myCols Sample, Type, BatchId,
PlateId, ShipDate, TSS
2018 06 05 18:03:38.630 DEBUG 6917b28273e6 readAsDataFrame - myRows
2018 06 05 18:03:38.707 DEBUG 6917b28273e6 readAsDataFrame - thePar -Xmx2000m
2018 06 05 18:03:38.709 DEBUG 6917b28273e6 readAsDataFrame - theFile
/BEI/OUTPUT/1528136174339/batches2.tsv
2018 06 05 18:03:38.711 DEBUG 6917b28273e6 readAsDataFrame - Calling .jinit
/usr/local/lib/R/site-library/MBatch/ReadRJava/ReadRJava.jar
2018 06 05 18:03:39.518 DEBUG 6917b28273e6 readAsDataFrame - .jinit complete
2018 06 05 18:03:39.520 DEBUG 6917b28273e6 readAsDataFrame before java
2018 06 05 18:03:39.548 DEBUG 6917b28273e6 readAsDataFrame after java
```

Notice that there is no error or any sort of exception in the log file. This generally means that the memory requirements for processing the data exceeded the memory available, and that R crashed. The system detected this crash, and marked the job as failed.

3.2 Step 4 (RBN with Pseudoreplicates)

Pseudoreplicates are used when you have replicates, or something you consider “close enough”, in the dataset, but they have different sample ids. When the Pseudoreplicates option comes up, it looks similar to the Replicates option, except for the lists from which to select Pseudoreplicates. We recommend at least 30 pseudoreplicates in each group.

Step 4: Select a Correction (optional)

Correction Type: RBN with Pseudoreplicates ?

Correlation Density Plot (CDP) for Original versus Corrected Data:
☒ Select to generate CDP ?

RBN Arguments:
RBN Invariant/Variant Data: ☒ If checked, use first dataset as invariant. Otherwise, use second. ?
RBN Invariant Id: ?
RBN Variant Id: ?
RBN Matched: ☐ ?

RBN Pseudoreplicants Arguments:
Select Multiple Invariant Pseudoreplicates from Sample

TCGA-3C-AAAU-01A-11R-A41B-07
TCGA-3C-AALI-01A-11R-A41B-07
TCGA-3C-AALJ-01A-31R-A41B-07
TCGA-3C-AALK-01A-11R-A41B-07
TCGA-4H-AAAK-01A-12R-A41B-07
TCGA-5L-AAT0-01A-12R-A41B-07
TCGA-5L-AAT1-01A-12R-A41B-07
TCGA-5T-A9QA-01A-11R-A41B-07
TCGA-A1-A0SB-01A-11R-A144-07
TCGA-A1-A0SD-01A-11R-A115-07
TCGA-A1-A0SE-01A-11R-A084-07
TCGA-A1-A0SF-01A-11R-A144-07
TCGA-A1-A0SG-01A-11R-A144-07
TCGA-A1-A0SH-01A-11R-A084-07
TCGA-A1-A0SI-01A-11R-A144-07
TCGA-A1-A0SJ-01A-11R-A084-07
TCGA-A1-A0SK-01A-12R-A084-07
TCGA-A1-A0SM-01A-11R-A084-07
TCGA-A1-A0SN-01A-11R-A144-07
TCGA-A1-A0SO-01A-22R-A084-07
TCGA-A1-A0SP-01A-11R-A084-07
TCGA-A1-A0SQ-01A-21R-A144-07
TCGA-A2-A04N-01A-11R-A115-07
TCGA-A2-A04P-01A-31R-A034-07
TCGA-A2-A04Q-01A-21R-A034-07
Please select at least one
RBN Invariant Pseudoreplicate

Select Multiple Variant Pseudoreplicates from Sample

TCGA-A1-A0SD-01A-11R-A115-07
TCGA-A1-A0SE-01A-11R-A084-07
TCGA-A1-A0SH-01A-11R-A084-07
TCGA-A1-A0SJ-01A-11R-A084-07
TCGA-A1-A0SK-01A-12R-A084-07
TCGA-A1-A0SM-01A-11R-A084-07
TCGA-A1-A0SO-01A-22R-A084-07
TCGA-A1-A0SP-01A-11R-A084-07
TCGA-A2-A04N-01A-11R-A115-07
TCGA-A2-A04P-01A-31R-A034-07
TCGA-A2-A04Q-01A-21R-A034-07
TCGA-A2-A04R-01A-41R-A109-07
TCGA-A2-A04T-01A-21R-A034-07
TCGA-A2-A04U-01A-11R-A115-07
TCGA-A2-A04V-01A-21R-A034-07
TCGA-A2-A04W-01A-31R-A115-07
TCGA-A2-A04X-01A-21R-A034-07
TCGA-A2-A04Y-01A-21R-A034-07
TCGA-A2-A0CL-01A-11R-A115-07
TCGA-A2-A0CM-01A-31R-A034-07
TCGA-A2-A0CP-01A-11R-A034-07
TCGA-A2-A0CQ-01A-21R-A034-07
TCGA-A2-A0CS-01A-11R-A115-07
TCGA-A2-A0CT-01A-31R-A056-07
TCGA-A2-A0CU-01A-12R-A034-07
Please select at least one RBN Variant Pseudoreplicate

7

Here, we have selected pseudoreplicates from each group.

