DSC Calculation Output
Tod Casasent
2022-03-01-0920

DSC calculations

Note that the DSC calculations use the Java 8 and rJava requirements from the install. The DSC can be calculated as part of creating the PCA+ diagram/files usually by calling PCA_Regular_Structures.

Documentation on using the PCA+ function (PCA_Regular_Structures) is available here: https://github.com/MD-Anderson-Bioinformatics/BatchEffectsPackage/blob/master/docs/MBatch/MBatch_04-02_PCA_Regular_Structures.pdf

Documentation is aimed at using MBatch with our Batch Effects Viewer, so some of the information described below is not found in other documentation.

Generally, the output is aimed at providing DSC data for the dynamic visualizations—we output enough data to generate dynamic versions of the output, as seen here: https://bioinformatics.mdanderson.org/MQA/ It may be useful to utilize either Batch Effects Interface (which incorporates the Batch Effects Viewer) to look at output or to use the ZIP generation functions available in the R packages, which lets you use a stand-alone version of the Batch Effects Viewer to look at the dynamic visualizations.

DSCOverview.tsv file

Summary output for the DSC values is written to a DSCOverview.tsv file.

The DSCOverview.tsv file is a tab delimited file. The first column (dataset) describes the dataset/batch being evaluated. The Overall-DSC gives the DSC for the dataset/batch overall, with Overall-DSC-pvalue giving the PValue. Other columns generate similar values for the first four principal components.

Dataset	Overall-DSC	Ov
/BEI/OUTPUT/1585774011149/MBatch/1585774011149/PCA/ShipDate/ManyToMany	0	<0
/BEI/OUTPUT/1585774011149/MBatch/1585774011149/PCA/TSS/ManyToMany	0	<0

PCAAnnotations.tsv file

Additional output for DSC is written to the PCA+ output directory (such as, within the output dir passed to PCA_Regular_Structures, PCA/ShipDate/ManyToMany/PCAAnnotations.tsv).

This file is also a TSV, and the output is reasonably self-evident, giving overall DSC, Dw (within groups), Db (between groups), and the same calculations for principal component pairs.

