



sMAP: Standard Microarray Analysis Pipeline

An R Shiny Educational App for Transcriptomic Analysis

GitHub: <https://github.com/BI-STEM-Away/sMAP>

Documentation: https://bi-stem-away.github.io/sMAP_doc/

sMAP Website: <https://bi-stem-away.github.io/sMAP/>



Welcome screen of sMAP

sMAP: Standard Microarray Analysis Pipeline

sMAP: Standard Microarray Analysis Pipeline

An application for processing quality control, statistical and functional analysis of a GEO dataset in order to find potential biomarkers.

This app is created by STEM-Away RShiny Project Team - Session 1, 2021

Begin

@BI-STEM-Away sMAP: 2021

Click on
Fullscreen mode
for better
visualization

Click on Begin to
get started with
sMAP.

Data Importation

SMAP: Standard Microarray Analysis Pipeline

Upload Data Files in sMAP to begin

Input Type: Expression Data File (CSV/TXT) | Metadata (CSV) | What microarray platform did you use? Affymetrix Human Gene 1.0 ST Array | Expression Data File (CSV/TXT)

Options in Input Type dropdown: Expression Data File (CSV/TXT), Expression Data File (CSV/TXT), Raw Affymetrix Data (CEL Files), GEO Accession Number, Load Demo Data (Raw)

Select one of the 3 data input methods sMAP provides or get started with an installed demo dataset.

Data Loading

SMAP: Standard Microarray Analysis Pipeline

Upload Data Files in sMAP to begin

Input Type: Load Demo Data (Raw) | Demo data selected. Download from here. | Load Data

	GSM494556	GSM494557	GSM494558	GSM494559	GSM494560	GSM494561	GSM494562	GSM494563	GSM494564	GSM494565	GSM494616
1	143	122	103	125	83	170	155	87	96	165	183
2	11664	10153	9156	15492	9941	17402	8521	10572	9373	8855	8673
3	173	167	123	155	100	194	126	85	98	208	172
4	11646	10607	9341	15444	9782	17385	8535	10833	9882	9233	9034
5	70	60	59	74	59	72	92	68	81	76	64
6	96	138	106	115	65	115	126	89	74	145	146
7	11751	10496	9158	15208	10010	17399	8758	10655	9738	8928	8424
8	113	151	121	159	87	163	191	83	111	186	162
9	11650	10317	9300	15093	9870	17307	8530	10581	9305	8757	8348
10	114	148	121	165	78	123	171	95	105	182	173

Showing 1 to 10 of 1,354,896 entries | Previous 1 2 3 4 5 ... 135,490 Next

Click on load data after selecting input method and uploading datasets.

Here we will use demo dataset that comes installed in sMAP.

QC: Visualization of raw data before normalization

The screenshot shows the SMAP interface with the 'QC & Preprocessing' menu open. The 'Visualize' button is highlighted with a red box. A red line points from this button to the 'Visualize' option in the 'Select QC visualization method before normalization' dropdown. Another red line points from the 'Visualize' button to the 'Visualize' option in the same dropdown. The main content area shows a boxplot titled 'Normalized Unscaled Standard Errors of Samples' with 'Standard Error Values' on the y-axis and 'Sample Names' on the x-axis.

Select one of the 4 visualization methods.

Click on visualize to load the plot.

QC: Normalization of raw data

The screenshot shows the SMAP interface with the 'Normalization' step. The 'Which normalization method do you want to use?' section has 'RMA' selected. The 'Begin Normalization' button is highlighted with a red box. A red line points from this button to the 'Begin Normalization' option in the 'Which normalization method do you want to use?' section. The main content area shows the 'Normalization' step with a 'Back' button and a 'Next' button. The text 'Background correction, normalization, and summarization have been performed.' is visible below the 'Begin Normalization' button.

Select one of the 3 methods for normalization of raw data.

Click on begin normalization to normalize and wait for the message to appear below indicating completion..

QC: Batch Correction + Visualization of normalized data

SMAP: Standard Microarray Analysis Pipeline

Batch Correction & Visualization After Processing

If samples come from different batches, specify which metadata feature indicates the batch each sample belongs to.

Cancer.State

Perform Batch Correction

Choose a QC visualization method after normalization.

Boxplot

Next

Boxplot of Gene Expression for Each Sample

Gene Expression Values

Sample Names

Here we will not use batch correction since demo dataset is from single batch.