Step 4: Select a Correction (optional)
Reset Defaults

Correction Type: RBN with Pseudoreplicates

Correlation Density Plot (CDP) for Original versus Corrected Data:
☒ Select to generate CDP

RBN Arguments:

RBN Invariant/Variant Data: ☒ If checked, use first dataset as invariant. Otherwise, use second.

RBN Invariant Id:

RBN Variant Id:

RBN Matched: ☐

RBN Pseudoreplicants Arguments:

Select Multiple Invariant Pseudoreplicates from Sample

Select Multiple Variant Pseudoreplicates from Sample

And we take the defaults for the assessments.

Step 5: Select PCA-Plus/DSC Arguments
Reset Defaults

\$root.selectedDSCPermutations=2000

Number of DSC Permutations:

Number of DSC Threads (1-5):

Minimum DSC Batch Size:

DSC Random Number Seed:

Maximum Number of Features:

Step 6: Select Boxplot Arguments
Reset Defaults

Max Number of Features:

Step 7: Select NGCHM Arguments
Reset Defaults

Perform NGCHM creation: ☒

Do MBatch Run
Save Configuration
Load Configuration
Reset Defaults 3-7

3.2.1 Do MBatch Run

See BEstack_02A_BEIUsingAssessments for more details about running and monitoring a run. We press Do MBatch Run from the configuration page.

THE UNIVERSITY OF TEXAS
MDAnderson Cancer Center
Making Cancer History®

Batch Effects Interface External 2018-04-25-1433

[Home](#) **Job Id:**1524687939743 **Job State:**MBATCHRUN_RUNNING_WAIT **Job Message:**MBatch Run in Progress [Edit Details](#)

The MBatch run is underway on a processing node.

Log File Tail (last 100 lines):

```
2018 04 25 20:50:44.485 DEBUG da7671a17f31 writeAsDataframe - thePar -Xmx2000m
2018 04 25 20:50:44.486 DEBUG da7671a17f31 writeAsDataframe - theFile /BEI/OUTPUT/1524687939743/MBatch/c
2018 04 25 20:50:44.487 DEBUG da7671a17f31 writeAsDataframe - length(myData) 12705
2018 04 25 20:50:44.487 DEBUG da7671a17f31 writeAsDataframe - length(myCols) 7
2018 04 25 20:50:44.488 DEBUG da7671a17f31 writeAsDataframe - length(myRows) 0
2018 04 25 20:50:44.489 DEBUG da7671a17f31 writeAsDataframe - Calling jinit /usr/local/lib/R/site-library/MBatch/R
2018 04 25 20:50:44.769 DEBUG da7671a17f31 writeAsDataframe - jinit complete
2018 04 25 20:50:44.770 DEBUG da7671a17f31 writeAsDataframe before java
2018 04 25 20:50:44.893 DEBUG da7671a17f31 writeAsDataframe after java
2018 04 25 20:50:44.895 DEBUG da7671a17f31 writeAsDataframe success= TRUE
```

The DSC Permutations step will take some time—30 minutes or more with the full run taking several hours.

THE UNIVERSITY OF TEXAS
MDAnderson Cancer Center
Making Cancer History®

Batch Effects Interface External 2018-04-24-1417

[Home](#) **Job Id:**1524669586504 **Job State:**MBATCHRUN_RUNNING_WAIT **Job Message:**MBatch Run in Progress [Edit Details](#)

The MBatch run is underway on a processing node.

