

Users Guide

We need to follow the corresponding procedure for multi-omics analysis of cancers. This document will introduce the use of the software through the example of bladder cancer.

The first step: Data Preparation

(1). Data preparation for TCGA dataset

Parameter Settings: Before data preparation, we should first write a mail to 'xlu.cpu@foxmail.com' (Name, Affiliation, purpose, etc.) to apply for a Username (E-mail Address) and a User Account. If we log in for the first time using the applied username and user account, "Whether log in for the first time using this user account" should choose "Yes", otherwise it should choose "No" after inputting "Username (E-mail Address)" and "User Account". Then, we fill in the name of the tumor in TCGA database ("Input Cancer Types From TCGA" enters "BLCA"). After that, we should specify a project name ("Project Name" entered "Bladder_Cancer") and choose at least 2 kinds of omics data types ("Multi-omics Types" chooses all omics data types except "radiomics" by default). Then, "Data Sources" chooses "External resources" and we use integrated dataset for analysis directly ("Use integrated datasets or not" chooses "Yes") due to many adjustments on clinical data. We click the "Browse" button to upload the

integrated dataset, and “External validation or not” should choose “Yes” for validation. We should prepare the integrated dataset by ourselves and save it as a “.RData” file. The integrated dataset is a data list containing information which should be named as “mRNA.expr”, “lncRNA.expr”, “meth.beta”, “cna”, “mut.staats”, “radiomics”, “count”, “tpm”/("fpkm_mrna"+“fpkm_Incrna”), “maf”, “segment”, “clin.info” if exists. Finally, we click the “Create” button to create an account, and then click the “Process” button to move the integrated dataset to the specified file path. If you want to continue the analysis using previous account, “Continue analysis on previous results or not” chooses “Yes” and “Please input the previously used user account” enters the previous account ID. Then, you should indicate the following parameters in turn: “The Number of Cancer Types”, “Input Cancer Types From TCGA”, “Project Name”, “Multi-omics Types” for TCGA dataset, “Multi-omics Types” for validation dataset, “Clustering algorithms” for “Consensus Clustering” in “GET Module” step. If you utilize user-defined clustering number, “The number of clusters” should choose “User defined” and you should indicate the clustering number for “User defined cluster number”.

Basic Settings

Continue analysis on previous results or not

Yes No

The Number of Cancer Types

1

Input Cancer Types From TCGA

	Type 1
Cancer Type	BLCA

Project Name

Bladder_Cancer

Now let's start to prepare the tcga datasets first.

Multi-omics Types (Choose at least two types)

- mRNA
- lncRNA
- DNA methylation
- copy number alterations
- binary somatic mutation
- radiomics

Data Sources

Internal resources External resources

Use integrated datasets (.RData) or not

Yes No

The approach to obtain integrated datasets

Download from specified uris Upload manually

Upload the integrated datasets (.RData)

Browse... blca_tcga.RData

Upload complete

External validation or not

Yes No

Create

Process

Parameter settings for TCGA dataset preparation

Result Display:

The software will give feedback to the users after finishing the process of preparing TCGA dataset.



(2). Data preparation for validation dataset

Parameter Settings: Since two validation cohorts only contain mRNA as well as clinical and survival information, “Multi-omics Types” should only choose “mRNA”. Then, we should upload the integrated dataset directly (“Use integrated datasets or not” chooses “Yes”) because some personalized processing needs to be done on the validation dataset. We select “Upload manually” for “The approach to obtain integrated datasets” and click the “Browse” button to upload the integrated validation dataset. Finally, we click the “Create” button to create a folder for validation dataset preparation, and then click the “Process” button to move the integrated validation dataset to the specified file path. It should be noted that only one validation dataset can be prepared at a time. We must finish the analysis of the current validation dataset, obtain as well as download

and save the results before the preparation and analysis of another validation dataset. Otherwise, it will easily cause errors and confusion.

Validation Datasets Preparation

Now let's start to prepare the validation datasets which will be used in 'RUN Module' and 'COMP Module'.

Multi-omics Types

mRNA
 lncRNA
 DNA methylation
 copy number alterations
 binary somatic mutation
 radiomics

Use integrated datasets (.RData) or not

Yes No

The approach to obtain integrated datasets

Download from specified urls Upload manually

Upload the integrated datasets (.RData)

blca.validation.affy.RData
Upload complete

Parameter settings for validation dataset preparation

Result Display:

The software will also give feedback to the users after finishing the process of preparing validation dataset.

Data Preparation (TCGA) Data Integration (TCGA) Data Preparation (Validation) Data Integration (Validation)

Now a folder storing validation datasets has been created and then let's start preparing validation datasets.

The validation integrated dataset (.RData) has been uploaded in the corresponding folder. Now let's start the first step of MOVICS--GET Module!

Result display for validation dataset preparation

The second step: GET Module

(1). Get Elites for TCGA dataset and validation dataset

Parameter Settings: “Module switching options” on the upper left of the software chooses “GET Module”, and then the “Steps” chooses “Get Elites”. After that, we should indicate the omics data types for “Get Elites” (choose “Yes” under the corresponding parameters) and then set the parameters for each omics data in turn. For continuous omics data including mRNA, lncRNA, DNA methylation, copy number alterations and radiomics, we utilize “mad” method (“Choose a Get Elites method” chooses “mad”) to extract top 1500 features (“Choose a filtration method” chooses “elite.num”, and then “An integer cutoff of exact number for selecting top elites” indicates 1500) for each omics data according to mad values. In particular, for mRNA and lncRNA, we also perform “log₂” transformation (“Perform log2 transformation for data before calculating statistics or not” chooses “Yes”) and set a numeric cutoff to remove features with low expression (“Set a numeric cutoff for removing low expression features or not” chooses “Yes”, and then “The cutoff for removing low expression features” indicates 0.1). For binary omics data containing somatic mutation, we extract the features whose mutation frequency divided by sample size is higher than 0.03 (“Choose a filtration method for binary somatic mutation dataset” chooses “elite.pct”, and then “A numeric cutoff of 'mutation / sample' frequency for selecting

elites” indicates 0.03). Additionally, “Use survival information or not” chooses “No”, “NA value action” chooses “Remove directly”, “Scaling the data after filtering or not” chooses “No”, and “Centering the data after filtering or not” chooses “No” for each omics data. Finally, we click the “Process” button to do “Get Elites”. If you want to do “Get Elites” on validation dataset, “Get elites on tcga datasets or validation datasets” should select “Validation”. Here we do not need to do “Get Elites” on validation dataset, thus “Get Elites for mRNA dataset or not” chooses “No”, and then we click the “Process” button.

The screenshot shows a user interface for configuring a 'Get Elites' step. At the top left, there is a sidebar titled 'Steps' containing the following options:

- Get Elites
- Get Clustering Number
- Consensus Clustering
- Silhouette
- Multi-omics Heatmaps

The main area is titled 'Get Elites' and contains the following instructions and configuration options:

In this step, we will filter out features that meet some stringent requirements as well as handle missing values. Now let's choose the datasets for 'Get Elites' first:

Get elites on tcga datasets or validation datasets

TCGA Validation

Get Elites for mRNA dataset or not

Yes No

Get Elites for lncRNA dataset or not

Yes No

Get Elites for DNA methylation dataset or not

Yes No

Get Elites for copy number alterations dataset or not

Yes No

Get Elites for binary somatic mutation dataset or not

Yes No

Get Elites settings for mRNA dataset

Use survival information or not

Yes No

NA value action

Remove directly KNN imputation No action

Perform log2 transformation for data before calculating statistics or not

Yes No

Set a numeric cutoff for removing low expression features or not

Yes No

The cutoff for removing low expression features

0.1

Choose a Get Elites method for mRNA dataset

mad sd pca

Choose a filtration method for mRNA dataset

elite.num elite.pct

An integer cutoff of exact number for selecting top elites

1500

Scaling the data after filtering or not

Yes No

Centering the data after filtering or not

Yes No

Get Elites settings for lncRNA dataset

Use survival information or not

Yes No

NA value action

Remove directly KNN imputation No action

Perform log2 transformation for data before calculating statistics or not

Yes No

Set a numeric cutoff for removing low expression features or not

Yes No

The cutoff for removing low expression features

0.1

Choose a Get Elites method for lncRNA dataset

mad sd pca

Choose a filtration method for lncRNA dataset

elite.num elite.pct

An integer cutoff of exact number for selecting top elites

1500

Scaling the data after filtering or not

Yes No

Centering the data after filtering or not

Yes No

Get Elites settings for DNA methylation dataset

Use survival information or not
 Yes No

NA value action
 Remove directly KNN imputation No action

Perform log2 transformation for data before calculating statistics or not
 Yes No

Set a numeric cutoff for removing low expression features or not
 Yes No

Choose a Get Elites method for DNA methylation dataset
 mad sd pca

Choose a filtration method for DNA methylation dataset
 elite.num elite.pct

An integer cutoff of exact number for selecting top elites
1500

Scaling the data after filtering or not
 Yes No

Centering the data after filtering or not
 Yes No

Get Elites settings for copy number alterations dataset

Use survival information or not
 Yes No

NA value action
 Remove directly KNN imputation No action

Perform log2 transformation for data before calculating statistics or not
 Yes No

Set a numeric cutoff for removing low expression features or not
 Yes No

Choose a Get Elites method for copy number alterations dataset
 mad sd pca

Choose a filtration method for copy number alterations dataset
 elite.num elite.pct

An integer cutoff of exact number for selecting top elites
1500

Scaling the data after filtering or not
 Yes No

Centering the data after filtering or not
 Yes No

Get Elites settings for binary somatic mutation dataset

Use survival information or not

Yes No

NA value action

Remove directly KNN imputation No action

Perform log2 transformation for data before calculating statistics or not

Yes No

Choose a filtration method for binary somatic mutation dataset

elite.num elite.pct

A numeric cutoff of 'mutation / sample' frequency for selecting elites

0.03

Scaling the data after filtering or not

Yes No

Centering the data after filtering or not

Yes No

Process Get Elites

Now click the 'Process' button below to process 'Get Elites' based on the settings above, and then integrate datasets for following steps.

Process

Parameter settings for “Get Elites” on TCGA dataset

Data Preparation GET Module COMP Module RUN Module Users

Steps

- Get Elites
- Get Clustering Number
- Consensus Clustering
- Silhouette
- Multi-omics Heatmaps

Get Elites

In this step, we will filter out features that meet some stringent requirements as well as handle missing values. Now let's choose the datasets for 'Get Elites' first:

Get elites on tcga datasets or validation datasets

TCGA Validation

Get Elites for mRNA dataset or not

Yes No

Process Get Elites

Now click the 'Process' button below to process 'Get Elites' based on the settings above, and then integrate datasets for following steps.

Process

Parameter settings for “Get Elites” on validation dataset

Result Display:

The software will give feedback to the users after finishing the process of “Get Elites” on TCGA dataset or validation dataset.



All omics datasets you specified have been processed 'Get Elites' on tcga datasets to extract important features, and you can download and check the obtained files you want below. Now let's start to get the clustering number using elites from tcga datasets.

[Download RData file of the integrated data after 'Get Elites'](#)

Result display for “Get Elites” on TCGA dataset



All omics datasets you specified have been processed 'Get Elites' on validation datasets to extract important features, and you can download and check the obtained files you want below. Now let's start to get the clustering number using elites from tcga datasets.

[Download RData file of the integrated data after 'Get Elites'](#)

Result display for “Get Elites” on validation dataset

(2). Get Clustering Number for TCGA dataset

Parameter Settings: The “Steps” chooses “Get Clustering Number”, and then we indicate the range of clustering number (“Minimum” indicates 2 and “Maximum” indicates 6). After that, both “Centering the data or not” and “Scaling the data or not” choose “Yes”. Finally, we enter the “Figure Name”, and click the “Process” button to determine the optimal clustering number.

Steps

- Get Elites
- Get Clustering Number
- Consensus Clustering
- Silhouette
- Multi-omics Heatmaps

Get Clustering Number

This step aims to search the optimal number for multi-omics integrative clustering determined by 'clustering prediction index' (CPI) and 'Gap-statistics' (Gapk).