Type SubType Annotation Value

Run - MBatch Version 1.6.0

Run - DSC Permutations 2000

Diagram - Disp. Sep. Crit. (DSC) 0

Diagram - Disp. within groups (Dw) 31.6580769770181

Diagram - Disp. between groups (Db) 0

Diagram - DSC pvalue <0.0005

Component PC1,PC2 Disp. Sep. Crit. (DSC) (1,2) 0

Component PC1,PC2 Disp. within groups (Dw) (1,2) 19.7082621306403

Component PC1,PC2 Disp. between groups (Db) (1,2) 0

Component PC1,PC2 DSC pvalue(1,2) <0.0005

Component PC1,PC3 Disp. Sep. Crit. (DSC) (1,3) 0

Component PC1,PC3 Disp. within groups (Dw) (1,3) 19.0787154683988

Component PC1,PC3 Disp. between groups (Db) (1,3) 0

Component PC1,PC3 DSC pvalue(1,3) <0.0005

Component PC1,PC4 Disp. Sep. Crit. (DSC) (1,4) 0

Component PC1,PC4 Disp. within groups (Dw) (1,4) 18.7506700741928

Component PC1,PC4 Disp. between groups (Db) (1,4) 0

Component PC1,PC4 DSC pvalue(1,4) <0.0005

Component PC2,PC3 Disp. Sep. Crit. (DSC) (2,3) 0

Component PC2,PC3 Disp. within groups (Dw) (2,3) 10.1945954488947

Component PC2,PC3 Disp. between groups (Db) (2,3) 0

Component PC2,PC3 DSC pvalue(2,3) <0.0005

Component PC2,PC4 Disp. Sep. Crit. (DSC) (2,4) 0

Component PC2,PC4 Disp. within groups (Dw) (2,4) 9.56660967499638

Component PC2,PC4 Disp. between groups (Db) (2,4) 0

Component PC2,PC4 DSC pvalue(2,4) <0.0005

Component PC3,PC4 Disp. Sep. Crit. (DSC) (3,4) 0

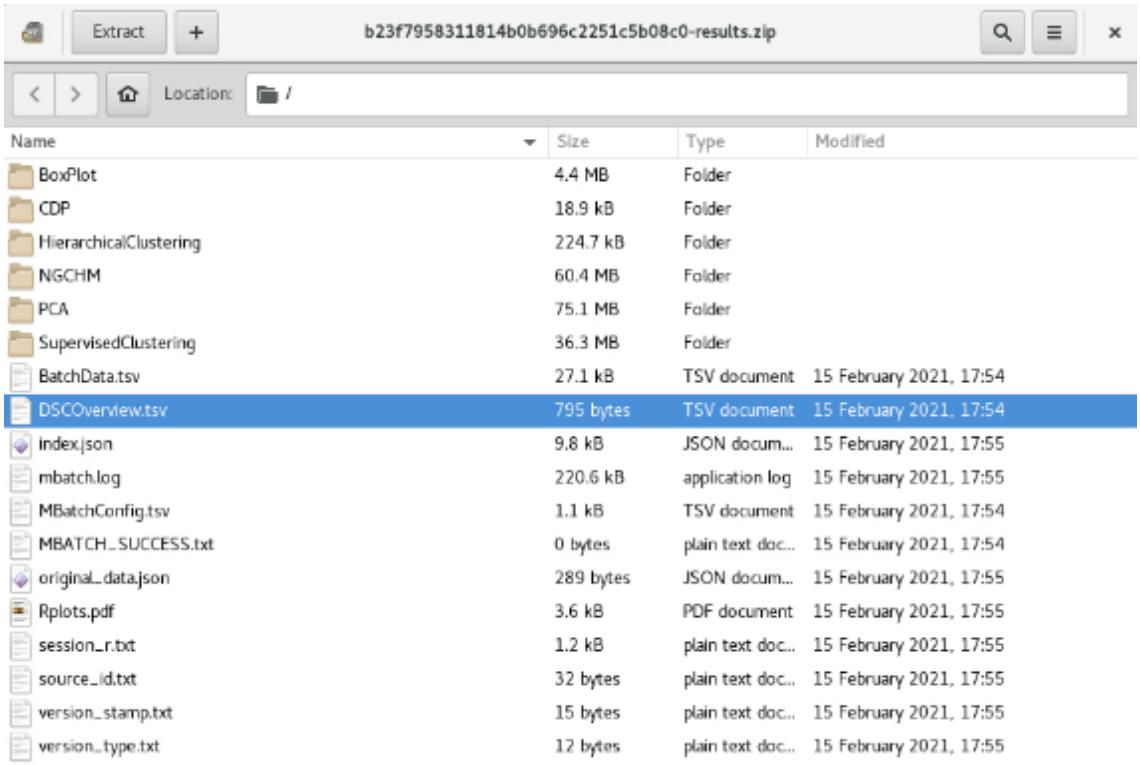
Component PC3,PC4 Disp. within groups (Dw) (3,4) 8.19156934828866

Component PC3,PC4 Disp. between groups (Db) (3,4) 0

Component PC3,PC4 DSC pvalue(3,4) <0.0005

Finding DSC Results from ZIP Archives

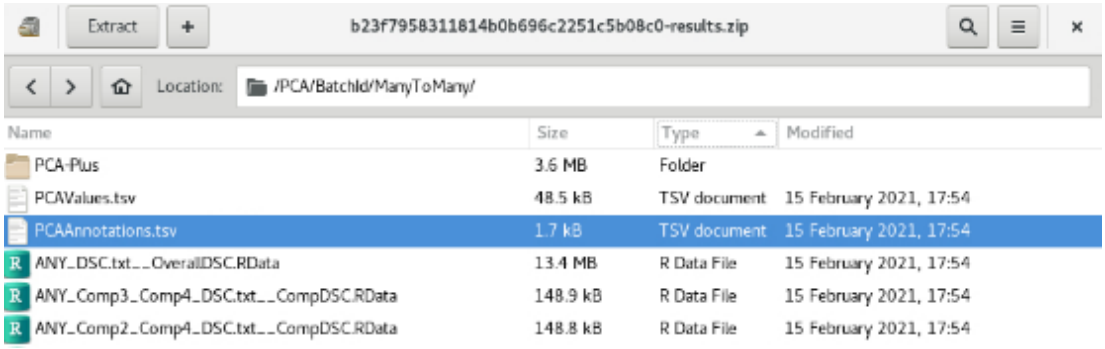
For solutions that use MBatchUtils to create a results ZIP (running the R package from a config file or using Batch Effects Interface), look for the DSCOverview.tsv file in the results ZIP file.



The first three columns of the file are dataset, Overall-DSC and Overall-DSC-pvalue. The other columns give the DSC and p-values for the requested principle components.

dataset	Overall-DSC	Overall-DSC-pvalue	DSC(1,2)
/PCA/BatchId/ManyToMany	0.507946205	<0.0005	1.130076
/PCA/ShipDate/ManyToMany	0.493434725	<0.0005	1.163249
/PCA/TSS/ManyToMany	0.391758406	<0.0005	0.762980

If you are just running the **MBatch R Package or otherwise not generating a results ZIP**, look in the output directory for the PCA directory, and then the batch type (BatchId in the example below) and then ManyToMany. There will be a file named PCAannotations.tsv.



This file is tab-delimited. The rows with Type equal to Diagram have the DSC information.

The row with Annotation equal to "Disp. Sep. Crit. (DSC)" has the DSC value in the Value column.

The row with annotation "DSC pvalue" has the pvalue in the Value column.

The Dw and Db values are also available.

Type	SubType	Annotation	Value
Run	-	MBatch Version	1.7.3
Run	-	DSC Permutations	2000
Diagram	-	Disp. Sep. Crit. (DSC)	0.507946204854667
Diagram	-	Disp. within groups (Dw)	22.6988960922084
Diagram	-	Disp. between groups (Db)	11.5298181244277
Diagram	-	DSC pvalue	<0.0005
Component	PC1,PC2	Disp. Sep. Crit. (DSC) (1,2)	1.13007684768579
Component	PC1,PC2	Disp. within groups (Dw) (1,2)	8.40926432601962