Log File Tail (last 100 lines):

```
2018 04 25 15:44:37.058 DEBUG 1549a9e9f75d pvalueDSCwithExcerpt start
2018 04 25 15:44:37.059 DEBUG 1549a9e9f75d nrow(thePcaDataExcerpt)= 1815
2018 04 25 15:44:37.060 DEBUG 1549a9e9f75d ncol(thePcaDataExcerpt)= 1815
2018 04 25 15:44:37.061 DEBUG 1549a9e9f75d length(theBatchIdsForSamples)= 1815
2018 04 25 15:44:37.064 DEBUG 1549a9e9f75d getDSCwithExcerpt before java
2018 04 25 15:44:38.491 DEBUG 1549a9e9f75d getDSCwithExcerpt after java
2018 04 25 15:44:38.494 DEBUG 1549a9e9f75d pvalueDSCwithExcerpt after getDSCwithExcerpt
2018 04 25 15:44:38.496 DEBUG 1549a9e9f75d pvalueDSCwithExcerpt length(unique(theBatchIdsForSamples))= 38
2018 04 25 15:44:38.499 DEBUG 1549a9e9f75d pvalueDSCwithExcerpt call doDscPerms
2018 04 25 15:44:38.502 DEBUG 1549a9e9f75d doDscPerms before java
```

3.2.2 Finished Job

After the job has finished, use the Download option to get the corrected data.

THE UNIVERSITY OF TEXAS
MDAnderson Cancer Center
Making Cancer History®

Batch Effects Interface External 2018-04-25-1433

Home Job Id:1524687939743 Job State:MBATCHRUN_END_SUCCESS Job Message:MBatch Run Finished Successfully Edit Details

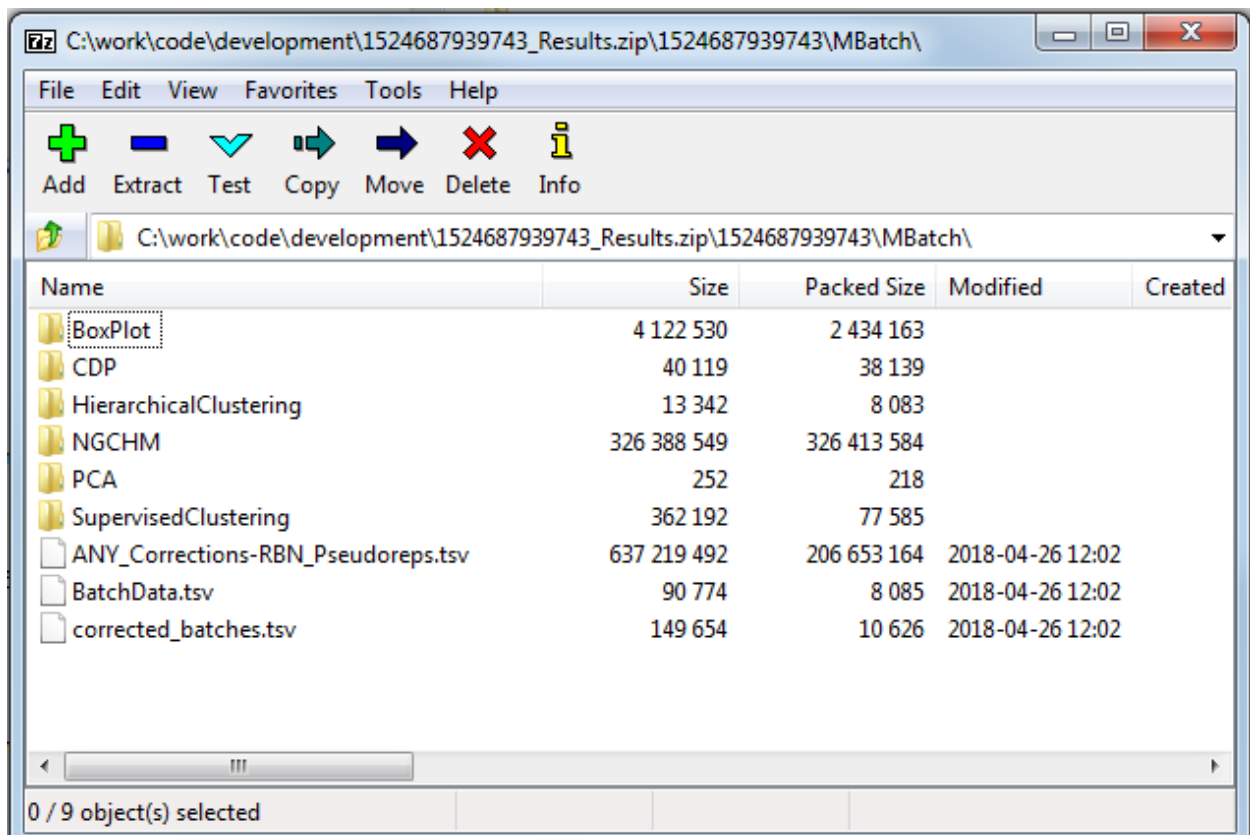
✓ MBatch succeeded.