Please indicate the range of clustering number first:

Minimum

Maximum

Centering the data or not

Yes No

Scaling the data or not

Yes No

Figure Name

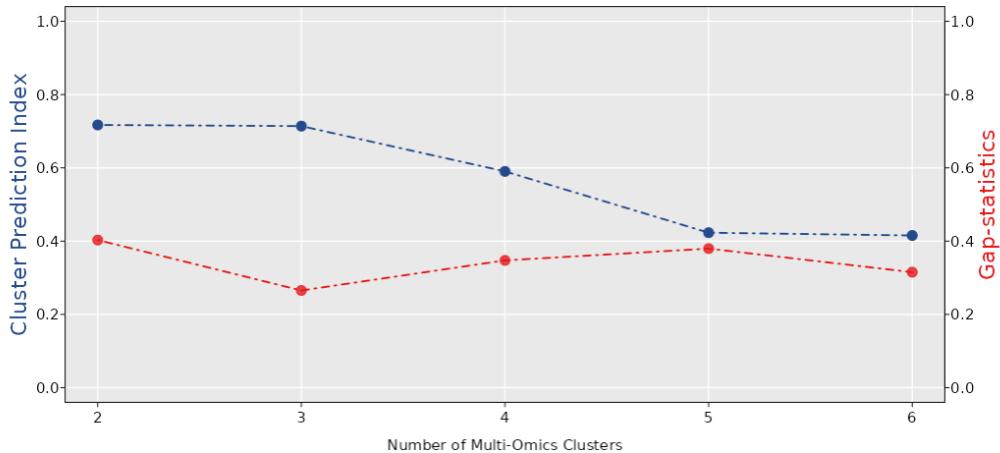
Process

Parameter settings for “Get Clustering Number” on TCGA dataset

Result Display:

The software will plot according to the values of CPI and Gaps-statistics and suggest the optimal clustering number when the sum of CPI and Gaps-statistics is the largest. However, we should also consider the prior information when determining the optimal clustering number, which will be used for the following “Consensus Clustering”. Here we combine prior information to determine “4” as the optimal clustering number.

[Download pdf of the figure](#)



The imputed optimal cluster number is 2 arbitrarily, but it would be better referring to other priori knowledge.

Now we have determined the optimal number of clustering, you can download and check the figure as well as the '.RData' file. Now let's start clustering based on all omics datasets.

[Download RData file of the results](#)

Result display for “Get Clustering Number” on TCGA dataset

(3). Consensus Clustering for TCGA dataset

Parameter Settings: The “Steps” chooses “Consensus Clustering”, and then we choose the clustering algorithms (10 algorithms are chosen by default). The parameters for each clustering algorithm are set by default, but you can set the parameters concretely if you just indicate one clustering algorithm. Since we determine the optimal clustering number based on the prior information, therefore the “The number of clusters” chooses “User defined” and “User defined cluster number” indicates 4. After that, we click the “Process” button to do “Consensus Clustering”. After we get the results of “Consensus Clustering”, we set the

corresponding parameters to plot the consensus heatmap. “Distance measurement for hierarchical clustering” indicates “euclidean”, “Clustering method for hierarchical clustering” indicates “ward.D2”, both “Heatmap mapping color settings” and “Clustering subtypes color settings” choose “System default”, and “Show sample ID or not” chooses “No”. Besides, we indicate the “Figure Name”, “The width of output figure” and “The height of output figure”. Finally, we click the “Process” button to plot the consensus heatmap.

The image consists of three vertically stacked screenshots of a software application's user interface, likely a web-based tool for multi-omics clustering analysis.

Screenshot 1: Step Selection

A sidebar titled "Steps" contains the following options:

- Get Elites
- Get Clustering Number
- Consensus Clustering
- Silhouette
- Multi-omics Heatmaps

Screenshot 2: Algorithm Selection - Step 1

A section titled "Consensus Clustering" contains the following text:

This step aims to perform multi-omics integrative clustering by specifying one or more algorithms at once.

Now let's choose clustering algorithms at first. If you want to get consensus results from different algorithms, please choose at least two algorithms!

Clustering algorithms (Choose at least two types if you want to get consensus results)

A list of checked checkboxes represents selected algorithms:

- iClusterBayes
- SNF
- PINSPlus
- NEMO
- COCA
- LRAcluster
- ConsensusClustering
- IntNMF
- CIMLR
- MoCluster

Screenshot 3: Algorithm Selection - Step 2

A section titled "Consensus Clustering" contains the following text:

You choose more than 1 algorithm and all of them shall be run with parameters by default.

The number of clusters

System optimal User defined

User defined cluster number

A text input field contains the value "4".

A blue "Process" button is located at the bottom of this panel.

Parameter settings for “Consensus Clustering” on TCGA dataset

Consensus Heatmap

Distance measurement for hierarchical clustering

Clustering method for hierarchical clustering

Heatmap mapping color settings

System default User defined

Clustering subtypes color settings

System default User defined

Show sample ID or not

Yes No

Figure Name

The width of output figure

The height of output figure

Process

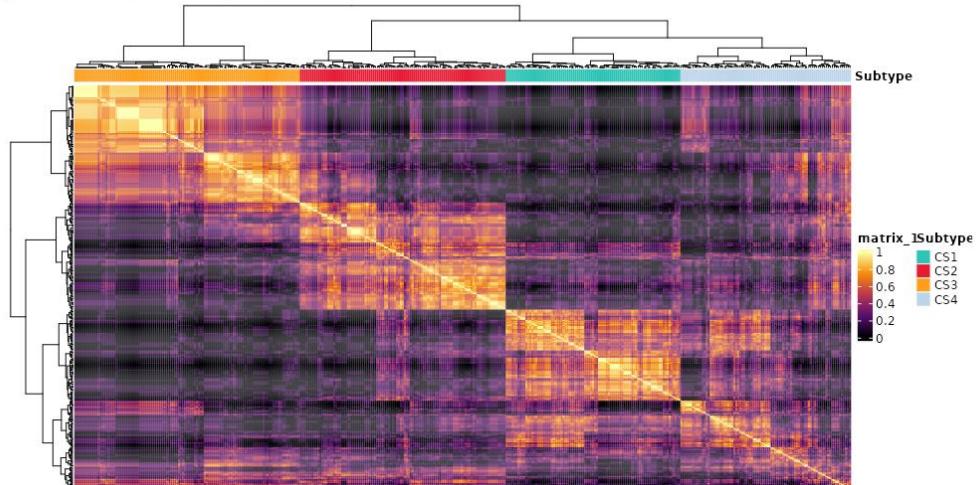
Parameter settings for consensus heatmap

Result Display:

The software will give feedback to the users if the “Consensus Clustering” is finished. Moreover, a consensus heatmap will be also displayed to reflect the quality of consensus clustering results.

[Get Elites](#)[Get Clustering Number](#)[Consensus Clustering](#)[Silhouette](#)[Multi-omics Heatmaps](#)

The process of clustering using multiple specified algorithms has been finished, you can download and check '.RData' file below. Now let's keep on the step of 'Consensus Heatmap'.

[Download RData file of the results](#)[Download pdf of the figure](#)

Result display for “Consensus Clustering” on TCGA dataset

(4). Silhouette for TCGA dataset

Parameter Settings: The “Steps” chooses “Silhouette”, and “Colors for annotating each cluster” chooses “System default”. Then, we indicate the “Figure Name”, “The width of output figure” and “The height of output figure”. Finally, we click the “Process” button to generate the Silhouette plot.

Steps

- Get Elites
- Get Clustering Number
- Consensus Clustering
- Silhouette
- Multi-omics Heatmaps

Silhouette

This step aims to visualize silhouette information from consensus clustering.

Colors for annotating each cluster

System default User defined

Figure Name

The width of output figure

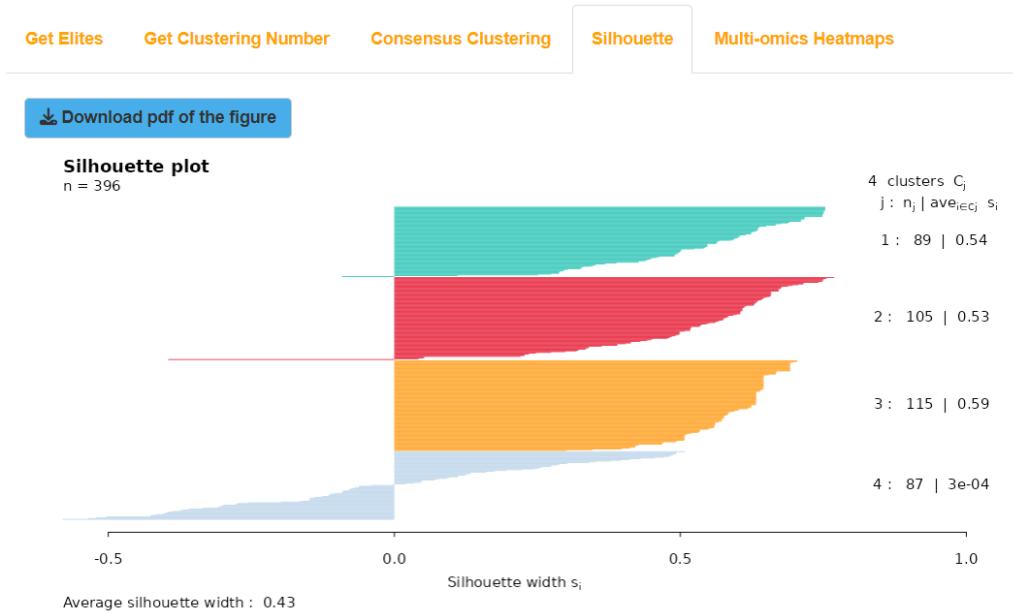
The height of output figure

Process

Parameter settings for “Silhouette” on TCGA dataset

Result Display:

The software will generate a Silhouette plot to evaluate the similarity between samples in each subtype derived from clustering results through Silhouette Coefficient, which can be used to evaluate the clustering results. Silhouette Coefficient ranges from -1 to 1, and the larger value indicates better clustering results.



The process of visualizing silhouette information from consensus clustering has been finished, you can download and check the figure. Now let's turn to the step of 'Multi-omics Heatmaps'.

Result display for “Silhouette” on TCGA dataset

(5). Multi-omics Heatmaps for TCGA dataset

Parameter Settings: There are many parameters that need to be set in this step. The “Steps” chooses “Multi-omics Heatmaps”, and we should specify whether to use the results of consensus clustering (“Heatmaps for consensus clustering results of multiple clustering algorithms or not” chooses “Yes” by default. Otherwise, we choose “No” and then indicate the clustering results derived from specified algorithm). Then, we need to standardize each omics data (“Indicate if each omic data should be centered”, “Indicate if each omic data should be scaled” and “Assign truncating values for extreme values in continuous normalized multi-omics data” use default settings). After that, we start to set the

parameters of the multi-omics heatmap such as “Row title settings for each omic data”, “Legend title settings for each omic data”, “Show the sample names for columns at the bottom of heatmap or not”, “Colors for annotating each cluster at the top of heatmap”, “Colors for heatmap of each omic data”, “The width for each heatmap”, “The height for each heatmap” and “Figure Name” (All these parameters use default settings). In addition, we should indicate the parameters of “Click the boxes below to indicate whether show the feature names for rows of each omic data”, “Click the boxes below to indicate whether show dendrogram for rows of each omic data”, “Input distance method for clustering each omic data at feature dimension” and “Input clustering method for each omic data at feature dimension” for each omics data (All these parameters use default settings). In particular, clustering algorithms of iClusterBayes, CIMLR, MoCluster contain the results of feature screening, and the software provides the option to display the results. Additionally, clustering algorithms of COCA, LRAcluster, ConsensusClustering, and MoCluster, as well as consensus clustering will generate a hierarchical clustering diagram of samples, and the users can decide whether to display it or not. Here we choose “No” for “Show the dendrogram for columns at the top of heatmap or not”. In addition to annotation for sample through clustering results, “Multi-omics Heatmaps” also allows users to select specific clinical features for annotation, and we select “PAM_Subtype”, “oneNN_Subtype”,

“Lund_Subtype” and “TCGA_Subtype”. Finally, we click the “Process” button to plot the multi-omics heatmap.

Steps

- Get Elites
- Get Clustering Number
- Consensus Clustering
- Silhouette
- Multi-omics Heatmaps

Multi-omics Heatmaps (multiple algorithms)

This step aims to vertically concatenate multiple heatmap derived from each omics data combined with clustering results and other annotation information.

