

# PROCEDURE

---

## Notice

The web version of PhosMap is for quick start of visualization only due to the low-level hardware of R shiny server. It is single-threaded and we recommend users to analyze small data sets using the demo server. For larger datasets, upgraded hardware is necessary according to the possible computational cost of the data. We recommend users to use the local docker version of PhosMap.

## Introduction of example data

Here, we reanalysis (phospho)proteomic profilings of WiDr colorectal cancer cells harbouring the BRAF(V600E) mutation after treatment using vemurafenibin a time course of 0, 2, 6, 24, and 48 hour<sup>[1]</sup>. The raw files were deposited in ProteomeXchange Consortium(PXD007740). The raw data were processed in Firmiana, a one-stop proteomic cloud platform<sup>[2]</sup>, to obtain quantitative peptide and protein files. You can download example data in <https://github.com/liuzan-info/PhosMap/tree/master/examplefile/mascot> and <https://github.com/liuzan-info/PhosMap/tree/master/examplefile/maxquant>.

## Analysis and visualization

PhosMap incorporated six analysis modules: dimension reduction analysis, differential expression analysis, time course analysis, kinase activity prediction, phosphorylation motif enrichment analysis and survival analysis.

### Upload data

#### Function

In this step, you can upload your preprocessed data to PhosMap, such as the phosphorylation dataframe. If you have not preprocessed your data, you must preprocess it with PhosMap (go to the 'Preprocessing' tab) or do it yourself.

#### How to upload your data

1. Go to the 'Import data' tab.
2. Choose 'Load example data' or follow the prompts to upload your own corresponding four files. If you have not preprocessed your data, you can click 'Go to preprocessing' to preprocess it.

The screenshot shows the PhosMap shiny app interface. On the left, there's a sidebar titled 'Import Data' with options for 'your data' (radio button) and 'example data' (radio button), and a link to 'Go to preprocessing'. Below this are three sections: '1. Experimental design file:' with a 'view' button, '2. Phosphorylation data frame:' with a 'view' button, and '3. Clinical data file[optional]:' with a 'view' button. On the right, the 'Data Overview' section displays a table of 39 entries. The columns are 'Experiment\_Code', 'Group', and 'Description'. The data includes rows like 'Exp027012' with Group 0 and Description 'ctr\_0h\_R1\_IMAC\_1.raw', and 'Exp027021' with Group 2 and Description 'PLX\_2h\_R1\_IMAC\_2.raw'. A search bar and navigation buttons (Previous, Next) are at the bottom of the table.

Experiment_Code	Group	Description
1 Exp027012	0	ctr_0h_R1_IMAC_1.raw
2 Exp027013	0	ctr_0h_R1_IMAC_2.raw
3 Exp027014	0	ctr_0h_R1_IMAC_3.raw
4 Exp027015	0	ctr_0h_R2_IMAC_1.raw
5 Exp027016	0	ctr_0h_R2_IMAC_2.raw
6 Exp027017	0	ctr_0h_R2_IMAC_3.raw
7 Exp027018	0	ctr_0h_R3_IMAC_1.raw
8 Exp027019	0	ctr_0h_R3_IMAC_2.raw
9 Exp027020	2	PLX_2h_R1_IMAC_1.raw
10 Exp027021	2	PLX_2h_R1_IMAC_2.raw

## Dimension reduction analysis

### Function

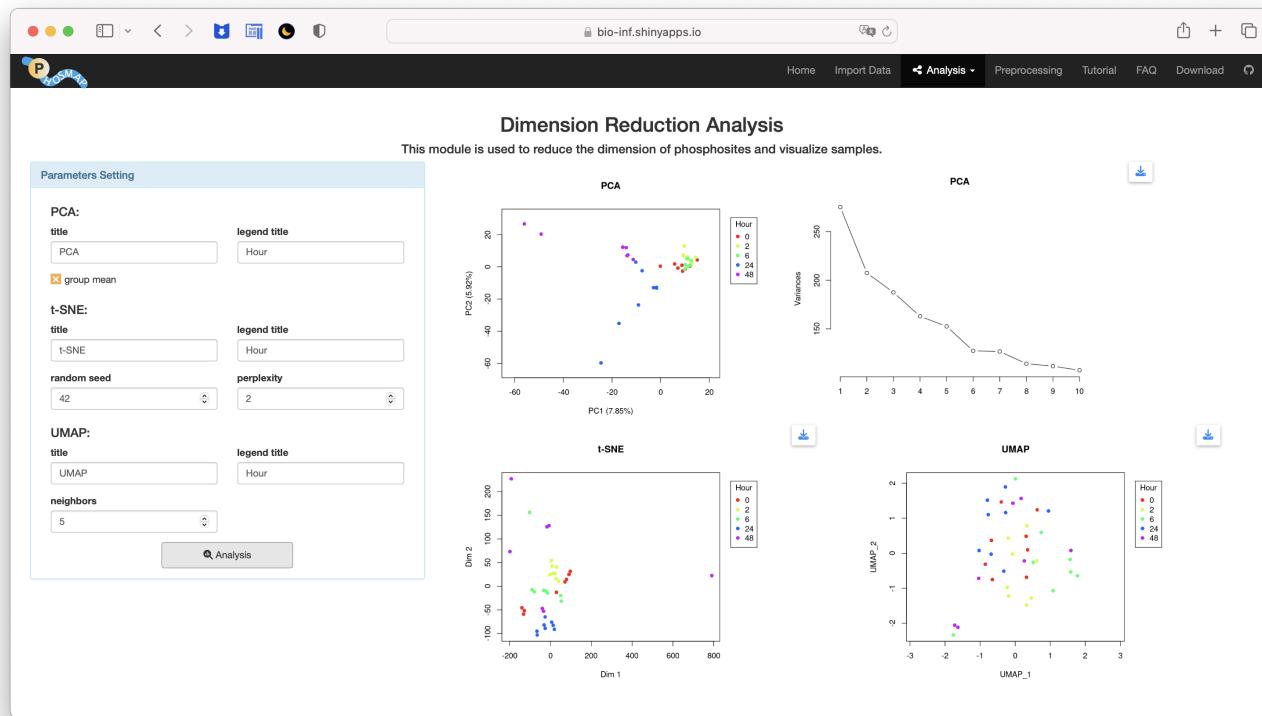
In PhosMap, Dimension reduction analysis methods allowed for PCA, t-SNE and UMAP.

### The meaning of the parameters

1. 'Title' refers to the main title of the plot.
2. 'Legend title' refers to the title of the legend in the plot.
3. 'Random seed' is a parameter for t-SNE that sets the seed for the random number generator. This can be used to ensure reproducibility of results.
4. 'Perplexity' is a numerical value for t-SNE, with a default value of 2. It balances the focus between preserving the local and global structure of the data.
5. 'Neighbors' is a parameter for UMAP that refers to the size of the local neighborhood (in terms of the number of neighboring sample points) used for manifold approximation. Larger values result in more global views of the manifold, while smaller values result in more local data being preserved.

### How to get analysis results

1. Go to the 'Dimension reduction analysis' under 'Analysis' tab.
2. Modify the parameters according to your needs.
3. Click the 'Analysis' button.
4. The PCA, t-SNE and UMAP plot after running will appear on the right.
5. Click the download button to download the plot file.



## Interpretation of analysis results

To extract an overview of the effect of the different time course treatments, we performed PCA analysis in the downstream analysis module of PhosMap. We could see that phosphorylation expression profiles of colorectal cancer cells after longer (24h and 48h) vemurafenib treatment were quite different from those after short treatment (2h and 6h). In addition, it shows that principal component 1 (PC1), with 31.77%, is superior to 20% from original literature and demonstrates phosphorylation expression profile normalized by matched proteomics data has an advantage over representing the variation over time in the BRAFi-treated samples.

## Differential expression analysis

### Function

In PhosMap, differential expression analysis methods allowed for limma, SAM and ANOVA Data analysis.