Select visualization method to check data after normalization and click next to generate plot.

QC: Detection and remove of outliers

SMAP: Standard Microarray Analysis Pipeline

Find Potential Outliers

Outlier Detection Method

KS

Find Potential Outliers Show Updated List of Samples

Potential Outliers

Select outlier candidates you would like to remove.

Potential Outliers

Value of Selected Statistic

Sample

@BI-STEM-Away

SMAP: 2021

Find potential outliers and if need remove them from dataset.

Sample Grouping for DEG Analysis

Select the feature you wish to analyze for differential gene expression.

- Cancer.State
- Cancer.State

Select column name in metadata file which contains sample grouping information.

Gene Filtering and Differentially Expressed Gene Analysis

What percent of genes should be filtered out?

Filter Out Genes

The percent of genes you choose will be the percent of genes that will be filtered because of the low levels of expression. Additionally any duplicates or NA gene symbols will be removed.

Cutoff LogFC Value: 0 to 5

Cutoff adjusted p-value: 0.01 to 0.2

Find DEGs

Genes with a lower adjusted p-value than the cutoff and with a higher magnitude log fold change will be considered significantly differentially expressed.

Show 10 entries

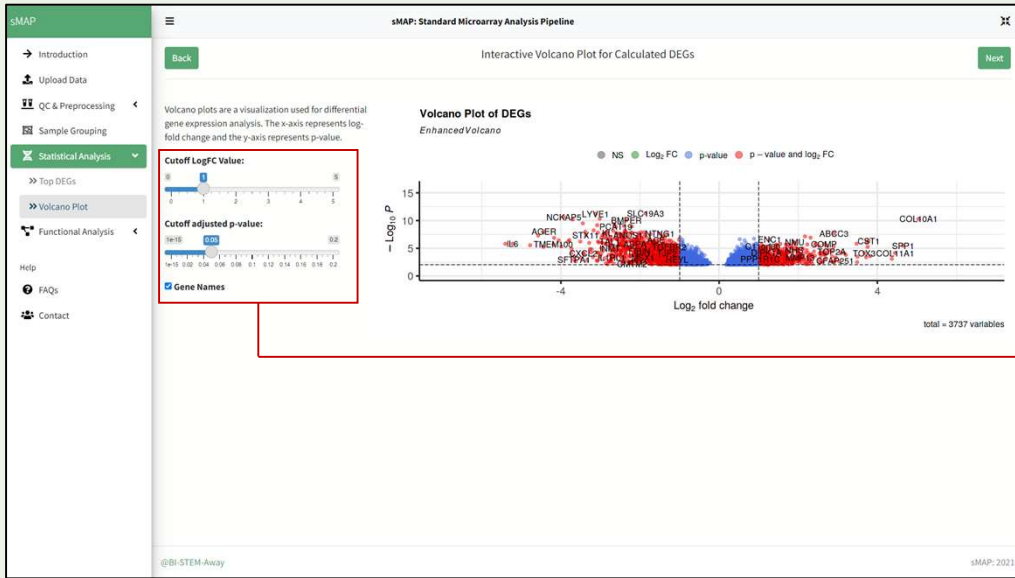
	logFC	AveExpr	t	PValue	adj.PVal	B
SLC19A3	-1.86433140321532	5.73388768404938	-14.1944049069961	5.13133190481571e-12	6.57115813523244e-8	17.0717847375712
LVE1	-3.12354095217871	8.23360047695039	-14.0576741592034	6.13983474443583e-12	6.57115813523244e-8	16.9168566892787
NCKAP5	-3.92068855210225	5.74962587819348	-13.0626544812896	2.3695222310211e-11	1.69095411183355e-7	15.73660196511
CCBE1	-3.01801219482324	7.73475406386807	-12.5982516044147	4.57869096984782e-11	2.2335463186728e-7	15.1522832469768
COL10A1	5.0382869564668	8.27574121251777	12.5077601380164	5.21734715877785e-11	2.2335463186728e-7	15.0358211015996
RTKN2	-3.70165819099124	5.69296533811371	-12.3596143997755	6.4711831604782e-11	2.3085945925006e-7	14.8432830223856
BMPEP	-2.34940052571804	5.14157289175724	-12.1061673308641	9.39803422939486e-11	2.8737846097111e-7	14.5084061524504
MED12L	-3.44550319113184	6.71913379584858	-11.2999382291861	3.20885450240023e-10	8.58569132788461e-7	13.3953269752683

Adjust required parameters for DEG analysis.