Heatmaps for consensus clustering results of multiple clustering algorithms or not

Yes No

First let's get standardized multi-omics data:

Indicate if each omic data should be centered

	mRNA	lncRNA	DNA methylation	copy number alterations
Center or not	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Indicate if each omic data should be scaled

	mRNA	lncRNA	DNA methylation	copy number alterations
Scale or not	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Assign truncating values for extreme values in continuous normalized multi-omics data (normalized values that exceed the truncating values will be replaced by truncating values, which is useful to map colors in heatmap)

	mRNA	lncRNA	DNA methylation	copy number alterations
Truncating value	2.00	2.00	2.00	2.00

Multi-omics Heatmaps (multiple algorithms)

Then let's set the parameters of the heatmap:

Row title settings for each omic data

System default User defined

Legend title settings for each omic data

System default User defined

Show the dendrogram for columns at the top of heatmap or not

Yes No

Show the sample names for columns at the bottom of heatmap or not

Yes No

Click the boxes below to indicate whether show the feature names for rows of each omic data

	mRNA	lncRNA	DNA methylation	copy number alterations	binary somatic mutation
Show rownames	<input type="checkbox"/>				

Input distance method for clustering each omic data at feature dimension

	mRNA	lncRNA	DNA methylation	copy number alterations	binary somatic mutation
Distance method	pearson	pearson	pearson	pearson	pearson

Input clustering method for each omic data at feature dimension

	mRNA	lncRNA	DNA methylation	copy number alterations	binary somatic mutation
Clustering method	ward.D	ward.D	ward.D	ward.D	ward.D

Click the boxes below to indicate whether show dendrogram for rows of each omic data

	mRNA	lncRNA	DNA methylation	copy number alterations	binary somatic mutation
Show row dendrogram	<input checked="" type="checkbox"/>				

Colors for annotating each cluster at the top of heatmap

System default User defined

Colors for heatmap of each omic data

System default User defined

Sample annotations from survival information for heatmap or not

Yes No

The number of sample annotations from survival information for heatmap

4

Input sample annotations from survival information for heatmap

First line: Please input the sample annotation variables from survival information

Second line: Please input 'Continuous' or 'Categorical' to indicate the type of each sample annotation variable

Last line: Please input the colors for each sample annotation variable (use hex color format, e.g. #000004FF and English semicolons should be used to separate the input colors)

Note1: If the sample annotation variable is continuous, the number of indicated colors should be equal to 3, which represents the minimum, median and maximum value of this variable)

Note2: If the sample annotation variable is categorical, the number of indicated colors should be equal to the number of categories for this variable)

	1	2	3	4
Sample annotation	PAM_Subty	oneNN_Sub	Lund_Subty	TCGA_Subt
Continuous or Categorical	Categorical	Categorical	Categorical	Categorical
Color settings	#2874C5;#E#2874C5;#C	#EABF00;#8	#2874C5;#8	

The width for each heatmap

6

The height for each heatmap

2

Figure Name

moheatmap

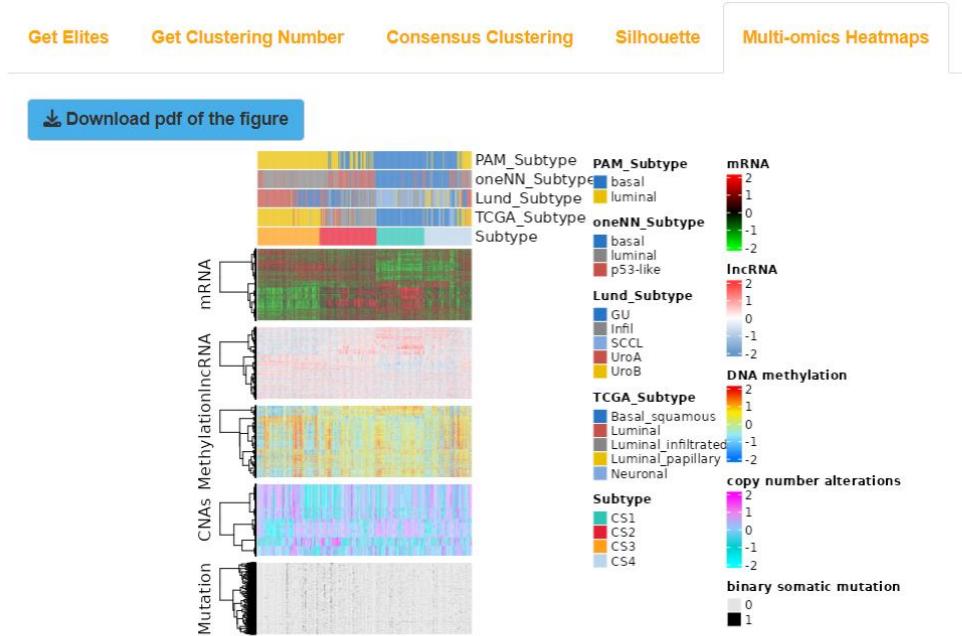
Process

Parameter settings for “Multi-omics Heatmaps” on TCGA dataset

Result Display:

The software will generate a multi-omics heatmap, and users can observe the expression differences of features in each omics data for each subtype. Because the generated multi-omics heatmap is relatively large, it may not be displayed clearly enough on the website. So, users can download the

heatmap, and then enlarge it for observation.



The process of getting multi-omics comprehensive heatmap has been finished, you can download and check the figure as well as the '.RData' file. Now all steps in 'GET Module' have been finished, let's start the second step of MOVICS--COMP Module!

[Download RData file of the results](#)

Result display for "Multi-omics Heatmaps" on TCGA dataset

The third step: COMP Module (TCGA dataset)

(1). Compare survival outcome for TCGA dataset

Parameter Settings: “Module switching options” on the upper left of the software chooses “COMP Module”, and then the “Steps” chooses “Compare survival outcome”, and “Compare survival outcome on TCGA datasets or Validation datasets” chooses “TCGA”. After that, “Format conversion of the survival time” chooses “Months” to convert data unit to month. Besides, we indicate the maximum value of the survival time as 10 years (“Setting for the x-axis cutoff for showing the maximal survival time”

chooses “System default”). Then, we will calculate the survival rate at 5-year and 10-year. “Estimate probability of surviving beyond a certain number of years” chooses “Yes”, and then we indicate 2 as “The number of years for surviving probability estimation” and enter 5 and 10 for “Input the year”. For the adjustment of the p value, the software provides 8 methods and we choose “none” (Method for adjusting p values). For the median survival time, we choose “Horizontal line” (The way for drawing a horizontal/vertical line at median survival). Additionally, “Setting for colors of clustering subtypes” and “Figure Name” use default settings. Finally, we click the “Process” button to compare survival outcomes among subtypes.

[Data Preparation](#) [GET Module](#) **COMP Module** [RUN Module](#) [Users Guide](#)

Steps

- Compare survival outcome
- Compare clinical features
- Compare mutational frequency
- Compare total mutation burden
- Compare fraction genome altered
- Compare drug sensitivity
- Compare agreement with other subtypes

Compare survival outcome

In this step, we will compare the prognosis of different subtypes based on the clustering results from 'GET Module' by Kaplan-Meier survival curve.

Pay attention: the format of survival time should be days and the values of survival status should be 0 or 1 (0: censoring; 1: event). Please make sure you provide the correct survival information first.

Compare survival outcome on TCGA datasets or Validation datasets

- TCGA
- Validation

Format conversion of the survival time

- Years
- Months
- No conversion

Setting for the x-axis cutoff for showing the maximal survival time

- System default
- User defined

Estimate probability of surviving beyond a certain number of years (Estimating x-year survival) or not

Yes No

The number of years for surviving probability estimation

2

Input the year

	1	2
Year	5	10

Setting for colors of clustering subtypes

System default User defined

Method for adjusting p values

- holm
- hochberg
- hommel
- bonferroni
- BH
- BY
- fdr
- none

The way for drawing a horizontal/vertical line at median survival

Horizontal line

Vertical line

Horizontal & Vertical

No lines

Figure Name

KAPLAN-MEIER_CURVE(TCGA)

Process

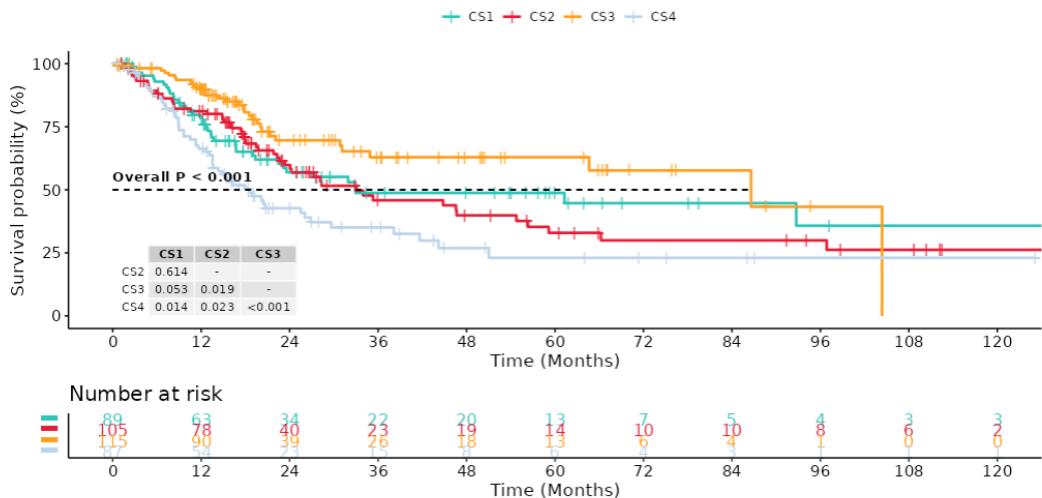
Parameter settings for “Compare survival outcome” on TCGA dataset

Result Display:

The software will generate Kaplan-Meier curve for each subtype derived from clustering results, and compare the survival differences among each subtype at the same time. “Overall P<0.05” reveals significant survival differences among each subtype.

Survival Clinical Features Mutational Frequency TMB FGA Drug Sensitivity Agreement

[Download pdf of the figure](#)



The process of comparing survival outcome on tcga datasets has been finished, you can download and check the figure as well as the '.RData' file. Now let's turn to the next step--'Compare clinical features'.

[Download RData file of the results](#)

Result display for "Compare survival outcome" on TCGA dataset

(2). Compare clinical features for TCGA dataset

Parameter Settings: The “Steps” chooses “Compare clinical features”, and “Compare clinical features on tcga datasets or validation datasets” chooses “TCGA”. First, we indicate 4 as the number of clinical features for comparison among subtypes derived from clustering results (“The number of variables in survival and clinical information chosen for summary and statistical tests”). After that, we enter the features containing “PAM_Subtype”, “oneNN_Subtype”, “Lund_Subtype”, and “TCGA_Subtype” for “Input the variables” and then define “Subtype” as

stratifying variable (“User defined stratifying variable or not” chooses “No” by default). Besides, we also indicate 4 as the number of categorical clinical features (“The number of the categorical variables in chosen variables”), and then enter all features (“Input the categorical variables”). Moreover, we indicate 0 as the number of features for nonparametric test (“The number of variables for which the p-values should be those of nonparametric tests”) and 4 as the number of features for exact test (“The number of variables for which the p-values should be those of exact tests”) respectively, and then enter the corresponding features (all features for “Input variables for which the p-values should be those of exact tests”). Additionally, “Whether NA should be handled as a regular factor level rather than missing value” chooses “No” and “Transform the '.txt' output file to a '.docx' WORD file or not” chooses “Yes” by default. Finally, we indicate the “The name of the output table” using default setting, and then click the “Process” button to compare clinical features among subtypes.

Steps

- Compare survival outcome
- Compare clinical features
- Compare mutational frequency
- Compare total mutation burden
- Compare fraction genome altered
- Compare drug sensitivity
- Compare agreement with other subtypes

Compare clinical features

In this step, a table that is easy to use in medical research papers will be created to summarize the specified baseline variables (continuous & categorical) stratified by specified categorical variable and then perform statistical tests.

Compare clinical features on tcga datasets or validation datasets

- TCGA Validation

The number of variables in survival and clinical information chosen for summary and statistical tests

4

Input the variables

Variable	1	2	3	4
	PAM_Subty	oneNN_Sub	Lund_Subty	TCGA_Subt

User defined stratifying variable or not

- Yes No

The number of the categorical variables in chosen variables

4

Input the categorical variables

Categorical variable	1	2	3	4
	PAM_Subty	oneNN_Sub	Lund_Subty	TCGA_Subt

The number of variables for which the p-values should be those of nonparametric tests

0

The number of variables for which the p-values should be those of exact tests

4

Input variables for which the p-values should be those of exact tests

Exact test variable	1	2	3	4
	PAM_Subty	oneNN_Sub	Lund_Subty	TCGA_Subt

Whether NA should be handled as a regular factor level rather than missing value

- Yes No

Transform the '.txt' output file to a '.docx' WORD file or not ('.txt' file will be also kept)

- Yes No

The name of the output table

Summarization_of_clinical_variables_stratified_by_current_subtypes(TCGA)

Process

Parameter settings for “Compare clinical features” on TCGA dataset

Result Display:

The software will generate a table to display the comparison results of clinical features among subtypes derived from clustering results. “ $p<0.05$ ” reveals the clinical feature that is significantly correlated with subtypes.