Download MBatch Results view on the visualization website.

Log File Tail (last 100 lines):

```
2018 04 25 20:57:08.904 INFO da7671a17f31 CDP_Plot theData1UnmatchedReplicates= 1000
2018 04 25 20:57:08.905 INFO da7671a17f31 CDP_Plot theData2UnmatchedReplicates= 1000
2018 04 25 20:57:10.851 INFO da7671a17f31 CDP_Plot pairedCorr= 0
2018 04 25 20:57:10.851 INFO da7671a17f31 CDP_Plot unmatchedCorr= 1000
2018 04 25 20:57:10.853 INFO da7671a17f31 CDP_Plot pairedDensity$x= 0
2018 04 25 20:57:10.854 INFO da7671a17f31 CDP_Plot pairedDensity$y= 0
2018 04 25 20:57:10.855 INFO da7671a17f31 CDP_Plot pairedDensity$bw=
2018 04 25 20:57:10.855 INFO da7671a17f31 CDP_Plot unmatchedDensity$x= 512
2018 04 25 20:57:10.856 INFO da7671a17f31 CDP_Plot unmatchedDensity$y= 512
2018 04 25 20:57:10.856 INFO da7671a17f31 CDP_Plot unmatchedDensity$bw= 0.0323481284011279
```

Click the Download MBatch Results button. Open or unzip the archive and enter the MBatch directory.



The ANY_Corrections-RBN_Pseudoreps.tsv file contains the corrected data. Looking at an excerpt from that file below, you see the group ids have been added to the end of the sample ids (with a period to separate them).

	TCGA-Z7-A8R5-01A-42R-A41B-07-INV	TCGA-Z7-A8R6-01A-11R-A41B-07-INV	TCGA-A1-A0SD-01A-11R-A115-07-VAR	TCGA-A1-A0SE-01A-11R-A084-07-VAR
A1BG 1	8.780670006	7.957533334	NA	NA
A1CF 29974	-2.016249352	-2.016249352	NA	NA
A2BP1 54715	-0.243830672	-2.016249352	NA	NA
A2LD1 87769	6.349581128	4.668675122	NA	NA
A1BG	NA	NA	0.949333333	0.279833333
A2BP1	NA	NA	0.539166667	-1.534666667
A2M	NA	NA	0.242	1.0385
A2ML1	NA	NA	0.4235	0.3715

The corrected_batches.tsv contains the combined batch files. Looking at an excerpt from that file below, you see the group ids have been added to the end of the sample ids (with a period to separate them).

Sample	Type	BatchId	PlateId	ShipDate	TSS	EBNPlus
TCGA-XX-A899-01A-11R-A36F-07-INV	01	00372	A36F	2014-01-29	XX - Spectrum Health	INV
TCGA-XX-A89A-01A-11R-A36F-07-INV	01	00372	A36F	2014-01-29	XX - Spectrum Health	INV
TCGA-Z7-A8R5-01A-42R-A41B-07-INV	01	00379	A41B	2014-05-28	Z7 - John Wayne Cancer Center	INV
TCGA-Z7-A8R6-01A-11R-A41B-07-INV	01	00379	A41B	2014-05-28	Z7 - John Wayne Cancer Center	INV
TCGA-A1-A0SD-01A-11R-A115-07-VAR	01	00085	A115	2011-01-12	A1 - UCSF	VAR
TCGA-A1-A0SE-01A-11R-A084-07-VAR	01	00072	A084	2010-09-29	A1 - UCSF	VAR
TCGA-A1-A0SH-01A-11R-A084-07-VAR	01	00072	A084	2010-09-29	A1 - UCSF	VAR
TCGA-A1-A0SJ-01A-11R-A084-07-VAR	01	00072	A084	2010-09-29	A1 - UCSF	VAR