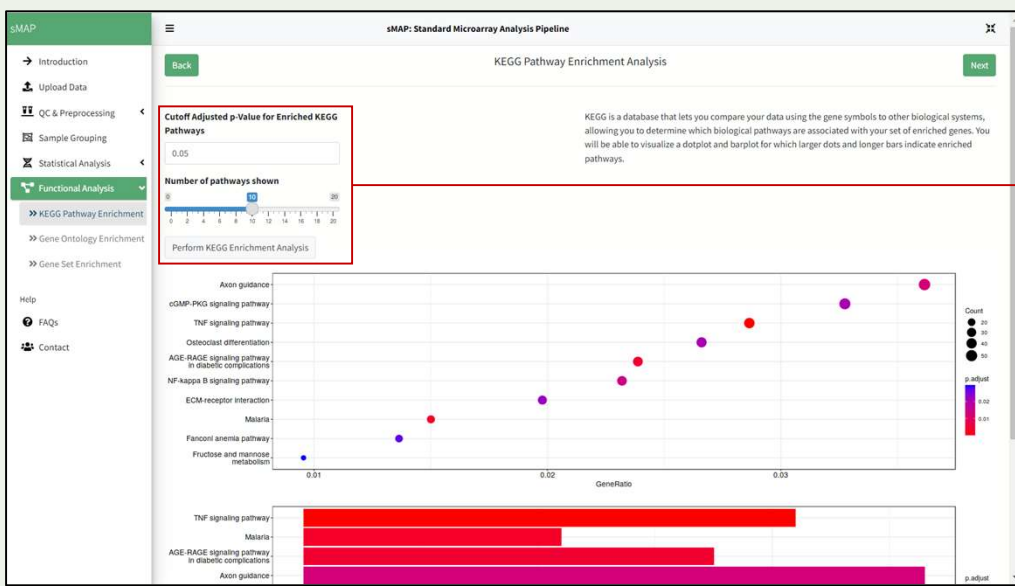
Click on find DEGs to start analysis.

* We didn't use gene filtering option for this demo but can be used as per requirement of user.