Survival	Clinical Features	Mutational Frequency	TMB	FGA	Drug Sensitivity	Agreement
Table: Summarization of clinical variables stratified by current subtypes (TCGA)						
Copy	CSV	Excel	Print			
	level	CS1	CS2	CS3	CS4	p
1	n	89	105	115	87	
2	PAM_Subtype (%)	basal	89 (100.0)	45 (42.9)	0 (0.0)	60 (69.0)
3		luminal	0 (0.0)	60 (57.1)	115 (100.0)	27 (31.0)
4	oneNN_Subtype (%)	basal	79 (88.8)	9 (8.6)	1 (0.9)	47 (54.0)
5		luminal	0 (0.0)	32 (30.5)	96 (83.5)	23 (26.4)
6		p53-like	10 (11.2)	64 (61.0)	18 (15.7)	17 (19.5)
7	Lund_Subtype (%)	GU	0 (0.0)	38 (36.2)	35 (30.4)	15 (17.2)
8		Infil	26 (29.2)	44 (41.9)	0 (0.0)	1 (1.1)
9		SCCL	60 (67.4)	5 (4.8)	0 (0.0)	38 (43.7)
10		UroA	0 (0.0)	17 (16.2)	79 (68.7)	15 (17.2)
11		UroB	3 (3.4)	1 (1.0)	1 (0.9)	18 (20.7)
12	TCGA_Subtype (%)	Basal_squamous	85 (95.5)	6 (5.7)	0 (0.0)	48 (55.2)
13		Luminal	0 (0.0)	20 (19.0)	3 (2.6)	3 (3.4)
14		Luminal_infiltrated	3 (3.4)	70 (66.7)	0 (0.0)	0 (0.0)
15		Luminal_papillary	0 (0.0)	6 (5.7)	111 (96.5)	22 (25.3)
16		Neuronal	1 (1.1)	3 (2.9)	1 (0.9)	14 (16.1)

Result display for “Compare clinical features” on TCGA dataset

(3). Compare mutational frequency for TCGA dataset

Parameter Settings: The “Steps” chooses “Compare mutational frequency”, and “Compare mutational frequency on tcga datasets or validation datasets” chooses “TCGA”. First, we indicate 0.03 as the frequency cutoff for mutation data. Only features mutated in over than

such proportion would be included in testing. Then, we choose “fisher” for “Statistical method for independence testing”. In addition, “The correction method for multiple comparison” chooses “BH”, “Transform the '.txt' output file to a '.docx' WORD file or not” chooses “Yes”, and we also use the default name for “The name of the output table”. After that, we set the parameters for oncoprint. We first choose “Yes” for “Perform clustering within each subtype for oncoprint or not”, and indicate 0.05 for “The nominal p value cutoff for significant mutations shown in oncoprint” and 0.25 for “The adjusted p value cutoff for significant mutations shown in oncoprint”. Besides, we use the default settings for the mutation and background color for oncoprint (Corresponding parameters choose “No”), as well as the annotation color for each subtype in oncoprint (“Setting for colors to annotate each subtype in oncoprint” chooses “System default”). In addition to annotation for sample through clustering results, this step also allows users to select specific clinical features for annotation, and we select “PAM_Subtype”, “oneNN_Subtype”, “Lund_Subtype” and “TCGA_Subtype”. Moreover, we utilize the default settings for “The width of output figure” and “Figure Name”, and then indicate 5 as “The height of output figure”. Finally, we click the “Process” button to compare mutational frequency among each subtype derived from clustering results.

Steps

- Compare survival outcome
- Compare clinical features
- Compare mutational frequency
- Compare total mutation burden
- Compare fraction genome altered
- Compare drug sensitivity
- Compare agreement with other subtypes

Compare mutational frequency

In this step, a table and an oncoprint will be generated to compare mutational frequency among different multi-omics integrative clusters to test the independency between subtypes and mutational status.

Compare mutational frequency on tcga datasets or validation datasets

- TCGA Validation

The frequency cutoff for mutation data (Only features that mutated in over than such proportion would be included in testing)

0.03

Statistical method for independence testing

- fisher chisq

The correction method for multiple comparison

- holm
- hochberg
- hommel
- bonferroni
- BH
- BY
- fdr

Transform the '.txt' output file to a '.docx' WORD file or not ('.txt' file will be also kept)

- Yes No

The name of the output table

Independent_test_between_subtype_and_mutation(TCGA)

Perform clustering within each subtype for oncoprint or not

- Yes No

The nominal p value cutoff for significant mutations shown in oncoprint

0.05

The adjusted p value cutoff for significant mutations shown in oncoprint

0.25

User defined mutation color for oncoprint or not

- Yes No

User defined background color for oncoprint or not

- Yes No

Setting for colors to annotate each subtype in oncoprint

- System default User defined

Sample annotations from survival information for oncoprint or not

Yes No

The number of sample annotations from survival information for oncoprint

4

Input sample annotations from survival information for oncoprint

First line: Please input the sample annotation variables from survival information

Second line: Please input 'Continuous' or 'Categorical' to indicate the type of each sample annotation variable

Last line: Please input the colors for each sample annotation variable (use hex color format, e.g. #000000FF and English semicolons should be used to separate the input colors)

Note1: If the sample annotation variable is continuous, the number of indicated colors should be equal to 3, which represents the minimum, median and maximum value of this variable)

Note2: If the sample annotation variable is categorical, the number of indicated colors should be equal to the number of categories for this variable)

	1	2	3	4
Sample annotation	PAM_Subty	oneNN_Sub	Lund_Subty	TCGA_Subt
Continuous or Categorical	Categorical	Categorical	Categorical	Categorical
Color settings	#2874C5;#E #2874C5;#C #EABF00;#C #2874C5;#8			

The width of output figure

8

The height of output figure

5

Figure Name

oncoprint_for_mutations_with_frequency_over_than_the_setting_cutoff(TCGA)

Process

Parameter settings for “Compare mutational frequency” on TCGA dataset

Result Display:

The software will generate a table to show the comparison of mutations among subtypes for specific genes with mutation frequency that is higher than the setting cutoff. Additionally, the software also generates an oncoprint to display the mutations of the genes whose mutation frequency is significantly different among subtypes on the basis of the table.

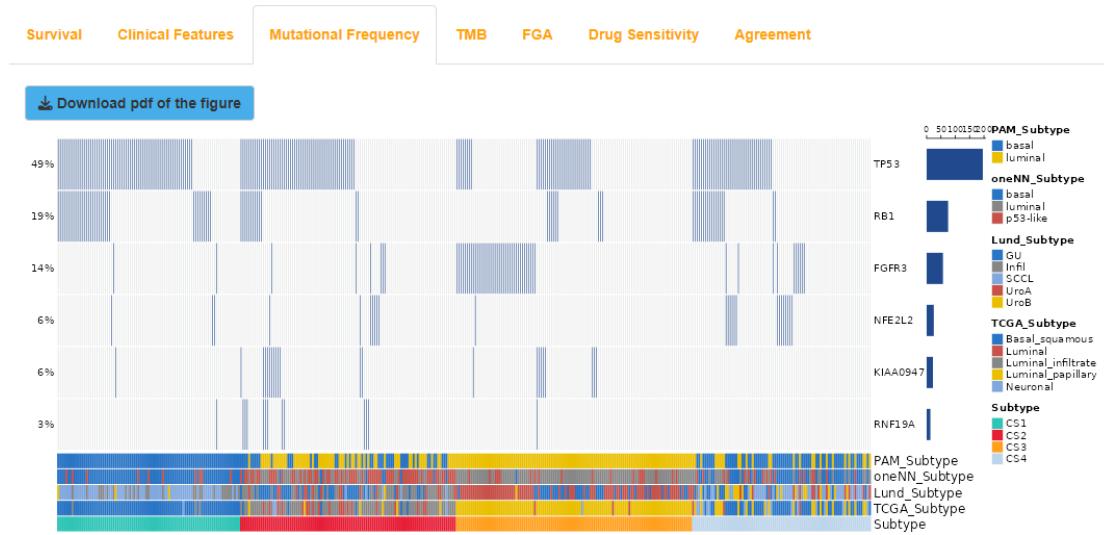


Table: Independent test between subtype and mutation (TCGA)

Gene (Mutated)	TMB	CS1	CS2	CS3	CS4	pvalue	padj
FGFR3	57 (14%)	2 (2.2%)	6 (5.7%)	39 (33.9%)	10 (11.5%)	3.44e-11	4.67e-08
TP53	196 (50%)	66 (74.2%)	56 (53.3%)	35 (30.4%)	39 (44.8%)	7.01e-09	4.76e-06
RB1	75 (19%)	35 (39.3%)	13 (12.4%)	9 (7.8%)	18 (20.7%)	1.52e-07	6.88e-05
RNF19A	13 (3%)	1 (1.1%)	11 (10.5%)	1 (0.9%)	0 (0.0%)	5.86e-05	1.99e-02
NFE2L2	25 (6%)	3 (3.4%)	7 (6.7%)	1 (0.9%)	14 (16.1%)	1.13e-04	3.07e-02
KIAA0947	22 (6%)	1 (1.1%)	12 (11.4%)	9 (7.8%)	0 (0.0%)	2.41e-04	5.45e-02

Result display for “Compare mutational frequency” on TCGA dataset

(4). Compare total mutation burden for TCGA dataset

Parameter Settings: The “Steps” chooses “Compare total mutation burden”, and “Compare total mutation burden on tcga datasets or validation datasets” chooses “TCGA”. First, we choose “Yes” for “Whether remove repeated variants in a particular sample”, and then we choose “No” for “Remove possible FLAGS”. After that, we choose “No” for “Whether user defined a list of variant classifications” and indicate 38 as “The estimation of exome size”. Moreover, “The method for statistical testing” chooses “nonparametric” and “Show the sample size within each subtype at the top of the figure or not” chooses “Yes”. Finally, we use default

settings for “Setting for colors to annotate each subtype” (System default), “The width of boxviolin plot”, “The height of boxviolin plot” as well as “Figure Name”, and then click the “Process” button to compare total mutation burden (TMB) among subtypes derived from clustering results.

Steps

- Compare survival outcome
- Compare clinical features
- Compare mutational frequency
- Compare total mutation burden
- Compare fraction genome altered
- Compare drug sensitivity
- Compare agreement with other subtypes

Compare total mutation burden

In this step, we will calculate Total Mutation Burden (TMB) and compare them among current subtypes.

Compare total mutation burden on tcga datasets or validation datasets

TCGA Validation

Whether remove repeated variants in a particuar sample, mapped to multiple transcripts of same gene

Yes No

Remove possible FLAGS (These FLAGS genes are often non-pathogenic and passengers, but are frequently mutated in most of the public exome studies, some of which are fishy. Examples of such genes include TTN, MUC16, etc)

Yes No

Whether user defined a list of variant classifications that should be considered as non-synonymous and the rest will be considered synonymous

Yes No

The estimation of exome size
38

The method for statistical testing
 parametric nonparametric

Show the sample size within each subtype at the top of the figure or not
 Yes No

Setting for colors to annotate each subtype
 System default User defined

The width of boxviolin plot
6

The height of boxviolin plot
6

Figure Name
distribution_of_TMB_and_titv(TCGA)

Process

Parameter settings for “Compare total mutation burden” on TCGA dataset

Result Display:

The software will generate a box-violin plot to show the comparison results of total mutation burden (TMB) among subtypes, and the total mutation burden (TMB) of each sample will be also displayed through a table. “ $p<0.05$ ” shows significantly different on total mutation burden (TMB) among subtypes.

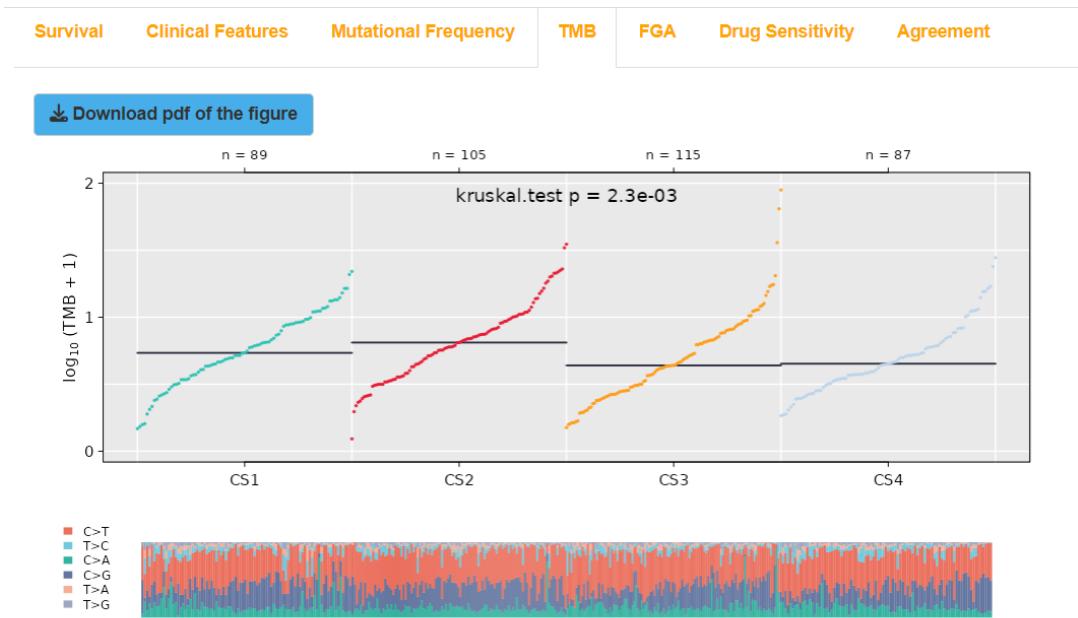


Table: Comparison of TMB among identified subtypes (TCGA)

Show entries Search:

samID	variants	TMB	log10TMB	Subtype
TCGA-DK-A3WY-01A	18	0.473684210526316	0.16840443038939	CS1
TCGA-XF-AAME-01A	20	0.526315789473684	0.183644396946127	CS1
TCGA-FD-A5BS-01A	22	0.578947368421053	0.198367653766833	CS1

Result display for “Compare total mutation burden” on TCGA dataset

(5). Compare fraction genome altered for TCGA dataset

Parameter Settings: The “Steps” chooses “Compare fraction genome altered”, and “Compare fraction genome altered on tcga datasets or validation datasets” chooses “TCGA”. First, we choose “segments value (segment_mean)” for “The type of the 'value' column in segmented copy number dataset”, and then indicate 0.3 as “The cutoff for identifying copy-number gain or loss”. After that, we choose “nonparametric” for “The method for statistical testing” and we utilize default settings for “Setting for mapping colors for bars of FGA, FGG and FGL” (System default) and

“Setting for colors to annotate each subtype” (System default). Finally, we indicate “The width of barplot” as well as “Figure Name” using default settings, and indicate 3 as “The height of barplot”. Then, we click the “Process” button to compare the fraction of genome altered (FGA), the fraction of genome gained (FGG) and the fraction of genome lost (FGL) among each subtype derived from clustering results.