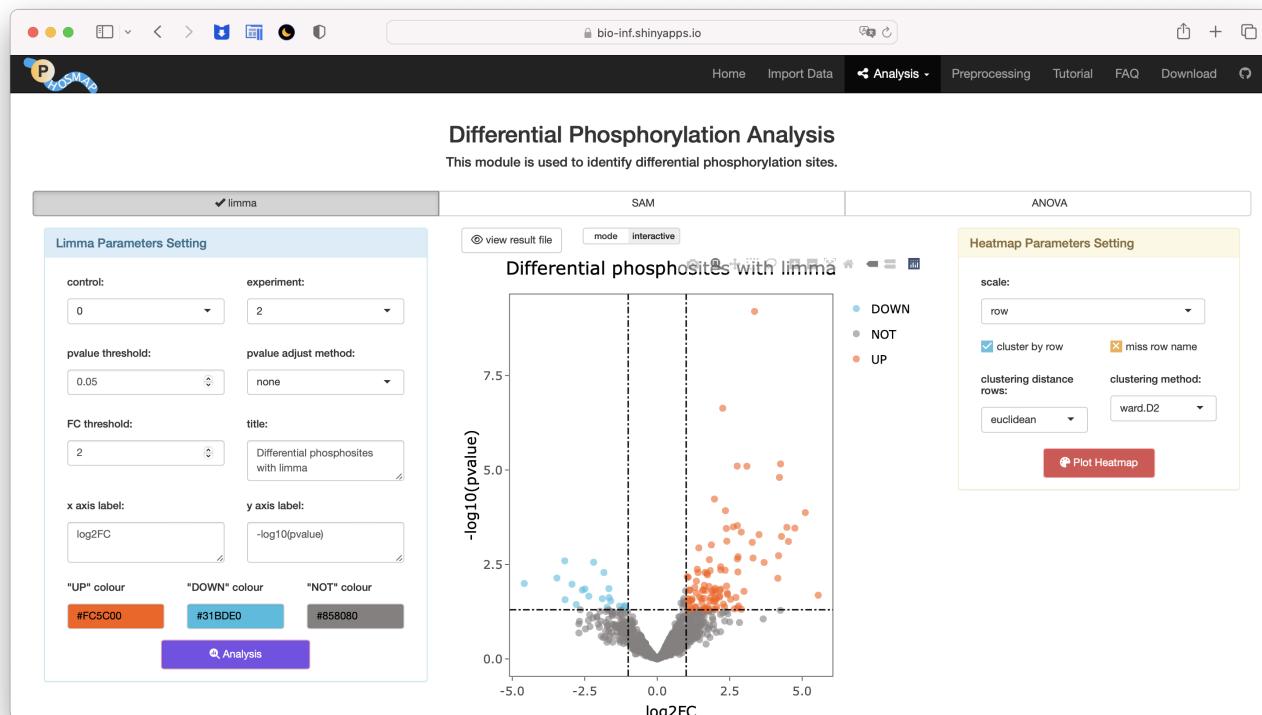
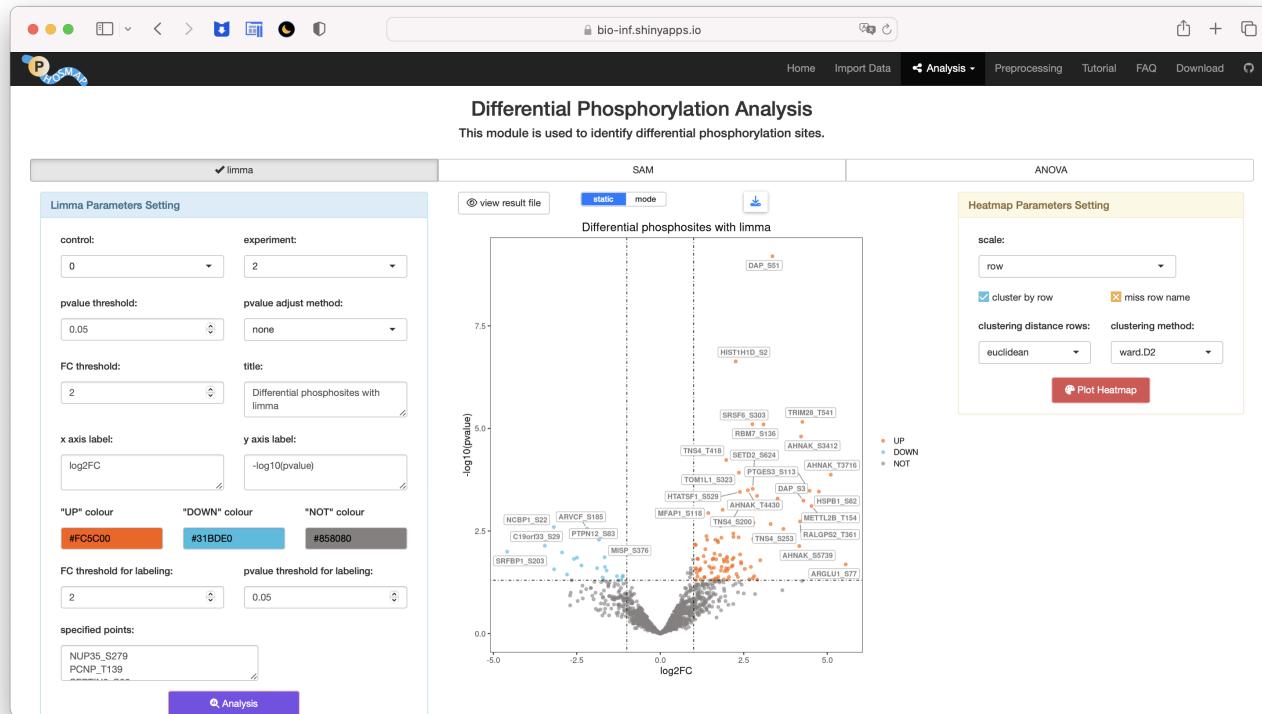
### The meaning of the parameters

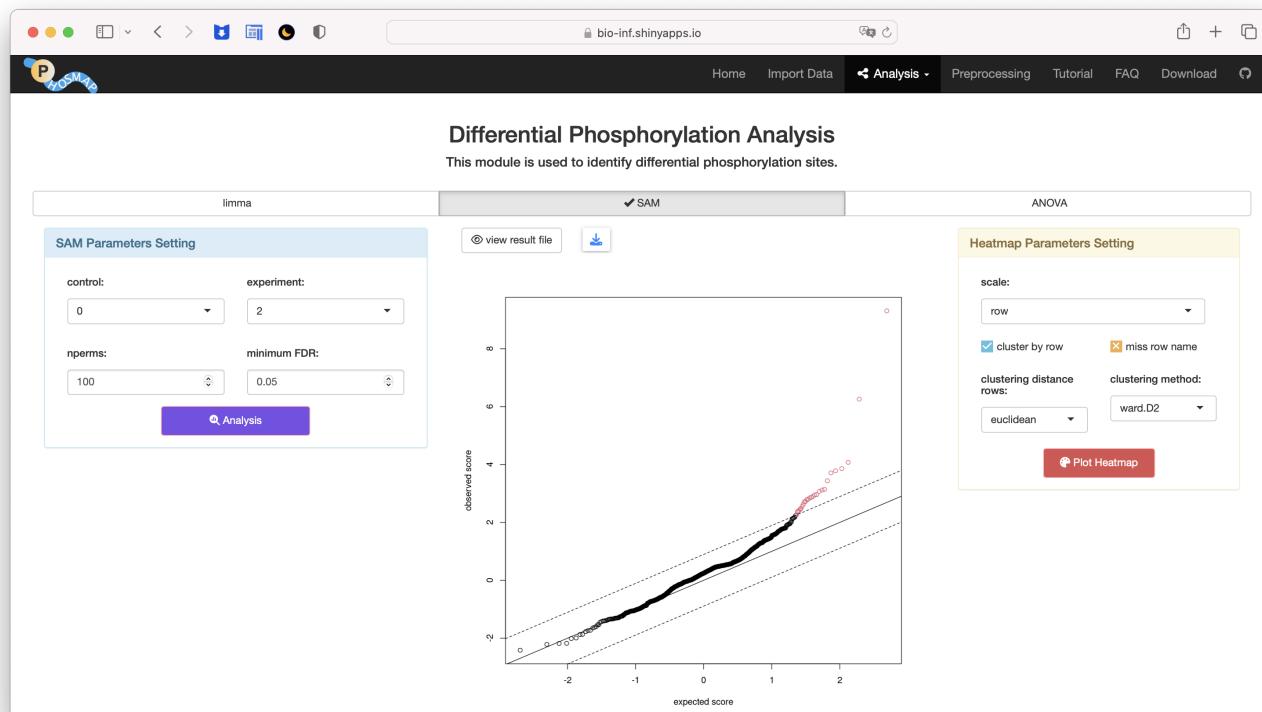
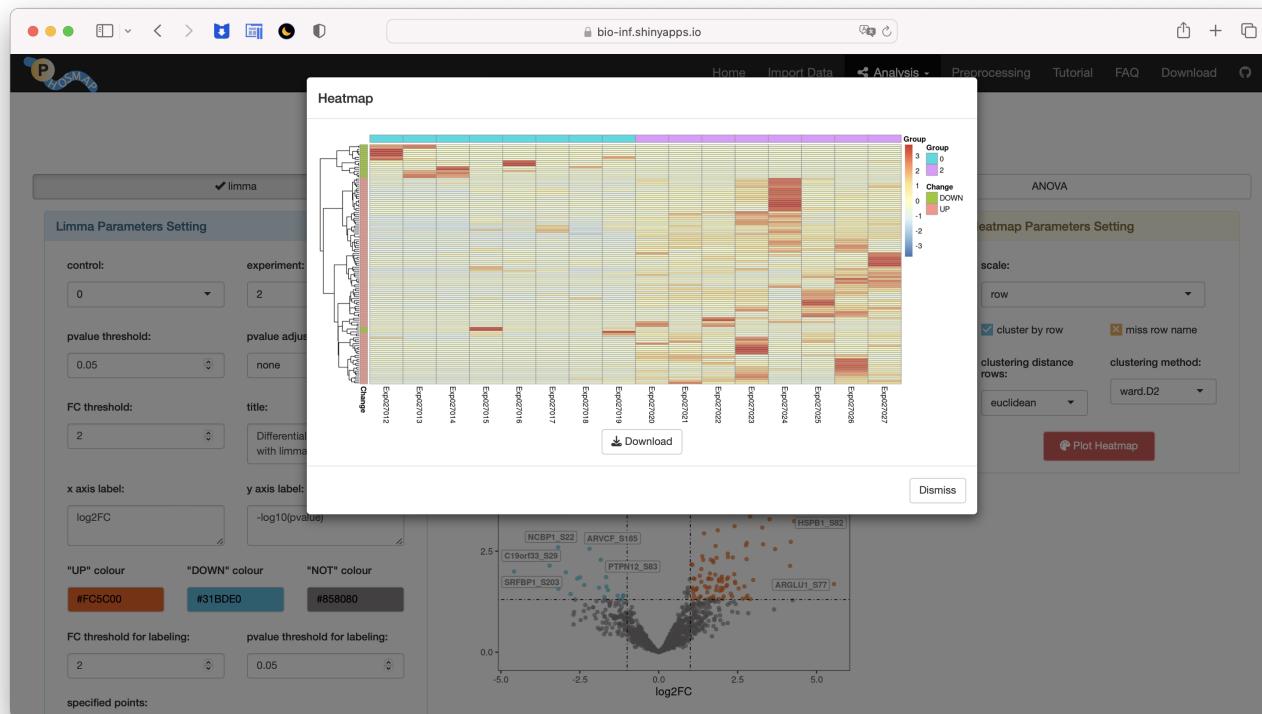
1. 'Control' refers to the control group in the experiment.
2. 'Experiment' refers to the experimental group in the experiment.
3. 'P-value threshold' is the threshold for determining statistical significance based on the p-value.
4. 'P-value adjust method' is the method used to adjust p-values for multiple comparisons.
5. 'FC threshold' is the fold change threshold for determining significant changes in phosphorylation levels.
6. 'nperms' is a parameter for the SAM method that specifies the number of permutations to perform.

