

## SubMapBrowser Documentation

**Description:** Surveys all possible subclass mappings  
**Author:** Yujin Hoshida (Broad Institute) [gp-help@broad.mit.edu](mailto:gp-help@broad.mit.edu)

**Summary:** Subclass Mapping is a method to search for matching pairs of subclasses between two independent data sets [1]. The SubMapBrowser module can be used to search for optimal subclassification in each data set (i.e. number of candidate subclasses) before applying Subclass Mapping (SubMap) module.

Input data sets should have a common identifier. The intersection of these data sets is automatically extracted.

### References:

1. Hoshida Y, et al. Subclass Mapping: Identifying Common Subtypes in Independent Disease Data sets. PLoS ONE 2(11): e1195, 2007

### Parameters:

Name	Description	Choices
datasetA file	Input dataset A - .gct <b>Note:</b> Remove spaces from sample names.	
datasetB file	Input dataset B - .gct <b>Note:</b> Remove spaces from sample names.	
classA file	Input class label A (format shown below)	
classB file	Input class label B ( format shown below )	
num marker genes	Number of marker genes to be mapped	
num perm	Number of random permutations for enrichment score (ES)	
num perm fisher	Number of random permutations for Fisher's statistics	
weighted score type	Weight enrichment by correlation vector (signal-to-noise ratio)	
null distribution	null distribution method	pool (default): pool permutations for all cells of SA matrix; each: use permutations for each cell
p value correction	p-value correction method	FDR (default): Benjamini and Hochberg, J Royal Stat Soc B, 1995. 57:289; Bonferroni
cluster rows	Cluster dataset A's subclass in heatmap of SA matrix	yes (default); no
cluster columns	Cluster dataset B's subclass in heatmap of	yes (default);

# GenePattern

	SA matrix	no
create legend	Create legend for heatmap	yes (default); no
random seed	Random seed	3876352 (default)
output filename	Name of output file containing summary of all possible subclass association (SA) matrices	

## Format for input class label file:

<b>SampleName</b>	<b>SampleName.1</b>	<b>No.of.classes_1</b>	<b>No.of.classes_2</b>	<b>...</b>
<i>id1</i>	<i>id1</i>	<i>class_label_id1</i>	<i>class_label_id1</i>	...
<i>id2</i>	<i>id2</i>	<i>class_label_id2</i>	<i>class_label_id2</i>	...
<i>id3</i>	<i>id3</i>	<i>class_label_id3</i>	<i>class_label_id3</i>	...
<i>id4</i>	<i>id4</i>	<i>class_label_id4</i>	<i>class_label_id4</i>	...
...	...	...	...	...

Tab-delimited text file (dChip sample info file). The first and second columns identify the samples in the dataset. Each subsequent column identifies the class labels which are used for a candidate subclassification. Class labels are sequential numbers beginning with one. For example:

<b>SampleName</b>	<b>SampleName.1</b>	<b>No.of.classes_2</b>	<b>No.of.classes_4</b>
S45	S45	2	2
S46	S46	2	3
S47	S47	2	2
S48	S48	1	4
S49	S49	2	1
...	...	...	...

SubMapBrowser creates an SA matrix for all possible combinations of the candidate subclassifications in *classA.file* and *classB.file*.

## Output files:

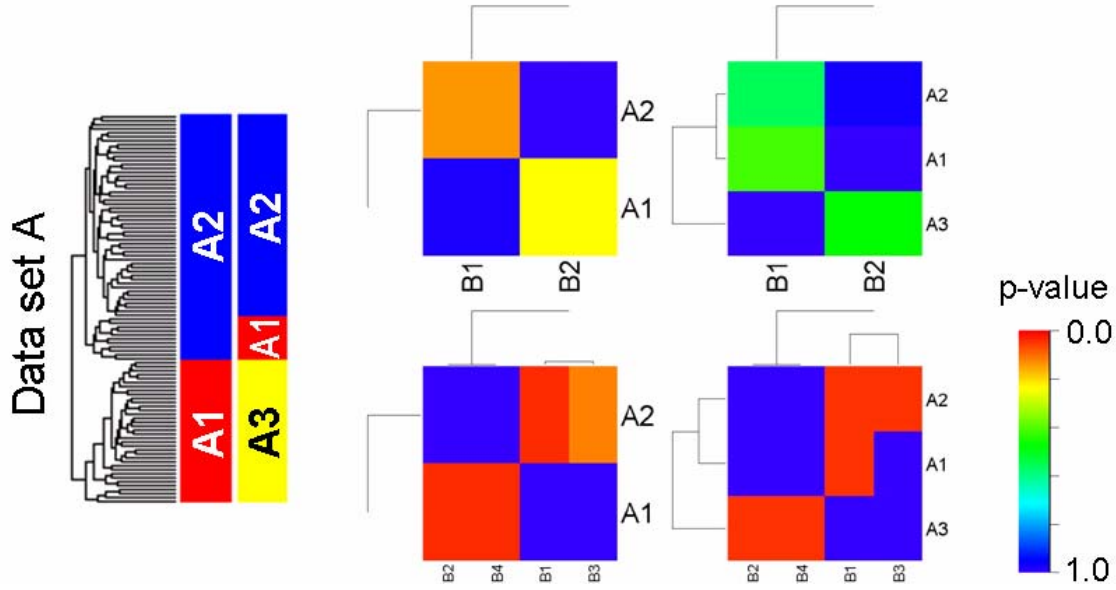
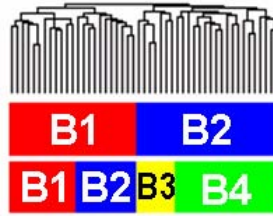
1. <outputfilename>.gct: SA matrix for each combination of candidate subclassifications
2. <output.filename>.jpeg: heatmap of the SA matrix for each combination of candidate subclassifications
3. <output.filename>.png: contains heatmaps of all the SA matrices for all combinations of candidate subclassifications
4. <output.filename>\_parameter.txt: a summary of the input parameters

## Sample output:

In the following example, two candidate subclassifications are provided for dataset A (A1,A2 and A1,A2,A3) and two for dataset B (B1,B2 and B1,B2,B3,B4). SubMapBrowser creates four SA matrices: A1\_B1 (top left), A1\_B2 (lower left), A2\_B1 (top right), and A2\_B2 (lower right).

# GenePattern

Data set B



Platform dependencies:

Module type: Clustering  
OS: Windows  
Language: R