Steps

- Compare survival outcome
- Compare clinical features
- Compare mutational frequency
- Compare total mutation burden
- Compare fraction genome altered
- Compare drug sensitivity
- Compare agreement with other subtypes

Compare fraction genome altered

This step calculates Fraction Genome Altered (FGA), Fraction Genome Gained (FGG) as well as Fraction Genome Lost (FGL) separately, and compares them among current subtypes.

Compare fraction genome altered on tcga datasets or validation datasets

TCGA Validation

The type of the 'value' column in segmented copy number dataset

copy-number value (original) segments value (segment_mean)

The cutoff for identifying copy-number gain or loss

The method for statistical testing

parametric nonparametric

Setting for mapping colors for bars of FGA, FGG and FGL

System default User defined

Setting for colors to annotate each subtype

System default User defined

The width of barplot

The height of barplot

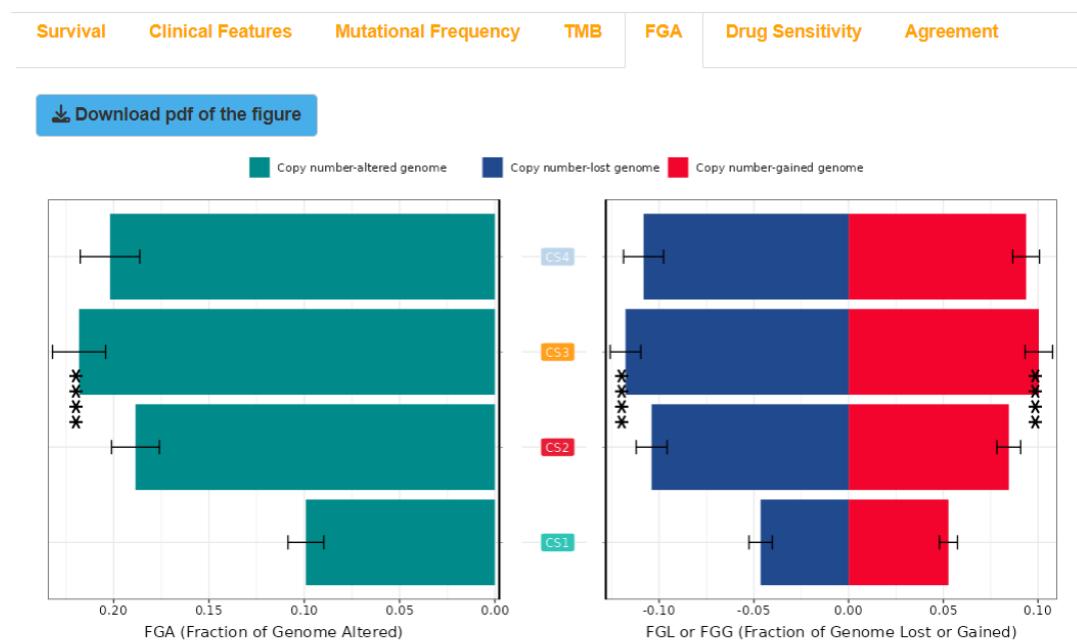
Figure Name

Process

Parameter settings for “Compare fraction genome altered” on TCGA dataset

Result Display:

The software will generate bar plots to show the comparison results of FGA, FGG and FGL among subtypes, and the values of FGA, FGG, and FGL of each sample will be also displayed through a table. The sign of “**” reveals significantly different on FGA, or FGG, or FGL among subtypes.



Result display for “Compare fraction genome altered” on TCGA dataset

(6). Compare drug sensitivity for TCGA dataset

Parameter Settings: The “Steps” chooses “Compare drug sensitivity”, and “Compare drug sensitivity on tcga datasets or validation datasets” chooses “TCGA”. First, we indicate the name of the drugs from GDSC, here

we indicate “Cetuximab” and “Erlotinib” separately. Then, we choose “Yes” for “Train the models on only a subset of the CGP cell lines or not”, and then choose “urogenital_system” for “The tissue type from which the cell lines originated”. After that, we choose “nonparametric” for “The method for statistical testing”. In addition, we use default settings for “Setting for colors to annotate each subtype” (System default), “The seed for reproducing the result of comparing drug sensitivity” (123456), “The width of boxviolin plot”, “The height of boxviolin plot” and “The prefix for the name of output boxviolin plot”. Finally, we click the “Process” button to compare drug sensitivity among subtypes derived from clustering results.

Steps

- Compare survival outcome
- Compare clinical features
- Compare mutational frequency
- Compare total mutation burden
- Compare fraction genome altered
- Compare drug sensitivity
- Compare agreement with other subtypes

Compare drug sensitivity

This step estimates the IC50 of specific drug for each subtype by developing a ridge regression predictive model based on all/specific cell lines derived from Genomics of Drug Sensitivity in Cancer (GDSC) and compares the IC50 among current subtypes.

Compare drug sensitivity on tcga datasets or validation datasets

TCGA Validation

The name of the drug from GDSC for which you would like to predict sensitivity

Train the models on only a subset of the CGP cell lines or not

Yes No

The tissue type from which the cell lines originated

- aero_digestive_tract
- blood
- bone
- breast
- digestive_system
- lung
- nervous_system
- skin
- urogenital_system

The method for statistical testing

parametric nonparametric

Setting for colors to annotate each subtype

System default User defined

The seed for reproducing the result of comparing drug sensitivity

123456

The width of boxviolin plot

5

The height of boxviolin plot

5

The prefix for the name of output boxviolin plot (the name of output boxviolin plot will be defined as 'prefix+the name of the indicated drug')

boxviolin_of_estimated_IC50(TCGA)

Process

Compare drug sensitivity

This step estimates the IC₅₀ of specific drug for each subtype by developing a ridge regression predictive model based on all/specific cell lines derived from Genomics of Drug Sensitivity in Cancer (GDSC) and compares the IC₅₀ among current subtypes.

Compare drug sensitivity on tcga datasets or validation datasets

TCGA Validation

The name of the drug from GDSC for which you would like to predict sensitivity

Erlotinib

Parameter settings for “Compare drug sensitivity” on TCGA dataset

Result Display:

The software will generate box-violin plots to show the IC₅₀ comparison results among subtypes for drugs, and the estimated IC₅₀ of each sample for drugs will be also displayed through tables. “*p*<0.05” reveals significance of differences on the response to drugs among subtypes.

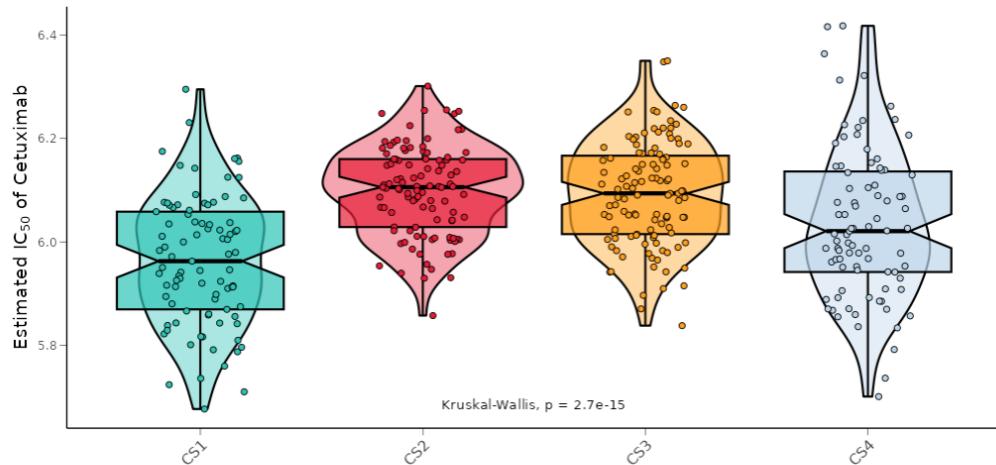
[Download pdf of the figure](#)


Table: Comparison of estimated IC50 (TCGA) for Cetuximab among identified subtypes

Copy CSV Excel Print Show 10 entries Search:

	Est.IC50	Subtype
TCGA-HQ-A5NE-01A	5.73612825749444	CS1
TCGA-FD-A5BT-01A	6.23068306420836	CS1
TCGA-XF-AAMW-01A	5.79149025378978	CS1

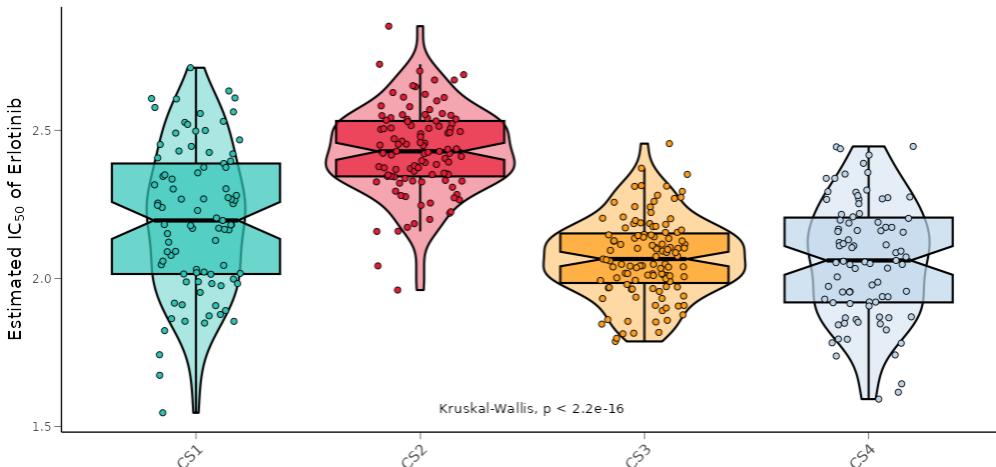
[Download pdf of the figure](#)


Table: Comparison of estimated IC50 (TCGA) for Erlotinib among identified subtypes

Copy CSV Excel Print Show 10 entries Search:

	Est.IC50	Subtype
TCGA-HQ-A5NE-01A	2.03047625576851	CS1
TCGA-FD-A5BT-01A	2.52634061080072	CS1
TCGA-XF-AAMW-01A	1.87375678066546	CS1

Result display for “Compare drug sensitivity” on TCGA dataset

(7). Compare agreement with other subtypes for TCGA dataset

Parameter Settings: The “Steps” chooses “Compare agreement with other subtypes”, and “Compare agreement with other subtypes on tcga datasets or validation datasets” chooses “TCGA”. We first indicate 5 as “The number of the traditional subtypes for comparison”, and then enter the names of the traditional subtypes containing “tumor_stage”, “PAM_Subtype”, “oneNN_Subtype”, “Lund_Subtype”, and “TCGA_Subtype” (“Input the variable name of traditional subtypes in survival and clinical information for comparison”). Besides, we indicate 0.2 as “The width for box in alluvial diagram”. Additionally, “Setting for colors to annotate each subtype”, “The width of the figure”, “The height of the figure”, and “Figure Name” use default settings. Finally, we click the “Process” button to compare agreement between subtypes derived from clustering results and other traditional subtypes.

Steps

- Compare survival outcome
- Compare clinical features
- Compare mutational frequency
- Compare total mutation burden
- Compare fraction genome altered
- Compare drug sensitivity
- Compare agreement with other subtypes

Compare agreement with other subtypes

This step aims to compute four evaluation indicators, including Rand Index, Jaccard Index, Fowlkes-Mallows, and Normalized Mutual Information for agreement of two partitions, then generate a barplot and an alluvial diagram for visualization.