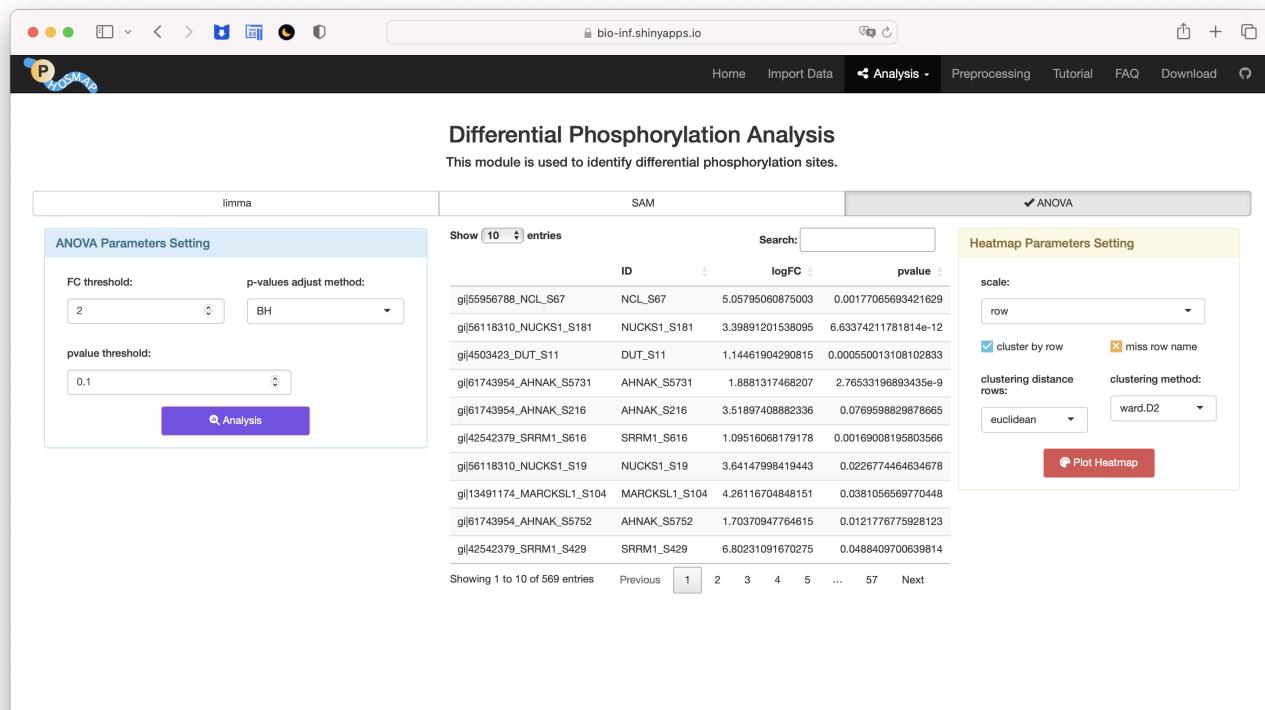
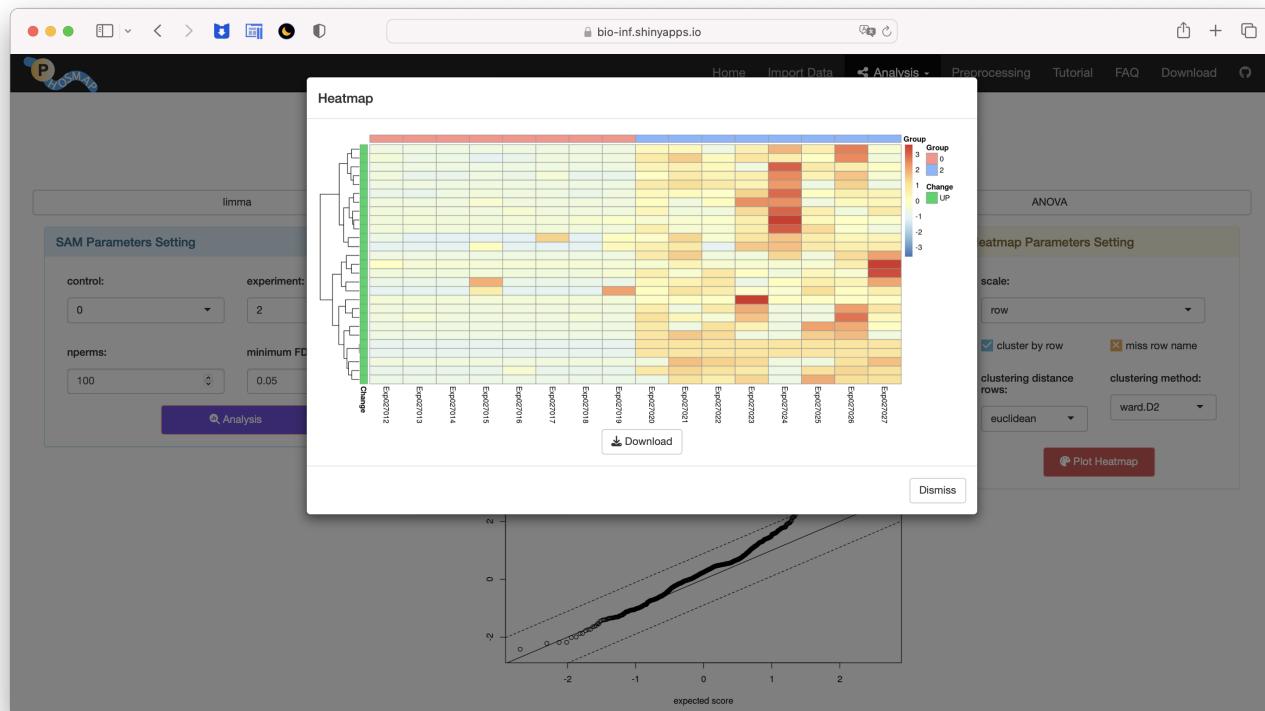
7. 'Minimum FDR' is the minimum false discovery rate threshold for determining statistical significance.
8. 'Clustering distance rows' is a parameter for heatmap generation that specifies the distance metric used for clustering rows.
9. 'Clustering method' is a parameter for heatmap generation that specifies the clustering method used to cluster rows and columns.

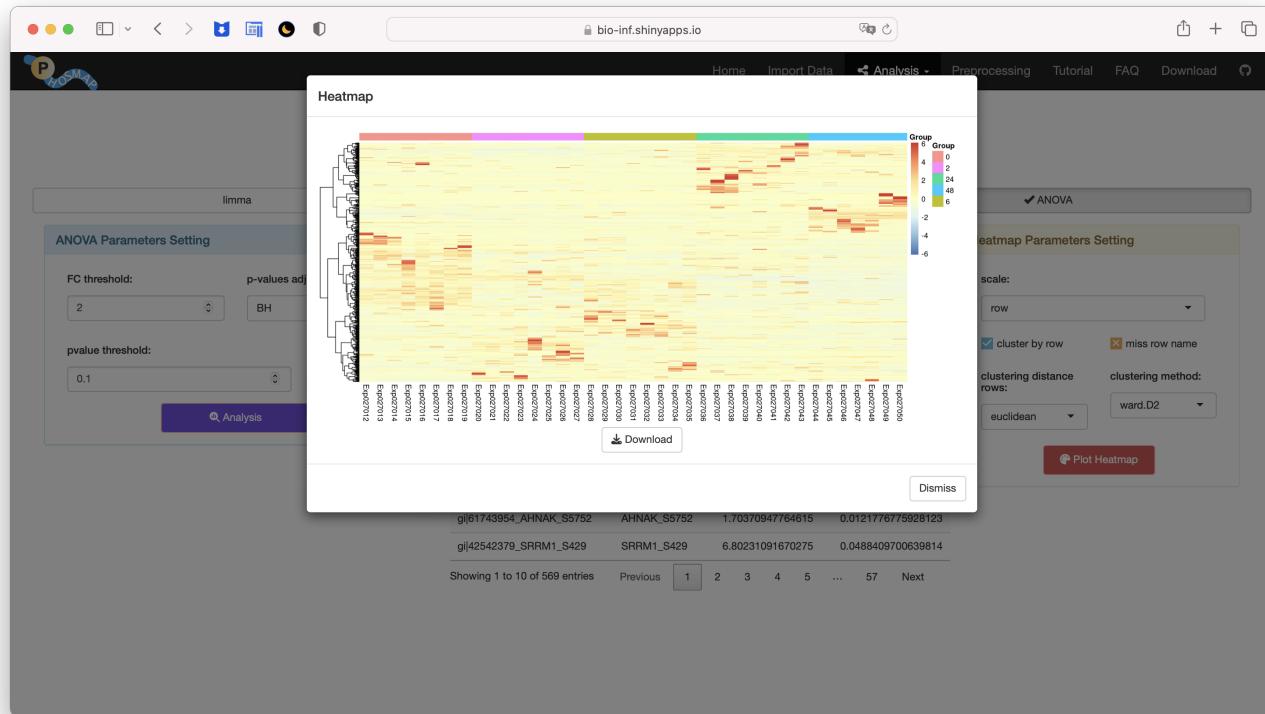
## How to get analysis results

1. Go to the 'Differential Expression Analysis' under 'Analysis' tab.
2. Go to the 'limma', 'SAM' or 'ANOVA' secondary tab.
3. Choose Control and Experiment used for differential Expression Analysis.
4. Choose 'Interactive mode' and click the 'Analysis' button. The interactive plot after running will appear on the right.
5. Choose 'Static mode' and click the 'Analysis' button. The static plot after running will appear on the right.
6. Click 'Plot Heatmap' button. The heatmap will appear in the pop-up window.
7. Click the download button to download the plot file.