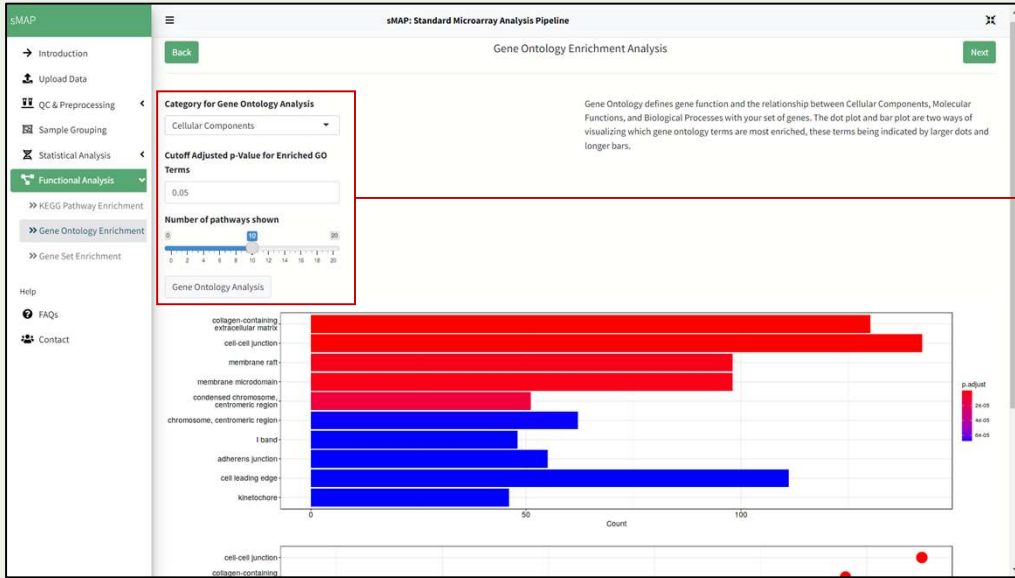
Volcano plot



Enrichment Analysis: KEGG Pathway Enrichment Analysis



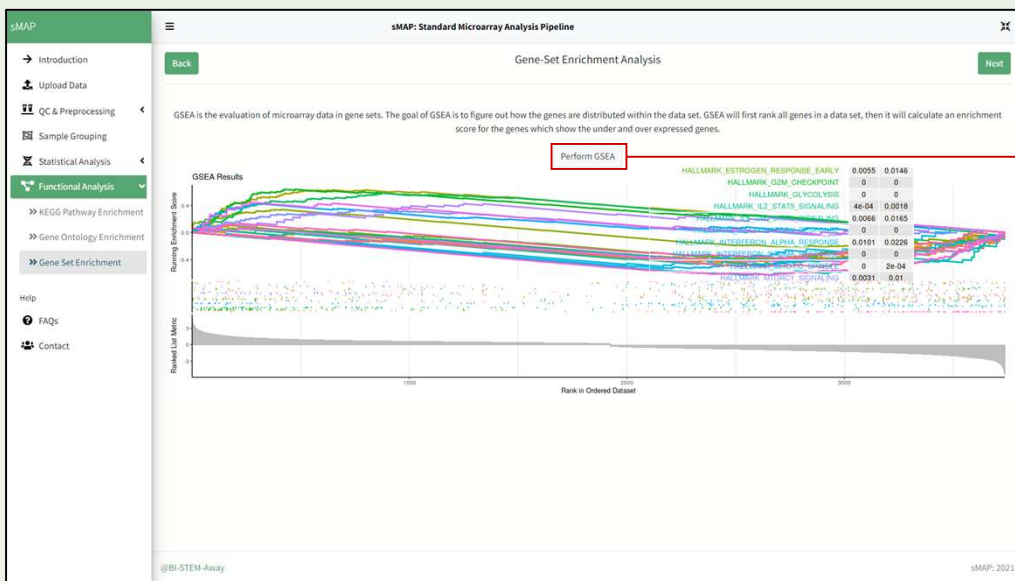
Enrichment Analysis: Gene Ontology Enrichment Analysis



Select one of the three GO categories.

Adjust P-value and other parameters for GO enrichment analysis.

Enrichment Analysis: Gene Set Enrichment Analysis



Perform GSEA analysis.

Help Sections: Frequently Asked Questions

The screenshot shows the SMAP help page with a sidebar on the left containing navigation links: Introduction, Upload Data, QC & Preprocessing, Sample Grouping, Statistical Analysis, Functional Analysis, KEGG Pathway Enrichment, Gene Ontology Enrichment, Gene Set Enrichment, Help, **FAQs**, and Contact. The main content area displays several FAQ cards with titles like 'What are the differences between all the file options in data importation?', 'What does NUSE and RLE do in quality control?', 'Why do you remove outliers?', 'How does batch correction work?', 'How do I analyze quality control boxplot and PCA plot?', 'How do I analyze volcano plots?', 'Can I get more info on enrichment KEGG analysis and its process getting to the output?', and 'What is the most important step in Functional Analysis?'. A red box highlights the 'FAQs' link in the sidebar, and a red line connects it to the text 'Use this section for FAQs/help' on the right side of the image.

Use this section for FAQs/help

Help section: Contact us and check out stemaway.com

The screenshot shows the SMAP help page with a sidebar on the left containing navigation links: Introduction, Upload Data, QC & Preprocessing, Sample Grouping, Statistical Analysis, Functional Analysis, KEGG Pathway Enrichment, Gene Ontology Enrichment, Gene Set Enrichment, Help, FAQs, and **Contact**. The main content area displays contact information including 'Project Technical Lead: Samuel Bharti', 'Team Members' (Nikita Krishnan, Disha Chauhan, Arian Veyssi, Maryam Momeni, Roman Ramirez, Hou Wang (Ivan) Lam, Aditi Verma, Huikun (Kelly) Li, Sneha Raj), 'Acknowledgement' text, 'How to cite us' (Updated soon), and 'Contact Us' (Email: Samuel Bharti, samuelbharti@gmail.com). A red box highlights the 'Contact' link in the sidebar. Two red lines point from the 'Contact' link to the text 'Visit STEM-Away to check out more interesting projects and see how STEM-Away is changing the shape of hiring industry and career advancements' and 'Visit our GitHub Page to clone repo and get docker image' on the right side of the image.

Visit STEM-Away to check out more interesting projects and see how STEM-Away is changing the shape of hiring industry and career advancements

Visit our GitHub Page to clone repo and get docker image