Compare agreement with other subtypes on tcga datasets or validation datasets

TCGA Validation

The number of the traditional subtypes for comparison (1-6)

5

Input the variable name of traditional subtypes in survival and clinical information for comparison

Variable name	1	2	3	4	5
tumor_stag	PAM_Subty	oneNN_Su	Lund_Subty	TCGA_Sub	

Setting for colors to annotate each subtype

System default User defined

The width for box in alluvial diagram

0.2

The width of the figure

6

The height of the figure

5

Figure Name

agreement_between_current_subtype_and_other_classifications(TCGA)

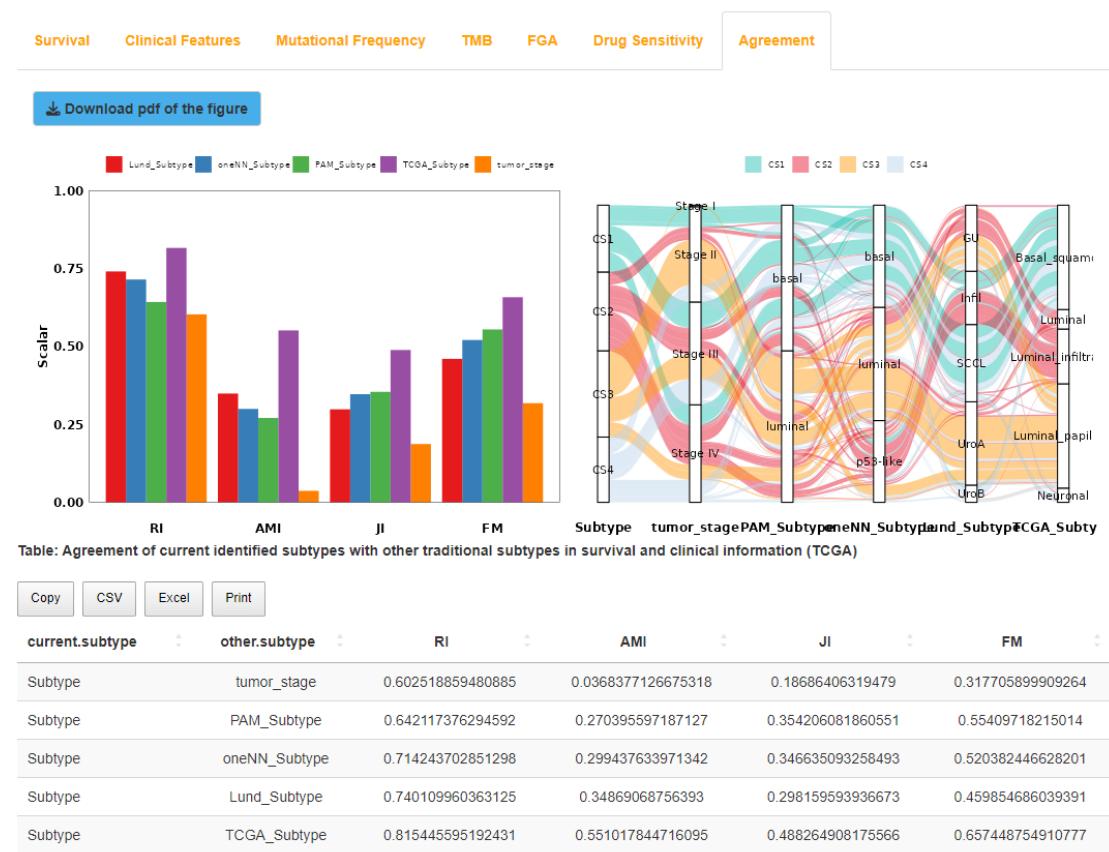
Process

Parameter settings for “Compare agreement with other subtypes” on TCGA dataset

Result Display:

The software will generate an alluvial diagram to compare agreement

between subtypes derived from clustering results and other traditional subtypes. In addition, a bar plot and a table are also generated to display the values of four statistical indicators which are utilized to evaluate the agreement, including Rand Index (RI), Adjusted Mutual Information (AMI), Jaccard Index (JI), and Fowlkes-Mallows (FM). All these indicators range from 0 to 1, and the larger the values are, the clustering results are more similar to the current classification results.



Result display for “Compare agreement with other subtypes” on TCGA dataset

The fourth step: RUN Module

(1). Run differential expression analysis for TCGA dataset

Parameter Settings: “Module switching options” on the upper left of the

software chooses “RUN Module”, and then the “Steps” chooses “Run differential expression analysis”. First, we choose “limma” as “Choose the algorithm for differential expression analysis”, and then we indicate the prefix of the output “.txt” file (we ignore “Indicate the prefix of output file” and type nothing). Additionally, “overwrite it or skip this step directly” chooses “Overwrite”, “Whether sort adjusted p value in ascending order for output table” chooses “Yes” and “Only select 'id', 'log2fc', 'pvalue' and 'padj' columns for output table or not” chooses “Yes”. Finally, we click the “Process” button to run differential expression analysis.

Data Preparation	GET Module	COMP Module	RUN Module	Users Guide
----------------------------------	----------------------------	-----------------------------	----------------------------	-----------------------------

Steps

- Run differential expression analysis
- Run biomarker identification procedure
- Run gene set enrichment analysis
- Run gene set variation analysis
- Run nearest template prediction
- Run partition around medoids classifier
- Run consistency evaluation using Kappa statistics

Run differential expression analysis

In this step, we will perform differential expression analysis using chosen algorithm (deseq2 or edger or limma) between two classes identified by multi-omics clustering process.

Choose the algorithm for differential expression analysis

deseq2 edger limma

Indicate the prefix of output file (e.g. if the name of output file is 'consensusMOIC_nscic_limma_test_result.CS1_vs_Others.txt', the prefix would be 'nscic'. If you do not want a prefix, ignore this parameter and type nothing)

The result already exists, overwrite it or skip this step directly

Overwrite Skip

Whether sort adjusted p value in ascending order for output table

Yes No

Only select 'id', 'log2fc', 'pvalue' and 'padj' columns for output table or not

Yes No

Process

Parameter settings for “Run differential expression analysis” on TCGA dataset

Result Display:

The software will generate a table to display the results of differential expression analysis.

id	fc	log2fc	pvalue	padj	compname
RBBP8NL	0.133991733910389	-2.8997840928187	1.06333887181011e-55	3.19001661543034e-52	CS1_vs_Others
PDCD1LG2	5.57988062645971	2.48023425799411	6.21685105116882e-55	9.32527657675324e-52	CS1_vs_Others
HID1	0.121084552014183	-3.04591327737074	6.52037558219528e-51	6.52037558219528e-48	CS1_vs_Others
BICDL2	0.142795180881191	-2.8079808034484	1.99093908906404e-46	1.49320431679803e-43	CS1_vs_Others
GRHL3	0.0765906088993295	-3.70668868176224	3.26943192651628e-46	1.96165915590977e-43	CS1_vs_Others
ZBTB7C	0.146095820512008	-2.77501318856558	6.45309821753183e-45	3.22654910876591e-42	CS1_vs_Others
KLHDC7A	0.0856928236853813	-3.54468179835427	8.65163843119724e-45	3.70784504194167e-42	CS1_vs_Others
TMPRSS2	0.0663575303039213	-3.91359599673876	3.38736810018551e-44	1.27026303756957e-41	CS1_vs_Others
MYCL	0.123221760700744	-3.02067103948097	2.60813525653578e-43	8.15926875472025e-41	CS1_vs_Others
FOXA1	0.0993353765235377	-3.33154859039027	2.77933802071992e-43	8.15926875472025e-41	CS1_vs_Others

Result display for “Run differential expression analysis” on TCGA dataset

(2). Run biomarker identification procedure for TCGA dataset

Parameter Settings: The “Steps” chooses “Run biomarker identification procedure”. First, we choose the algorithm (limma) that we use for differential expression analysis. Then, we indicate 0.05 for both the “nominal p value” and “adjusted p value” to identify significant markers for each subtype derived from clustering results. After that, we choose “up-regulated” and “down-regulated” for “Indicate the direction of identifying significant marker” in turn to screen significant up-regulated and down-regulated markers for each subtype respectively. Furthermore,

we indicate 30 and 50 as “the number of top markers sorted by log2fc should be identified for each subtype” for up-regulated and down-regulated markers respectively. For the processing of expression data, we normalize the data by default (both “Indicate if expression data should be centered” and “Indicate if expression data should be scaled” choose “Yes”). After that, we choose “Yes” for “Assign marginal cutoff for truncating values in data or not” and indicate 3 as “Marginal cutoff for truncating values in data”. For the settings of the heatmap, “Colors for annotating each cluster at the top of heatmap” (System default), “Show rownames (feature names) in heatmap or not” (No), “Show colnames (sample ID) in heatmap or not” (No), and “Colors for heatmap” (System default) use default settings. We indicate 12 for “The width of output figure” and “The height of output figure”. In addition to annotation for sample through clustering results, this step also allows users to select specific clinical features for annotation, and we select “PAM_Subtype”, “oneNN_Subtype”, “Lund_Subtype” and “TCGA_Subtype”. Finally, we indicate the prefix of the output figure (we ignore “Indicate the prefix of output figure” and type nothing), and then click the “Process” button to run biomarker identification procedure.

Steps

- Run differential expression analysis
- Run biomarker identification procedure
- Run gene set enrichment analysis
- Run gene set variation analysis
- Run nearest template prediction
- Run partition around medoids classifier
- Run consistency evaluation using Kappa statistics

Run biomarker identification procedure

This step aims to identify uniquely and significantly expressed (overexpressed or downexpressed) biomarkers for each subtype identified by multi-omics clustering process. A template including top markers will be generated for subtype external verification and a heatmap will also be generated.

Indicate the algorithm for completed differential expression analysis

- deseq2
- edger
- limma

Indicate the nominal p value for identifying significant markers

0.05

Indicate the adjusted p value for identifying significant markers

0.05

Indicate the direction of identifying significant marker

- up-regulated
- down-regulated

Indicate the number of top markers sorted by log2fc should be identified for each subtype

30

Sample annotations from survival information for heatmap or not

- Yes
- No

The number of sample annotations from survival information for heatmap

4

Input sample annotations from survival information for heatmap

First line: Please input the sample annotation variables from survival information

Second line: Please input 'Continuous' or 'Categorical' to indicate the type of each sample annotation variable

Last line: Please input the colors for each sample annotation variable (use hex color format, e.g. #000000FF and English semicolons should be used to separate the input colors)

Note1: If the sample annotation variable is continuous, the number of indicated colors should be equal to 3, which represents the minimum, median and maximum value of this variable)

Note2: If the sample annotation variable is categorical, the number of indicated colors should be equal to the number of categories for this variable)

	1	2	3	4
Sample annotation	PAM_Subty	oneNN_Sub	Lund_Subty	TCGA_Subt
Continuous or Categorical	Categorical	Categorical	Categorical	Categorical
Color settings	#2874C5;#E #2874C5;#C #EABF00;#R #2874C5;#8			

Colors for annotating each cluster at the top of heatmap

- System default
- User defined

Indicate if expression data should be centered

Yes No

Indicate if expression data should be scaled

Yes No

Assign marginal cutoff for truncating values in data or not

Yes No

Marginal cutoff for truncating values in data

3

Show rownames (feature names) in heatmap or not

Yes No

Show colnames (sample ID) in heatmap or not

Yes No

Colors for heatmap

System default User defined

The width of output figure

12

The height of output figure

12

Indicate the prefix of output figure (e.g. if the name of output figure is 'markerheatmap_using_upregulated_genes.pdf', the prefix would be 'markerheatmap'. If you do not want a prefix but using the system default prefix 'markerheatmap', ignore this parameter and type nothing)

Process

Indicate the direction of identifying significant marker

up-regulated down-regulated

Indicate the number of top markers sorted by log2fc should be identified for each subtype

50

Parameter settings for “Run biomarker identification procedure” on TCGA dataset

Result Display:

The software will generate heatmaps to show the expression of the screening up-regulated and down-regulated markers in each subtype

(nominal p value < 0.05 & adjusted p value < 0.05). In addition, the screening markers for each subtype are also displayed through tables.