## Interpretation of analysis results

In order to show differential expression analysis between two experimental conditions. We use the limma method integrated into differential expression analysis module of PhosMap to identify 128 significant differently expressed p-sites (DEPs) between the samples with BRAFi-treated for two hours and control samples (P value < 0.05 and fold change > 2). 139 p-sites were up-regulated in the BRAFi-treated samples. The most disparate difference is observed in DAP\_S51, whose phosphoserine is related to the MTOR pathway. 99 p-sites were down-regulated in the BRAFi-treated samples.

For the multiple experimental conditions, we leveraged the embedded ANOVA analysis of PhosMap and identified 548 DEPs among the five time points (P value < 0.1 and fold change > 2).

## Time Course Analysis

### Function

Fuzzy clustering was applied to time course analysis for discovering patterns associated with time points in PhosMap. The corresponding line chart combined with membership for each cluster was also drawn.

### The meaning of the parameters

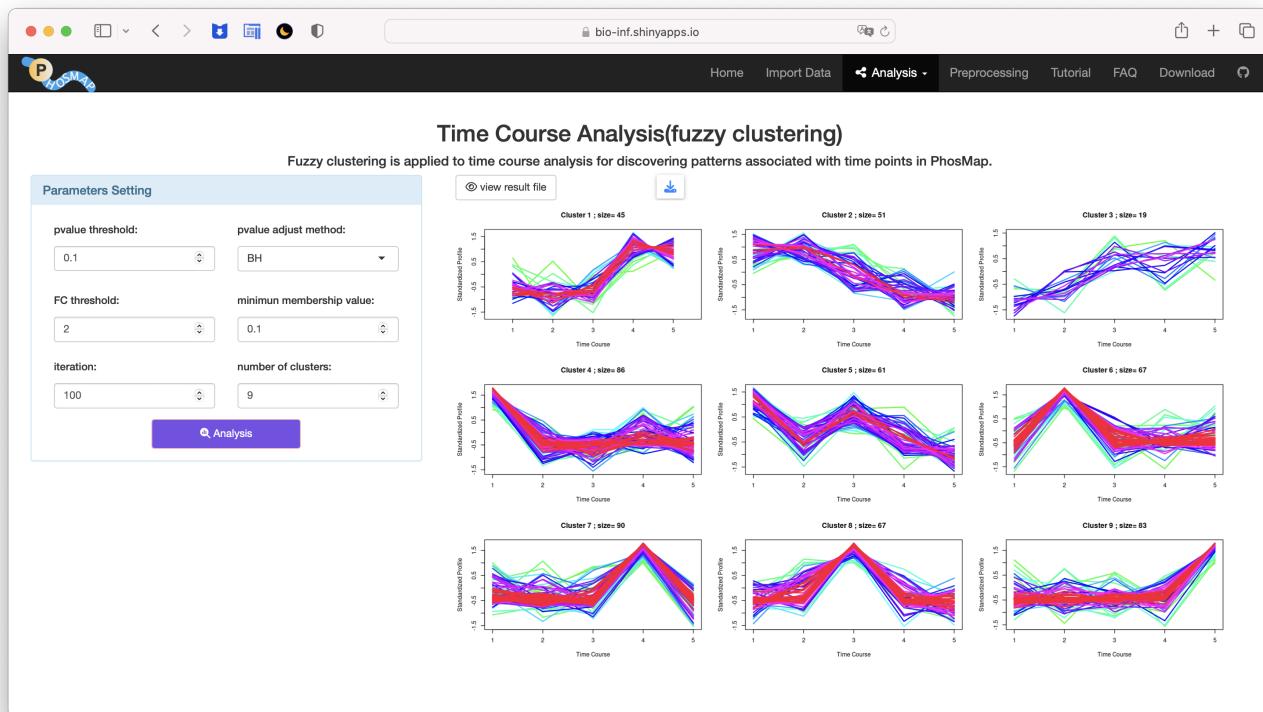
1. 'Minimum membership value' is a threshold for determining the minimum membership value for a data point to be included in a cluster.
2. 'Iteration' is the number of iterations to perform in the clustering algorithm.
3. 'Number of clusters' is the number of clusters to generate in the clustering algorithm.

### How to get analysis results

1. Go to the 'Time course Analysis (fuzzy clustering)' under 'Analysis' tab.
2. Modify the parameters according to your needs.
3. Click the 'Analysis' button. The plot after running will appear on the right.
4. Click the download button to download the plot file.

## Interpretation of analysis results

These 548 DEPs were used as inputs in the time course analysis module of PhosMap, then 9 strong expression patterns were generated. Two major clusters show significant downregulation at the phosphoproteomics signalling level upon BRAFi treatment in line with the original literature. Cluster 1 responds within 2 hours, an early treatment response. Cluster 2 responds within 24 hours, a late treatment response.



## Kinase activity prediction (KSEA)

### Function

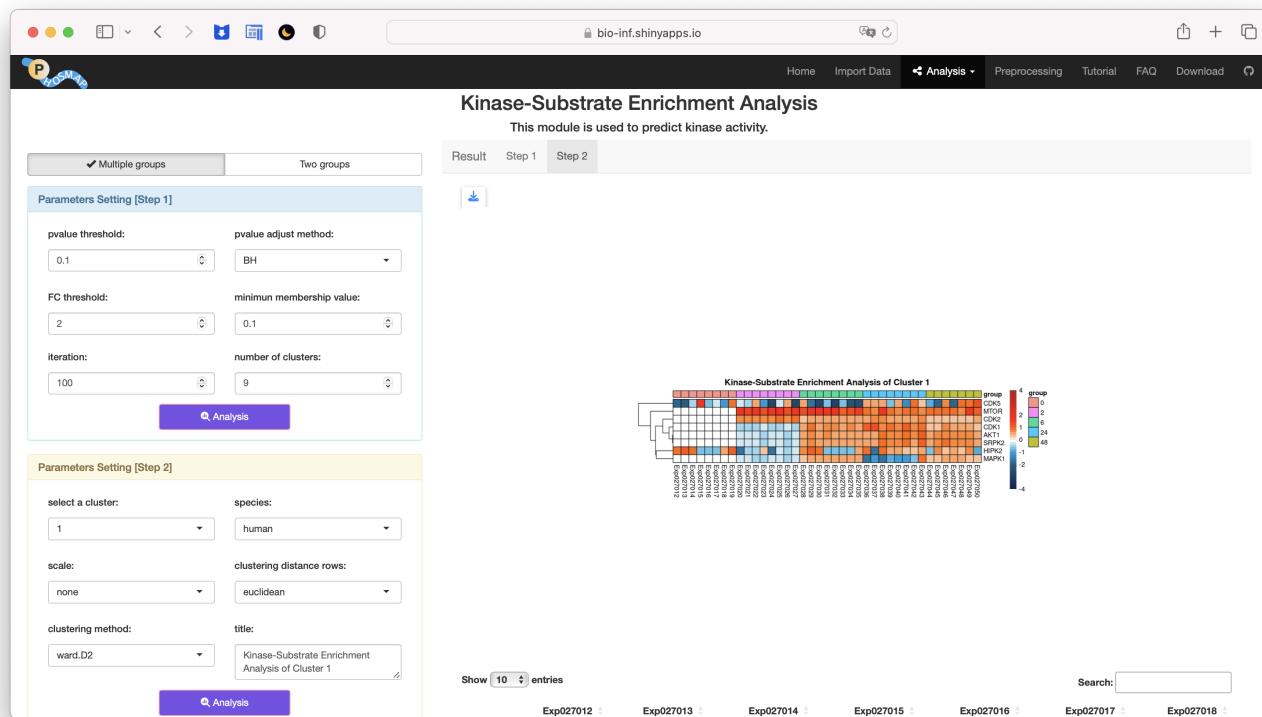
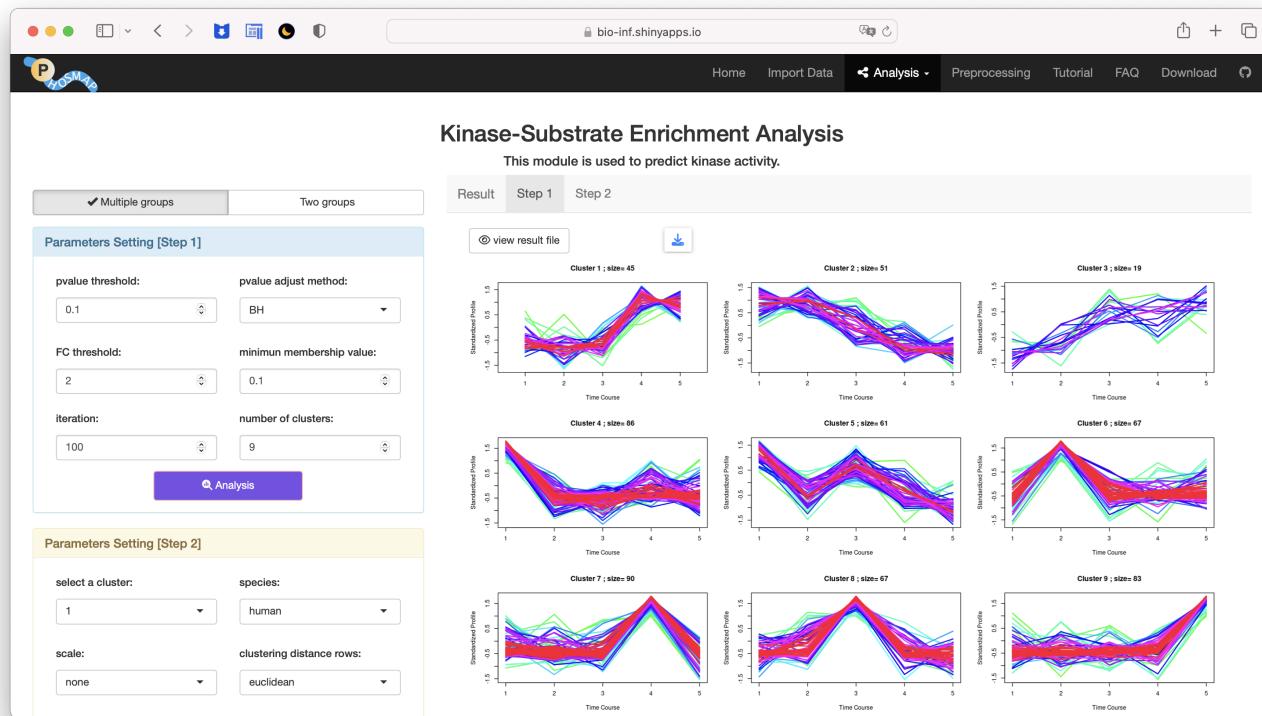
In PhosMap, KSEA was used to predict kinase activity.