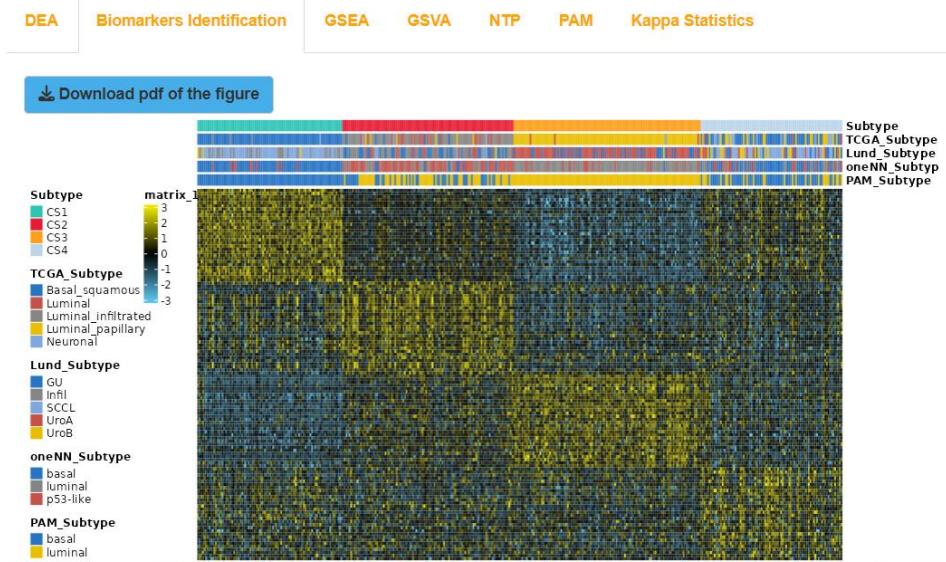


Table: Subtype-specific upregulated biomarkers for each identified subtype based on differential expression analysis results using 'limma' algorithm

	Copy	CSV	Excel	Print	Show 10 entries	Search:
probe					class	
SAA1					CS1	up
MYOSLID					CS1	up

DEA Biomarkers Identification GSEA GSVA NTP PAM Kappa Statistics

Download pdf of the figure

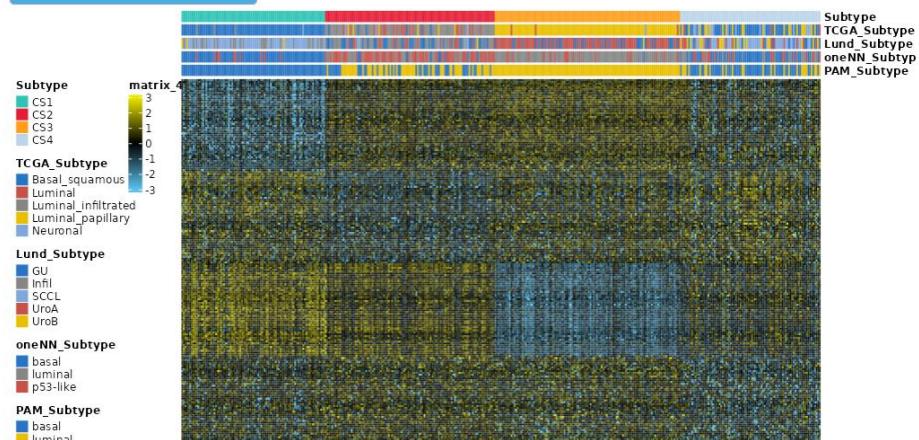


Table: Subtype-specific downregulated biomarkers for each identified subtype based on differential expression analysis results using 'limma' algorithm

	Copy	CSV	Excel	Print	Show 10 entries	Search:
probe					class	
GPX2					CS1	down
UPK1B					CS1	down

Result display for “Run biomarker identification procedure” on TCGA dataset

(3). Run gene set enrichment analysis for TCGA dataset

Parameter Settings: The “Steps” chooses “Run gene set enrichment analysis”. First, we upload the gene set background file (reactome.hallmark.v7.3.symbols.gmt), and choose “Upload manually” for “The manner for gene set background file preparation”. Then, we click the “Browse” button to upload the gene set background file and click the “Process” button to finish the preparation of the gene set background file. After that, we choose the algorithm (limma) that we use for differential expression analysis, and then choose “up-regulated” and “down-regulated” for “Indicate the direction of identifying significant pathway” in turn to screen significant up-regulated and down-regulated pathways for each subtype derived from clustering results respectively. After that, we indicate 10 as “the number of top pathways sorted by NES should be identified for each subtype”. Besides, we indicate 1000 as “the number of permutations for gene set enrichment analysis”, 10 as “minimal size of each gene set for analysis” and 500 as “maximal size of each gene set for analysis”. Furthermore, we indicate 0.05 for “nominal p value” and 0.25 for “adjusted p value” to identify significant pathways for each subtype derived from clustering results. Additionally, “Indicate the method to employ in the estimation of gene set enrichment scores per sample” chooses “gsva” and “Indicate the method to calculate subtype-specific

pathway enrichment scores” chooses “mean”. For the settings of the heatmap, we utilize default settings for “Colors for annotating each cluster at the top of heatmap” (System default), “Colors for heatmap” (System default). Moreover, we indicate 10 for “The width of output figure” and 20 for “The height of output figure”. Finally, we indicate the prefix of the output figure (we ignore “Indicate the prefix of output figure” and type nothing), and then click the “Process” button to run gene set enrichment analysis.

The screenshot shows two stacked panels of a software application. The top panel has a blue header labeled "Steps" and contains a list of seven items, each with a radio button and a description:

- Run differential expression analysis
- Run biomarker identification procedure
- Run gene set enrichment analysis
- Run gene set variation analysis
- Run nearest template prediction
- Run partition around medoids classifier
- Run consistency evaluation using Kappa statistics

The bottom panel has a blue header labeled "Prepare a gene set background file" in orange text. It contains the following sections:

- First we need to prepare a gene set background file for gene set enrichment analysis.**
- The manner for gene set background file preparation**
- System default Download from specified url Upload manually
- Upload the gene set background file**
- A file input field with the text "reactome.hallmark.v7.3.symbols.gmt" and a "Browse..." button.
- A blue button labeled "Upload complete".
- A large blue button labeled "Process".

Run gene set enrichment analysis

This step aims to perform gene set enrichment analysis using a background file to identify subtype-specific (overexpressed or downexpressed) functional pathways for each subtype.

Indicate the algorithm for completed differential expression analysis

- deseq2 edger limma

Indicate the direction of identifying significant pathway

- up-regulated down-regulated

Indicate the number of top pathways sorted by NES should be identified for each subtype

10

Indicate the number of permutations for gene set enrichment analysis (1000 by default and 10000 will be better for reproducibility)

1000

Indicate minimal size of each gene set for analysis

10

Indicate maximal size of each gene set for analysis

500

Indicate the nominal p value for identifying significant pathways

0.05

Indicate the adjusted p value for identifying significant pathways

0.25

Indicate the method to employ in the estimation of gene set enrichment scores per sample

- gsva ssgsea zscore plage

Indicate the method to calculate subtype-specific pathway enrichment scores

- mean median

Colors for annotating each cluster at the top of heatmap

- System default User defined

Colors for heatmap

- System default User defined

The width of output figure

10

The height of output figure

20

Indicate the prefix of output figure (e.g. if the name of output figure is 'gseaheatmap_using_upregulated_pathways.pdf', the prefix would be 'gseaheatmap'. If you do not want a prefix but using the system default prefix 'gseaheatmap', ignore this parameter and type nothing)

Process

Indicate the direction of identifying significant pathway

up-regulated down-regulated

Parameter settings for “Run gene set enrichment analysis” on TCGA dataset

Result Display:

The software will generate heatmaps to show the enrichment scores of the screening up-regulated and down-regulated pathways in each subtype (nominal p value < 0.05 & adjusted p value < 0.25). In addition, the gene set enrichment analysis results as well as enrichment scores of the screening up-regulated and down-regulated pathways in each subtype are also displayed through tables.

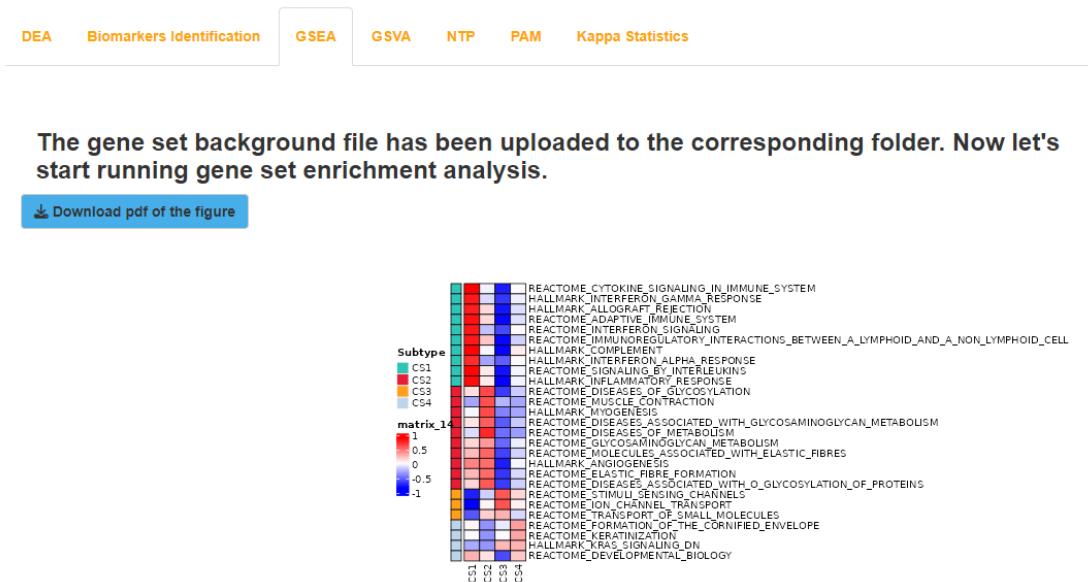


Table 1: GSEA results with upregulated pathways for each identified subtype based on differential expression analysis results using 'limma' algorithm

	setSize	enrichmentScore	NES	pvalue	p.adjust	qvalue	Subtype
	127	0.65455508768311	3.10065410909388	0.00198807157057654	0.00628930817610063	0.0028649357034869	CS1
	70	0.730702519862898	3.08165589822424	0.00204081632653061	0.00628930817610063	0.0028649357034869	CS1
	56	0.742781213281168	3.0209333303717	0.00205761316872428	0.00628930817610063	0.0028649357034869	CS1
	88	0.673704881640801	2.98102220743303	0.00204081632653061	0.00628930817610063	0.0028649357034869	CS1
	44	0.725415383190205	2.7792422795105	0.00209643605870021	0.00628930817610063	0.0028649357034869	CS1
MPHOID_CELL	36	0.743003578658666	2.70561420645678	0.0020746887966805	0.00628930817610063	0.0028649357034869	CS1
	53	0.653268909894806	2.64155332490265	0.00200400801603206	0.00628930817610063	0.0028649357034869	CS1
	26	0.774246038549765	2.63409256567362	0.00195694716242661	0.00628930817610063	0.0028649357034869	CS1
	78	0.600584493808231	2.602166178291969	0.00196463654223969	0.00628930817610063	0.0028649357034869	CS1
	67	0.620408260146148	2.60202928159608	0.00204918032786885	0.00628930817610063	0.0028649357034869	CS1

Showing 1 to 10 of 75 entries

Previous 1 2 3 4 5 ... 8 Next

Table 2: subtype-specific enrichment scores on upregulated pathways among identified subtypes based on differential expression analysis results using 'limma' algorithm

	CS1	CS2	CS3	CS4
_SYSTEM	0.908587480157412	-0.00703097246683867	-0.731196664584179	0.0415393887437293
	0.827772425351802	-0.0829658839413201	-0.645342447420137	-0.0139278750780818
	0.808304282740688	0.164237954622694	-0.761357249150002	-0.0709937319385595
	0.851339340824841	0.166880881588706	-0.807107961642894	-0.0664391382709674
	0.829096184105986	-0.170831669511776	-0.58425531884131	0.0361532811033746
DNs_BETWEEN_A_LYMPHOID_AND_A_NON_LYMPHOID_CELL	0.823931335409712	0.234259403977811	-0.805856843564454	-0.118173362312457
	0.896054194449996	0.0123733956545938	-0.81574426580805	0.105893684035972
	0.761569274386063	-0.270164805705296	-0.507306660782453	0.0467695684322682
	0.892778683272245	0.109958331913571	-0.77140562167291	0.00970130141281623
	0.873311796778305	0.104031583633586	-0.818095354801434	0.00250461236052602

Showing 1 to 10 of 27 entries

Previous 1 2 3 Next

DEA Biomarkers Identification GSEA GSVA NTP PAM Kappa Statistics

The gene set background file has been uploaded to the corresponding folder. Now let's start running gene set enrichment analysis.