### The meaning of the parameters

1. 'Control' refers to the control group in the experiment.
2. 'Experiment' refers to the experimental group in the experiment.
3. 'Species' refers to the species of the organism being studied.
4. 'Scale' is a parameter for scaling the data before generating the heatmap.
5. 'Clustering distance rows' is a parameter for heatmap generation that specifies the distance metric used for clustering rows.
6. 'Clustering method' is a parameter for heatmap generation that specifies the clustering method used to cluster rows and columns.

## How to get analysis results

1. Go to the 'Kinase-Substrate Enrichment Analysis' under 'Analysis' tab.
2. Select 'Multiple groups' or 'Two groups' according to the number of groups of your data.
3. Click the first 'Analysis' button. If 'Multiple groups' is selected, after running, the plot will appear on the right. Click 'view result' to view and download the kinase prediction time course result. If 'Two groups' is selected, only the phosphorylation dataframe will appear on the right.
4. Select a cluster if 'Multiple groups' is selected. Click the second 'Analysis' button. After running, the heatmap will appear on the right.
5. Click the download button to download the plot file.



Kinase-Substrate Enrichment Analysis  
This module is used to predict kinase activity.

Result Step 1 Step 2

Show 10 entries Search:

ID	Exp027012	Exp027013	Exp027014	Exp027015	Exp027016
gi 134152708_ARGLU1_S77	ARGLU1_S77	0.00122345917384075	1.90761488725083	2.10581483285106	0.00122345917384075
gi 103471995_SRFBP1_S203	SRFBP1_S203	0.693174887039922	0.00199973904878981	0.00199973904878981	6.38109883743686
gi 157649073_CAP1_S308	CAP1_S308	2.82446144974694	2.83969640165206	1.08349213539089	0.000935231246922691
gi 65549640_SEPTIN2_S218	SEPTIN2_S218	1.4728663540898	2.36419485031788	1.9508567526532	0.287797559649905
gi 10835067_SSB_S366	SSB_S366	1.60023531503568	1.1359856502941	1.1164413268142	0.00129819757220116
gi 61743954_AHNAK_S135	AHNAK_S135	0.00192066914680648	0.00192066914680648	0.00192066914680648	4.05382391891923
gi 13491174_MARCKSL1_S104	MARCKSL1_S104	0.83577283596371	0.988007631083198	0.912591979250879	1.19788073167183
gi 42542379_SRRM1_T220	SRRM1_T220	6.8818305219933	0.159738492542953	0.159738492542953	0.159738492542953
gi 23308579_PTGES3_S113	PTGES3_S113	0.03968808271462743	0.304635153687522	0.03968808271462743	0.651020845870127
gi 118572613_SRRM2_T1177	SRRM2_T1177	0.0455467033094671	0.0455467033094671	6.99531969332298	0.731400086820214

Showing 1 to 10 of 409 entries Previous 1 2 3 4 5 ... 41 Next

Kinase-Substrate Enrichment Analysis  
This module is used to predict kinase activity.

Result Step 1 Step 2

Parameters Setting [Step 1]

control: experiment:  
0 2

FC threshold:  
4

Analysis

Parameters Setting [Step 2]

species:  
human

scale: clustering distance rows:  
none euclidean

clustering method: title:  
ward.D2 Kinase-Substrate Enrichment Analysis

Analysis

## Interpretation of analysis results

Afterwards, the substrates from the two clusters are imported into the KSEA module of PhosMap to infer kinase activities. The results indicate that CDK1/2, MAPK1/3 and AKT1 are suppressed during BRAFi treatment.

## Motif enrichment analysis

## Function

PhosMap allowed for performing MEA on user defined phosphopeptides lists to provide clues for finding candidate kinases that are not present in the database.

## The meaning of parameters

1. 'Fasta type' refers to the type of fasta file used as input for the analysis.
2. 'Selected row number for plotting motif logo' is the number of rows to be selected for generating the motif logo plot.
3. 'Matched seqs threshold' is the threshold for determining the minimum number of matched sequences required for a motif to be considered significant.
4. 'Scale' is a parameter for scaling the data before generating the heatmap.
5. 'Distance metric' is a parameter for heatmap generation that specifies the distance metric used for clustering rows.
6. 'Clustering method' is a parameter for heatmap generation that specifies the clustering method used to cluster rows and columns.

## How to get analysis results

1. Go to the 'Motif Enrichment Analysis' under 'Analysis' tab.
2. Modify the parameters according to your needs.
3. Click the 'Analysis' button.
4. The foreground dataframe mapped to motifs is shown on the right after running.
5. Select row number for plotting logo.
6. Click the first 'Plot' button, and the logo will appear on the right.
7. Modify the parameters below.
8. Click the second 'Plot' button.
9. The heatmap will appear on the right.
10. Click the download button to download the plot file.

**Motif Enrichment Analysis**

This module is used to find and visualize enriched motifs.

**Motif enrichment analysis result:**

motif	score	foreground_matches	foreground_size	background_matches	background_size	fold_increase
.....SP..S..	32	13	238	0	370	
S.....SP....	32	10	225	0	370	
.....SSP.....	32	8	215	0	370	
.....SP...R.	32	7	207	0	370	
.....W...S.....	16	2	200	0	370	
.....K....SP....	30.8084617345128	6	198	0	370	
.....L.R.S.....	27.3492739476456	11	192	0	370	
.....SD.E....	27.000081356622	8	181	0	370	
.....SP..R..	28.1272204971372	5	173	0	370	
.....K.R....S....	26.8407667007037	6	168	0	370	
.....S.DD....	25.8688269706461	5	162	0	370	
.....SP.G...	24.2416535623975	4	157	0	370	
.....S.D....E	24.3283175400327	5	153	0	370	
.....S.S...E...	24.587942252058	6	148	0	370	
.....R.R....S....	23.5187257244898	5	142	0	370	
.....E.SE.....	22.8036017111921	6	137	0	370	
.....D....SP....	23.0911568071752	3	131	0	370	
.....SDS.....	22.4247740888602	4	128	0	370	

**Motif Selection**

selected row number for plotting motif logo: 1

**Heatmap Parameters Setting**

Assign quantitative values of peptides to their motif

matched seqs threshold: 5

scale: none distance metric: euclidean

clustering method:

**Motif logo**

**Motif-Kinase Relation**

motif	score	foreground_matches	foreground_size	background_matches	background_size	fold_increase
.....SP..S..	32	13	238	0	370	
S.....SP....	32	10	225	0	370	
.....SSP.....	32	8	215	0	370	
.....SP...R.	32	7	207	0	370	
.....W...S.....	16	2	200	0	370	
.....K....SP....	30.8084617345128	6	198	0	370	
.....L.R.S.....	27.3492739476456	11	192	0	370	
.....SD.E....	27.000081356622	8	181	0	370	
.....SP..R..	28.1272204971372	5	173	0	370	
.....K.R....S....	26.8407667007037	6	168	0	370	
.....S.DD....	25.8688269706461	5	162	0	370	
.....SP.G...	24.2416535623975	4	157	0	370	
.....S.D....E	24.3283175400327	5	153	0	370	
.....S.S...E...	24.587942252058	6	148	0	370	
.....R.R....S....	23.5187257244898	5	142	0	370	
.....E.SE.....	22.8036017111921	6	137	0	370	
.....D....SP....	23.0911568071752	3	131	0	370	
.....SDS.....	22.4247740888602	4	128	0	370	

**Motif Selection**

selected row number for plotting motif logo: 1

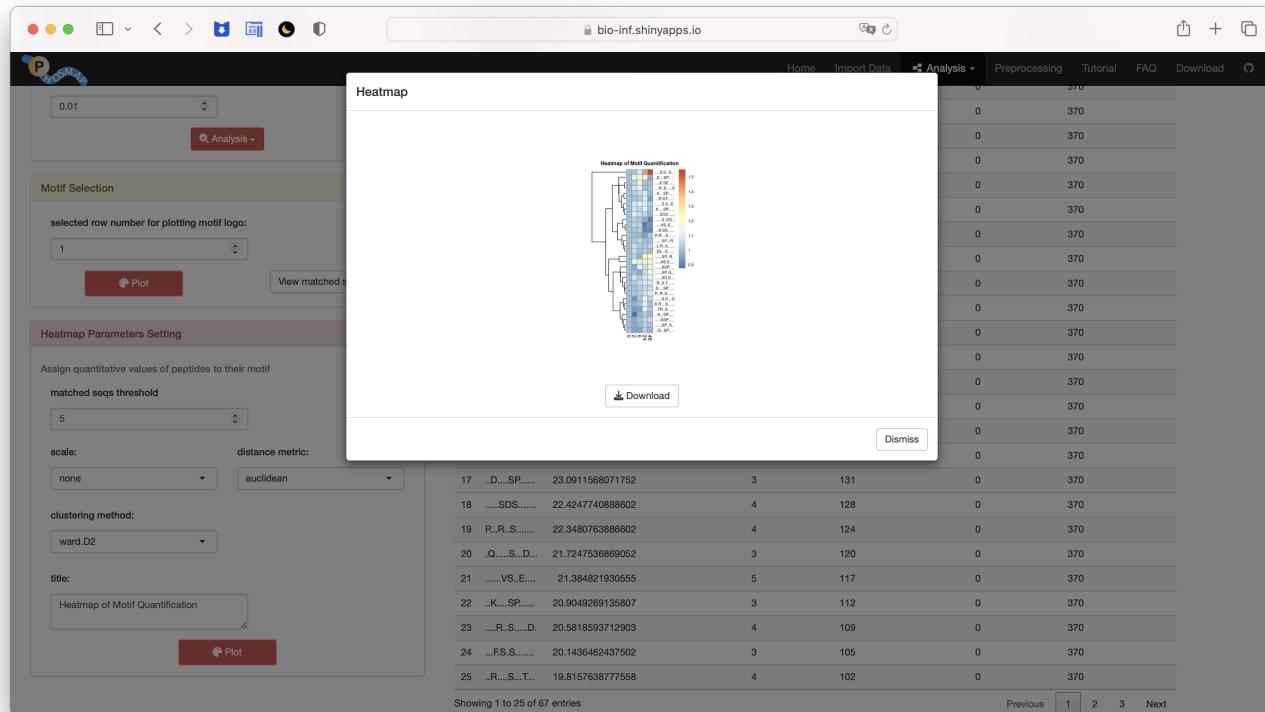
**Heatmap Parameters Setting**

Assign quantitative values of peptides to their motif

matched seqs threshold: 5

scale: none distance metric: euclidean

clustering method:



## Interpretation of analysis results

The 3,649 identified phosphor-peptides as foreground sequences are used for MEA of PhosMap and the results further strengthen the evidence of CDK and MAPK pathway deactivation in BRAF mutant CRC cells in response to BRAFi treatment.

## Survival analysis

### Function

This module is used to identify phosphorylation sites or kinases associated with clinical outcomes of patients. Using kinases or phosphorylation locations files and patients' survival information as input matrices, coxph function from survival R package was used to calculate the hazard ratio (HR) and P-value.

### How to get analysis results

1. Go to the 'Survival Analysis' under 'Analysis' tab.
2. Modify the parameters according to your needs.
3. Click the 'Analysis' button.
4. The summary dataframe list will appear on the right.
5. Click the 'Plot' button.
6. The plot after running will appear on the right.
7. Click the download button to download the plot file.

**Survival Analysis**

This module is used to identify phosphorylation sites or kinases associated with clinical outcomes of patients.

Parameters Setting

pvalue adjust method: BH pvalue threshold: 0.01

Analysis

Feature Selection

Select a feature: gi|110556636\_TNKS1BP1\_S872

"high" colour #3300CC "low" colour #CC3300

Plot

Show 10 entries

Feature	HR	lower_95	upper_95	pvalue	adj_pvalue
gi 55956788_NCL_S67	1.3259	0.8589	2.0467	0.2028	0.937276345933562
gi 156523970_AHSG_S138	1.2348	0.9708	1.5708	0.0857	0.822954794520548
gi 56118310_NUCKS1_S214	1.086	0.5269	2.2386	0.8231	0.994815689655172
gi 56118310_NUCKS1_S181	1.1024	0.3474	3.4982	0.8686	1
gi 5803165_SEC61B_S17	1.1138	0.9135	1.358	0.2864	0.937276345933562
gi 24234699_KRT19_S35	0.887	0.427	1.8423	0.7478	0.98558386396396
gi 61743954_AHNAK_S5841	1.2589	0.9846	1.6098	0.0664	0.791298305084746
gi 45359846_G3BP2_T227	0.4852	0.1615	1.4583	0.1978	0.937276345933562
gi 42542379_SRMM1_S450	0.6556	0.3289	1.3067	0.2302	0.937276345933562
gi 4503423_DUT_S11	0.5296	0.1759	1.5943	0.2583	0.937276345933562

Showing 1 to 10 of 3,505 entries

Previous 1 2 3 4 5 ... 351 Next

**Survival Analysis**

This module is used to identify phosphorylation sites or kinases associated with clinical outcomes of patients.

Parameters Setting

pvalue adjust method: BH pvalue threshold: 0.01

Analysis

Feature Selection

Select a feature: gi|110556636\_TNKS1BP1\_S872

"high" colour #3300CC "low" colour #CC3300

Plot

Show 10 entries

Feature	HR	lower_95	upper_95	pvalue	adj_pvalue
gi 55956788_NCL_S67	1.3259	0.8589	2.0467	0.2028	0.937276345933562
gi 156523970_AHSG_S138	1.2348	0.9708	1.5708	0.0857	0.822954794520548
gi 56118310_NUCKS1_S214	1.086	0.5269	2.2386	0.8231	0.994815689655172
gi 56118310_NUCKS1_S181	1.1024	0.3474	3.4982	0.8686	1
gi 5803165_SEC61B_S17	1.1138	0.9135	1.358	0.2864	0.937276345933562
gi 24234699_KRT19_S35	0.887	0.427	1.8423	0.7478	0.98558386396396
gi 61743954_AHNAK_S5841	1.2589	0.9846	1.6098	0.0664	0.791298305084746
gi 45359846_G3BP2_T227	0.4852	0.1615	1.4583	0.1978	0.937276345933562
gi 42542379_SRMM1_S450	0.6556	0.3289	1.3067	0.2302	0.937276345933562
gi 4503423_DUT_S11	0.5296	0.1759	1.5943	0.2583	0.937276345933562

Showing 1 to 10 of 3,505 entries

gi|110556636\_TNKS1BP1\_S872

Distribution

Maximally Selected Rank Statistics

Cutoff pt: 1.62

Survival probability

p < 0.0001

Number at risk

Strata	Group-High	Group-Low
0	12	27
500	6	24
1000	3	18
1500	0	3
2000	0	0

## Preprocessing for Maxquant data

### Import MaxQuant data

#### How to import your MaxQuant data

1. Go to the 'Preprocessing' tab.
2. Choose 'Maxquant' to start with data from Maxquant.

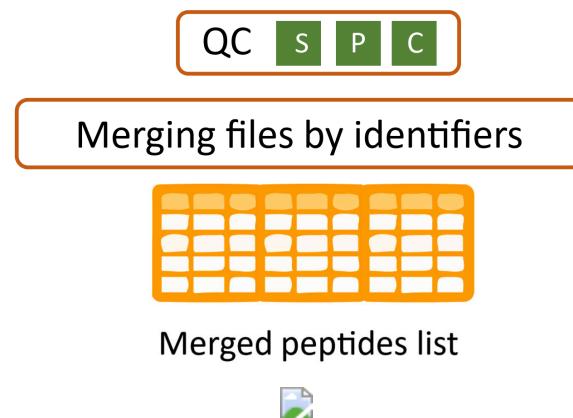
3. Click 'Browse' to upload phosphoproteomics experimental design file in .txt format, and phospho (STY)Sites.txt. Proteomics experimental design file is optional.
4. Uploaded data will be shown in the 'Data Overview' secondary tab.
5. You can also choose 'load example data' to use example files.

Experiment_Code	Group	Description
1 Exp027015	0	ctr_0h_R2_IMAC_1.raw
2 Exp027016	0	ctr_0h_R2_IMAC_2.raw
3 Exp027017	0	ctr_0h_R2_IMAC_3.raw
4 Exp027031	6	PLX_6h_R2_IMAC_1.raw
5 Exp027032	6	PLX_6h_R2_IMAC_2.raw
6 Exp027033	6	PLX_6h_R2_IMAC_3.raw
7 Exp027046	48	PLX_48h_R2_IMAC_1.raw
8 Exp027047	48	PLX_48h_R2_IMAC_2.raw
9 Exp027048	48	PLX_48h_R2_IMAC_3.raw

## Quality control and merging

### Function

Generate merged phosphoproteomics data frame based on peptides files.



### How to get analysis results

1. Go to the 'Preprocessing' secondary tab.
2. Modify the parameters in Step1 according to your needs.
3. Click the running button in Step1 and the file will appear on the right.

The screenshot shows the PhosMap preprocessing interface on a Mac OS X system. The main window has a title bar "bio-inf.shinyapps.io" and a menu bar with Home, Import Data, Analysis, Preprocessing (selected), Tutorial, FAQ, Download, and Help.

**Upload Data Panel:** Contains sections for "Load example data" (MaxQuant, Firmiana), "Experimental design file" (Phospho (STY)Sites.txt), "Proteomics data[optional]" (Proteomics experimental design file, proteinGroups.txt), and "QC result" (Step 1, Step 2, Step 3).

**Data Overview Panel:** Shows tabs for "Data Overview" (selected) and "Preprocessing". The "Data Overview" tab displays a table of QC results with columns: ID, Exp027015, Exp027016, Exp027017. The table lists 10 entries from a total of 4,862.

**QC result Panel:** Shows a table titled "QC result" with columns: ID, Exp027015, Exp027016, Exp027017. The table lists 10 entries from a total of 4,862. The table includes a search bar and navigation links for previous and next pages.

ID	Exp027015	Exp027016	Exp027017
A0AVK6_E2F8_S102	E2F8_S102	0	15484000
A0AVK6_E2F8_S71	E2F8_S71	0	0
A0FGR8_ESYT2_S743	ESYT2_S743	0	0
A0JLT2_MED19_S226	MED19_S226	0	96244000
A0JNW5_UHRF1BP1L_S989	UHRF1BP1L_S989	24650000	43735000
A0MZ66_KIAA1598_S506	KIAA1598_S506	36034000	24778000
A0MZ66_KIAA1598_S494	KIAA1598_S494	0	28114000
A1KXE4_FAM168B_S6	FAM168B_S6	8743300	0
A1L170_C1orf226_S223	C1orf226_S223	0	0
A1L390_PLEKHG3_S76	PLEKHG3_S76	31566000	30739000

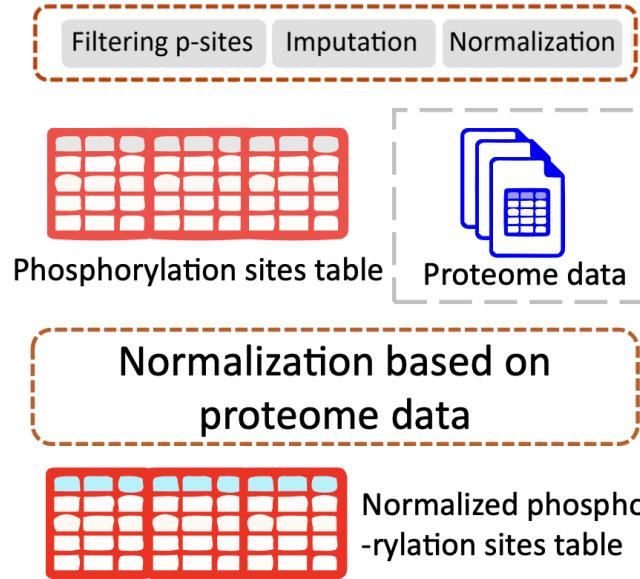
## Interpretation of analysis results

We performed quality control for identified phosphopeptides using PhosMap, those phosphopeptides that met 1% FDR at peptide level and had ion score greater than 20 and the highest confidence probabilities of p-sites computed by Mascot, were kept. We merged phosphopeptides list with quantitative value from all experiments to generate a matrix for analysis.

## Data normalization

### Function

PhosMap provides two kinds of normalizations, a total sum scaling normalization and normalizing phosphoproteomics data based on proteomics data.



## How to get analysis results

1. Go to the 'Preprocessing' secondary tab.
2. Modify the parameters according to your needs.
3. Click the running button in the Step2 and the normalized data of p-sites based on a total sum scaling will appear on the right.
4. Click the running button in Step3 and normalized data of p-sites based on proteomics data will appear on the right.

The screenshot shows the 'Preprocessing' tab of the bio-inf.shinyapps.io application. The interface is divided into several sections:

- Upload Data:** Includes options to 'Load example data' (MaxQuant or Firmiana), 'Experimental design file', 'Phospho (STY)Sites.txt', and 'Proteomics data[optional]' (including 'Proteomics experimental design file' and 'proteinGroups.txt').
- Data Overview:** Shows the current status of the upload.
- Step1: Quality Control:** Contains fields for 'minimum score' (40), 'minimum localization probability' (0.75), and 'minimum detection frequency' (1).
- Step2: Normalization & Filtering:** Includes 'normalization method' (global), 'imputing jointly' (checkbox), 'imputation method' (minimum/10), and a 'top' filter (100).
- Step3: Normalization based on proteomics data:** Includes 'With proteomics data' checkbox, 'Proteomics data preprocessing parameters' (IBAQ), 'minimum unique peptide' (1), 'minimum detection frequency' (1), 'normalization method' (global), 'imputation method' (minimum/10), and a 'control label' field (0).
- Result:** Displays a table titled 'Phosphorylation data frame:' showing 10 entries. The columns include Position, AA\_in\_protein, Sequence, and others. The first few rows of data are:

Position	AA_in_protein	Sequence
P09651_HNRNPA1_S6	S6	SEsPKEPEQLRK
P08238_HSP90AB1_S226	S226	EKEIsDDEAEEEK
P35579_MYH9_S1943	S1943	GAGDGsDEEVDGKADGAEAKP
Q9UHD8_SEPT9_S82	S82	HVDLSLQRSPK
P19338_NCL_S67	S67	KVVvsPTK
Q9UHD8_SEPT9_S85	S85	HVDLSLSRSPK
P27824_CANX_S583	S583	AEEDELNRsPRNR
Q01518_CAP1_T307	T307	PFSAPKPKQSPSPK
Q15019_SEPT2_S218	S218	IYHLPLDAEsDEDEDFK
P02545_LMNA_S392	S392	LSPsPTSQR

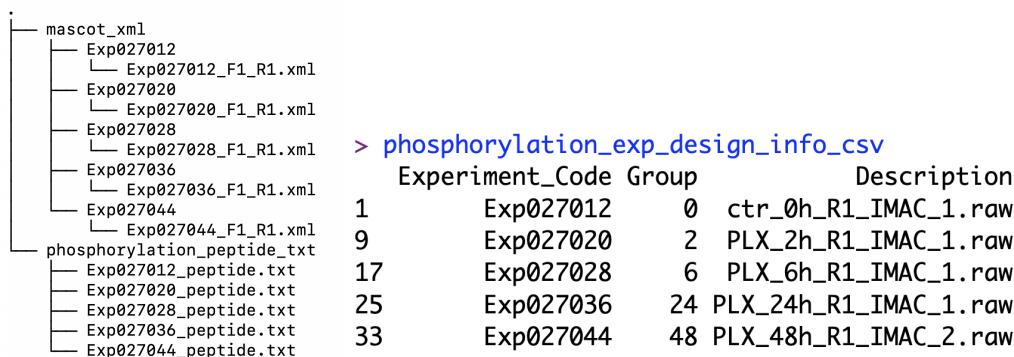
The screenshot shows the 'Preprocessing' tab of the bio-inf.shinyapps.io interface. On the left, there's a sidebar for 'Upload Data' with options like 'Load example data', 'MaxQuant', and 'Firmiana'. Below it are sections for 'Experimental design file', 'Phospho (STY)Sites.txt', 'Proteomics data[optional]', 'Proteomics experimental design file', and 'proteinGroups.txt'. The main area has tabs for 'Data Overview' and 'Preprocessing'. Under 'Preprocessing', there are three steps: 'Step1: Quality Control', 'Step2: Normalization & Imputation Filtering', and 'Step3: Normalization based on proteomics data'. Step3 includes fields for 'minimum score', 'minimum localization prob', 'minimum detection frequency', 'normalization method' (set to 'global'), 'imputing jointly', 'imputation method' (set to 'minimum/10'), 'top' (set to 100), and 'control label'. To the right, there are two tables: 'Phosphorylation data frame' showing 10 entries with columns for Position, AA\_in\_protein, and Sequence; and 'Proteomics data frame' showing 10 entries with columns for Symbol, Exp026982, Exp026983, and Exp026995.

## Preprocessing for Firmiana data

### Import Firmiana data

#### How to import your Firmiana data

1. Go to the 'Preprocessing' tab.
2. Choose 'Firmiana' to start with data from Firmiana.
3. Click 'Browse' to upload phosphoproteomics experimental design file in .txt format.
4. Zip your Mascot xml files and Phosphoproteomics peptide files, and then upload. The folder tree is shown below. File names of Mascot xml files and Phosphoproteomics peptide files must be consistent with 'Experiment\_Code' of phosphoproteomics experimental design file.



6. Proteomics data is optional. Click 'Browse' to upload proteomics experimental design file in .txt format. Zip your Profiling\_gene\_txt and upload. The folder tree is shown below. File names of Profiling\_gene\_txt must be consistent with 'Experiment\_Code' of proteomics experimental design file.

> profiling_exp_design_info				
Profiling_gene_txt	Experiment_Code	Group	Description	
Exp026921_gene.txt	1	Exp026921	0	ctr_0h_R1_injection_2.raw
Exp026986_gene.txt	2	Exp026986	2	PLX_2h_R1_injection_1.raw
Exp026993_gene.txt	3	Exp026993	6	PLX_6h_R1_injection_1.raw
Exp026999_gene.txt	4	Exp026999	24	PLX_24h_R1_injection_1.raw
Exp027006_gene.txt	5	Exp027006	48	PLX_48h_R1_injection_3.raw

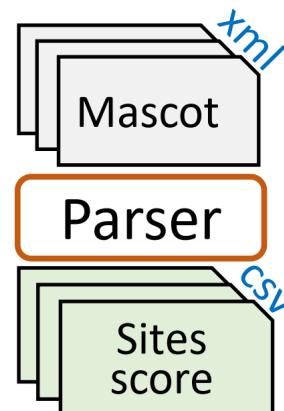
7. Uploaded data will be shown in the 'Data Overview' secondary tab.

8. You can also choose 'load example data' to use example files.

## Parser

### Function

If you start with .xml files from mascot results, you can run this button to parser them to sites score files, based on which .csv files of phosphorylation sites with confidence score will be generated.



### How to get analysis results

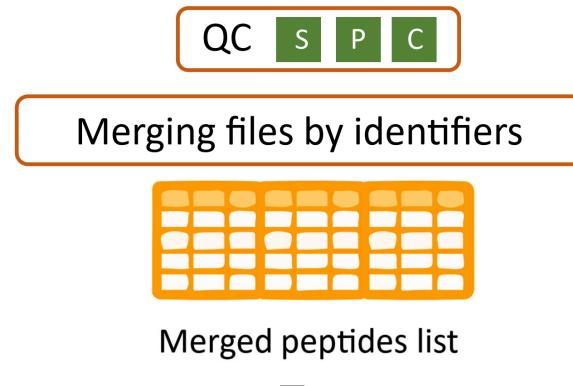
1. Go to the 'Preprocessing' secondary tab.
2. Click the running button in Step1 and the file will appear on the right.

pep_seq	pep_var_mod_conf
1 ITLPVDFVTADKFDENAK	
2 AATCKVSSINQETIQTYCVEDTPICFSR	
3 NQTAEEKEEFEHQKQ	
4 DAVITVPVFNFNQAERR	
5 ALLTLADGRRDESILK	
6 GFTGIDSEYEKPEAPELVLK	
7 EILEQQGLFSK	
8 PAYQRQPENALK	
9 VEEEIQTLSQVLAAK	
10 YEVFRTEEEEK	

## Quality control and merging

### Function

Generate merged phosphoproteomics data frame based on peptides files.



### How to get analysis results

1. Go to the 'Preprocessing' secondary tab.
2. Modify the parameters according to your needs.
3. Click the running button in Step2 and the file will appear on the right.

The screenshot shows the PhosMap software interface. The top navigation bar includes Home, Import Data, Analysis, Preprocessing (selected), Tutorial, FAQ, Download, and Help. The main area is divided into several tabs: Data Overview, Preprocessing, Step1: Parser, Step2: Quality Control & Merging, Step3: Mapping, Step4: Filtering & Normalization & Imputation, and Step5: Normalization based on proteomics data. The Step4 tab is active, showing parameters for minimum detection frequency (1), normalization method (global), imputing jointly, imputation method (minimum'10), and top (100). The Step5 tab shows a checkbox for 'With proteomics data' and a dropdown for Proteomics data preprocessing parameters with US cutoff set to 1. On the right, a results table titled 'Peptide data frame through phosphorylation sites quality control' lists 10 entries, each with ID, sequence, and various statistics like count and length.

## Interpretation of analysis results

We performed quality control for identified phosphopeptides using PhosMap, those phosphopeptides that met 1% FDR at peptide level and had ion score greater than 20 and the highest confidence probabilities of p-sites computed by Mascot, were kept. We merged phosphopeptides list with quantitative value from all experiments to generate a matrix for analysis.

## Mapping p-sites to protein

### Function

Mapping protein gi number to gene symbol and outputting expression profile matrix with gene symbol. Constructing the data frame with unique phosphorylation site for each protein sequence.

### Mapping p-sites to protein



Raw phosphorylation sites table

## How to get analysis results

1. Go to the 'Preprocessing' secondary tab.
2. Modify the parameters according to your needs.
3. Click the running button in Step3 and the file will appear on the right.

The screenshot shows the PhosMap software interface in a web browser. The main menu at the top includes Home, Import Data, Analysis, Preprocessing (selected), Tutorial, FAQ, and Download. The Preprocessing tab is active, showing a flow of five steps: Step1: Parser, Step2: Quality Control & Merging, Step3: Mapping, Step4: Filtering & Normalization & Imputation, and Step5: Normalization based on proteomics data. The Step4 section is currently selected. On the left, there's a 'Upload Data' panel with sections for Experimental design file, Mascot xml file, Phosphoproteomics peptide file, and Proteomics data[optional]. The right side displays a table of results with columns for AA\_in\_protein, AA\_in\_peptide, and Sequence.

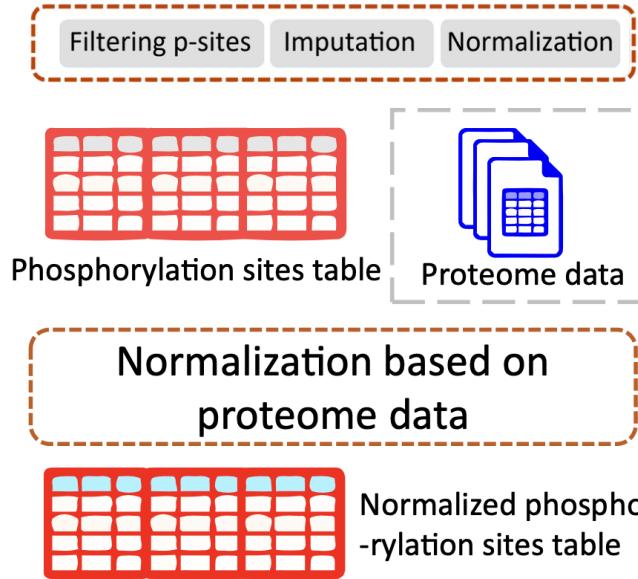
## Interpretation of analysis results

Combining the phosphopeptides sequence, modification position, attached protein ID and the built-in human protein reference database of PhosMap, all p-sites were mapped to the corresponding protein sequence and represented by unique p-sites identifier (upsID) that consisted of a protein GI number/accession, gene symbol and location of the p-site in the protein sequence. In addition, the matched proteome data with phosphoproteome were collected at each time point in Ressa, et al. study. Finally, 3,649 unique p-sites were obtained and their quantitative values were normalized by matched protein profiling data using PhosMap.

## Data normalization

### Function

PhosMap provides two kinds of normalizations, a total sum scaling normalization and normalizing phosphoproteomics data based on proteomics data.



## How to get analysis results

1. Go to the 'Preprocessing' secondary tab.
2. Modify the parameters according to your needs.
3. Click the running button in the Step4 and the normalized data of p-sites based on a total sum scaling will appear on the right.
4. Click the running button in Step5 and normalized data of p-sites based on proteomics data will appear on the right.

The screenshot shows the PhosphoSitePlus software interface in a web browser. The top navigation bar includes links for Home, Import Data, Analysis, Preprocessing (which is the active tab), Tutorial, FAQ, and Download. The main workspace is divided into several sections:

- Upload Data:** Includes fields for Experimental design file (Exp027015\_F1\_R1), Mascot xml file (Exp027015\_peptide), Phosphoproteomics peptide file (Exp027015\_peptide), and Proteomics data[optional] (Exp026982\_gene).
- Data Overview:** Shows a preview of the experimental design file.
- Processing Steps:**
  - Step1: Parser:** no parameter
  - Step2: Quality Control & Merging:** minimum score: 20, minimum FDR: 0.01
  - Step3: Mapping:** species: human, id type: RefSeq\_Protein\_GI, fasta type: refseq
  - Step4: Filtering & Normalization & Imputation:** minimum detection frequency: 1, normalization method: global, imputing jointly, imputation method: minimum/10, top: 100
  - Step5: Normalization based on proteomics data:** With proteomics data checked, US cutoff: 1
- Result:** Displays the 'Phosphorylation data frame' table.

The 'Phosphorylation data frame' table has columns for Position, AA\_in\_protein, and Sequence. It shows 10 entries of phosphorylation sites across different proteins (e.g., NCL, AHSG, KRT19, G3BP2, DUT, AHNK) at specific positions (e.g., S67, S138, S214, S181, S35, T227, S11, S5731, S5841, S552) with their corresponding sequences.

The screenshot shows a web-based bioinformatics tool with a light gray background. In the center, a white modal window titled 'All Done' contains a green checkmark icon and the text 'You can now download the 'Phosphorylation data frame' for further analysis.' Below this is a blue 'OK' button. Behind the modal, the main interface has several sections:

- 3. Phosphoproteomics peptide file:** A dropdown menu showing 'Exp027015\_peptide'.
- 4. Proteomics data[optional]**
- 4.1 Proteomics experimental design file:** A dropdown menu showing 'Exp view'.
- 4.2 Profiling file:** A dropdown menu showing 'Exp026982\_gene'.
- Step3: Mapping** section with fields for:
  - species:** human
  - id type:** RefSeq\_Protein\_GI
  - fasta type:** refseq
- imputation method:** dropdown set to 'minimum/10'.
- control label:** dropdown set to '0'.

To the right of the main interface, there is a table titled 'Proteomics data frame:' showing 10 entries. The columns are 'Symbol', 'Exp026982', 'Exp026983', and 'Exp026995'. The data includes various protein symbols and their corresponding values across three replicates. Below this table is another table showing 10 entries from a different dataset, with columns 'Symbol', 'Exp026982', 'Exp026983', and 'Exp026995'.

## References

1. Feng, J., Ding, C., Qiu, N., Ni, X., Zhan, D., Liu, W., Xia, X., Li, P., Lu, B. and Zhao, Q. (2017) Firmiana: towards a one-stop proteomic cloud platform for data processing and analysis. *Nature biotechnology*, 35, 409-412.
2. Ressa, A., Bosdriesz, E., De Ligt, J., Mainardi, S., Maddalo, G., Prahallad, A., Jager, M., De La Fonteijne, L., Fitzpatrick, M. and Grotel, S. (2018) A system-wide approach to monitor responses to synergistic BRAF and EGFR inhibition in colorectal cancer cells. *Molecular & Cellular Proteomics*, 17, 1892-1908.