[Download pdf of the figure](#)

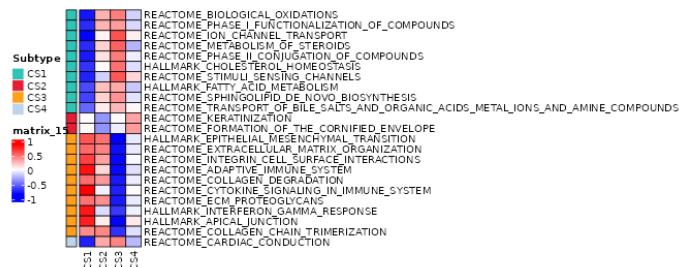


Table 1: GSEA results with downregulated pathways for each identified subtype based on differential expression analysis results using 'limma' algorithm

	Copy	CSV	Excel	Print	Show 10 entries	Search:	
	setSize	enrichmentScore	NES	pvalue	p.adjust	qvalue	Subtype
	33	-0.632603318795763	-2.26590982745526	0.00197628458498024	0.00628930817610063	0.0028649357034869	CS1
	23	-0.621685385219035	-2.01957180484275	0.00199203187250996	0.00628930817610063	0.0028649357034869	CS1
	22	-0.603304886642277	-1.93222015885375	0.00200803212851406	0.00628930817610063	0.0028649357034869	CS1
	14	-0.65048218418364	-1.85006402143245	0.00823045267489712	0.0178152836380684	0.00811530310382927	CS1
	10	-0.702378459319984	-1.80063327244324	0.00628930817610063	0.0151474887058198	0.00690005641262337	CS1
	11	-0.679493282592888	-1.80042966122803	0.00814663951120163	0.0178152836380684	0.00811530310382927	CS1
	14	-0.620989130910023	-1.76618157534796	0.00823045267489712	0.0178152836380684	0.00811530310382927	CS1
	18	-0.563601457413615	-1.74973682801151	0.0115606936416185	0.0232573954437266	0.0105943198697185	CS1
	10	-0.635059117971616	-1.62805188942618	0.0188679245283019	0.0350697292863002	0.0159751306074867	CS1
COMPOUNDS	11	-0.582961427097482	-1.5446525691807	0.0427698574338086	0.0769857433808554	0.0350689135745355	CS1

Showing 1 to 10 of 105 entries

Previous 1 2 3 4 5 ... 11 Next

Table 2: subtype-specific enrichment scores on downregulated pathways among identified subtypes based on differential expression analysis results using 'limma' algorithm

	Copy	CSV	Excel	Print	Show 10 entries	Search:		
					CS1	CS2	CS3	CS4
					-0.770627938433205	0.30078898877846	0.478862151853589	-0.117353320001267
OF_COMPOUNDS					-0.690282557518981	0.2902722223263	0.371670638297782	-0.0934623463135406
					-0.82968840107253	0.0759021720188665	0.630546992600368	0.0919210701592323
					-0.694107591835339	0.191254563852626	0.577466370026564	-0.216799817204467
OMPOUNDS					-0.752876549162572	0.135499724942267	0.483673390997174	-0.0169441639514944
					-0.661275657421243	0.0435428334732109	0.518727959054105	-0.0774361504756129
					-0.720677139563765	-0.122166013194812	0.6146682723452	0.177473197275284
					-0.587882839698883	0.252331456205827	0.329353092081203	-0.149628316911315
YNTHESIS					-0.614055252758763	0.192551894749849	0.40828299947314	-0.0631777007983775
ND_ORGANIC_ACIDS_METAL_IONS_AND_AMINE_COMPOUNDS					-0.491767197799781	0.101597406419201	0.277918102843314	0.0609146468705173

Showing 1 to 10 of 23 entries

Previous 1 2 3 Next

Result display for “Run gene set enrichment analysis” on TCGA dataset

(4). Run gene set variation analysis for TCGA dataset

Parameter Settings: The “Steps” chooses “Run gene set variation analysis”.

First, we upload the gene set list of interest (h.all.v7.2.symbols.gmt), and choose “Upload manually” for “The manner for the gene set list of interest preparation”. Then, we click the “Browse” button to upload the gene set list of interest and click the “Process” button to finish the preparation of

the gene set list of interest. After that, “Indicate the method to employ in the estimation of gene set enrichment scores per sample” chooses “gsva” by default. Besides, we choose “Yes” for both “Indicate if enrichment scores should be centered or not” and “Indicate if enrichment scores should be scaled or not”. Furthermore, we choose “Yes” for “Assign marginal cutoff for truncating enrichment scores or not”, and then indicate 1 as “Marginal cutoff for truncating enrichment scores”. For the settings of the heatmap, “Colors for annotating each cluster at the top of heatmap” (System default), “Colors for heatmap” (System default), “Distance measurement for hierarchical clustering” (euclidean), “Clustering method for hierarchical clustering” (ward.D), “Show rownames (feature names) in heatmap or not” (Yes), “Show colnames (sample ID) in heatmap or not” (No) use default settings. Moreover, we indicate 12 for “The width of output figure” and 10 for “The height of output figure”. In addition to annotation for sample through clustering results, this step also allows users to select specific clinical features for annotation, and we select “PAM_Subtype”, “oneNN_Subtype”, “Lund_Subtype” and “TCGA_Subtype”. Finally, we indicate the prefix of the output figure (we ignore “Indicate the prefix of output figure” and type nothing), and then click the “Process” button to run gene set variation analysis.

Steps

- Run differential expression analysis
- Run biomarker identification procedure
- Run gene set enrichment analysis
- Run gene set variation analysis
- Run nearest template prediction
- Run partition around medoids classifier
- Run consistency evaluation using Kappa statistics

Prepare a gene set list of interest

First we need to prepare a gene set list of interest for gene set variation analysis.

The manner for the gene set list of interest preparation

- System default
- Download from specified url
- Upload manually

Upload the gene set list of interest

Browse... h.all.v7.2.symbols.gmt

Upload complete

Process

Run gene set variation analysis

This step aims to use gene set variation analysis to calculate enrichment score of each sample in each subtype based on a given gene set list of interest.

Indicate the method to employ in the estimation of gene set enrichment scores per sample

- gsva
- ssgsea
- zscore
- plage

Indicate if enrichment scores should be centered or not

- Yes
- No

Indicate if enrichment scores should be scaled or not

- Yes
- No

Assign marginal cutoff for truncating enrichment scores or not

- Yes
- No

Marginal cutoff for truncating enrichment scores

1

Sample annotations from survival information for heatmap or not Yes No**The number of sample annotations from survival information for heatmap**

4

Input sample annotations from survival information for heatmap

First line: Please input the sample annotation variables from survival information

Second line: Please input 'Continuous' or 'Categorical' to indicate the type of each sample annotation variable

Last line: Please input the colors for each sample annotation variable (use hex color format, e.g. #0000004FF and English semicolons should be used to separate the input colors)

Note1: If the sample annotation variable is continuous, the number of indicated colors should be equal to 3, which represents the minimum, median and maximum value of this variable)

Note2: If the sample annotation variable is categorical, the number of indicated colors should be equal to the number of categories for this variable)

	1	2	3	4
Sample annotation	PAM_Subty	oneNN_Sub	Lund_Subty	TCGA_Subt
Continuous or Categorical	Categorical	Categorical	Categorical	Categorical
Color settings	#2874C5;#E #2874C5;#C #EABF00;#G #2874C5;#8			

Colors for annotating each cluster at the top of heatmap System default User defined**Colors for heatmap** System default User defined**Distance measurement for hierarchical clustering**

euclidean

Clustering method for hierarchical clustering

ward.D

Show rownames (feature names) in heatmap or not Yes No**Show colnames (sample ID) in heatmap or not** Yes No**The width of output figure**

12

The height of output figure

10

Indicate the prefix of output figure (e.g. if the name of output figure is 'enrichment_heatmap_using_gsva.pdf', the prefix would be 'enrichment_heatmap_using'. If you do not want a prefix but using the system default prefix 'enrichment_heatmap_using', ignore this parameter and type nothing)

Process**Parameter settings for “Run gene set variation analysis” on TCGA dataset**

Result Display:

The software will generate a heatmap to show the enrichment scores of the pathways of interest in each sample. In addition, the raw enrichment scores as well as the z-scored enrichment scores of the pathways of interest in each sample are also displayed through tables.

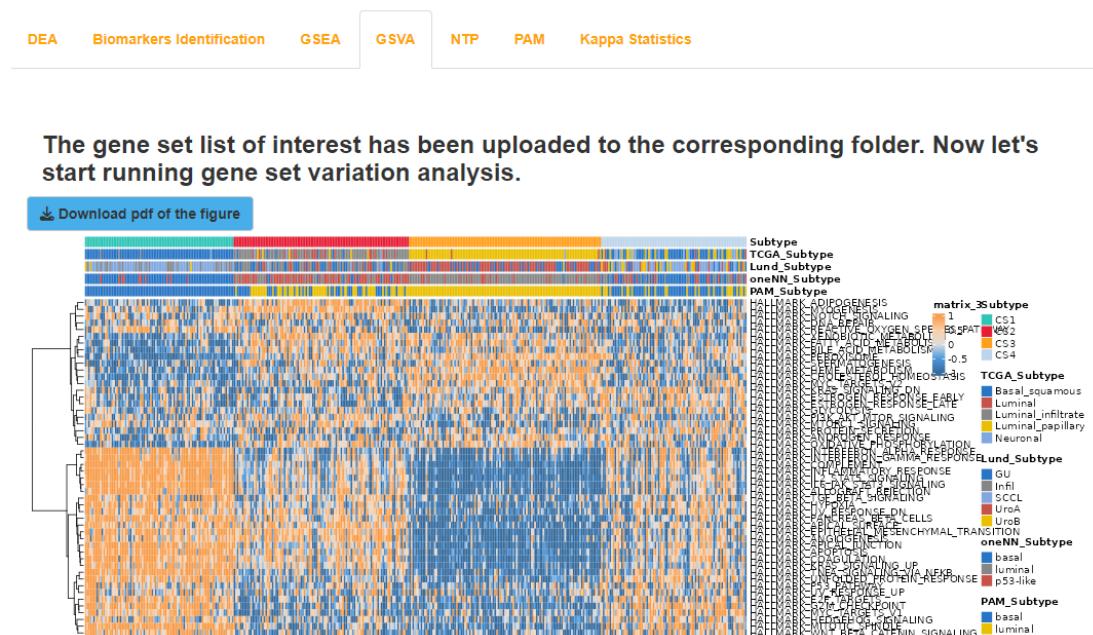


Table 1: Raw enrichment score based on the given gene set list of interest by using gsva method

Copy	CSV	Excel	Print	Show	10	entries	Search:	
TCGA-HQ-A5NE-01A		TCGA-FD-A5BT-01A		TCGA-DK-A3IQ-01A		TCGA-ZF-AA4X-01A		TCGA-GD-A301A
HALLMARK_TNFA_SIGNALING_VIA_NFKB	0.609218342433021	0.342134708499665	-0.378235223995979	-0.415108127429381	0.449228402			
HALLMARK_HYPOXIA	0.358997941976495	0.47732064628403	-0.0368311341472628	-0.453998566810624	0.461055813			
HALLMARK_CHOLESTEROL_HOMEOSTASIS	0.0323779511502909	0.00213634297680262	-0.533605330050768	0.235882964475868	-0.442244592			
HALLMARK_MITOTIC_SPINDLE	0.597602119146635	0.278140073467567	0.559150124604204	-0.582666033541275	0.552710255			
HALLMARK_WNT_BETA_CATEININ_SIGNALING	0.418776091290295	0.327385512194761	0.0453084471775966	-0.828641349786719	0.452445899			
HALLMARK_TGF_BETA_SIGNALING	0.646759875102749	0.33520775642067	0.482956598665959	-0.27711834193699	0.529793887			
HALLMARK_IL6_JAK_STAT3_SIGNALING	0.191113230567938	0.541986286185345	-0.595254171404916	-0.37553307776739	-0.149116146			
HALLMARK_DNA_REPAIR	-0.141512544244856	0.300803310529454	0.179262027213133	-0.673758559800182	-0.473760174			
HALLMARK_G2M_CHECKPOINT	0.394217500582665	0.625125334451975	-0.67334000667331	-0.798083798083763	0.107774441			
HALLMARK_APOPTOSIS	0.398781383159376	0.391502718290118	0.019325332332012	-0.448959828281674	0.192061819			

Table 2: z-scored enrichment score based on the given gene set list of interest by using gsva method

	TCGA-HQ-A5NE-01A	TCGA-FD-A5BT-01A	TCGA-DK-A3IQ-01A	TCGA-ZF-AA4X-01A	TCGA-GD-A3O-01A
HALLMARK_TNFA_SIGNALING_VIA_NFKB	1	1	-1	-1	1
HALLMARK_HYPOXIA	1	1	-0.160501448405905	-1	1
HALLMARK_CHOLESTEROL_HOMEOSTASIS	0.1318780794035	0.0283053056664498	-1	0.828850891782026	-1
HALLMARK_MITOTIC_SPINDLE	1	0.612744740885331	1	-1	1
HALLMARK_WNT_BETA_CATENIN_SIGNALING	1	0.768397855210029	0.130803092463057	-1	1
HALLMARK_TGF_BETA_SIGNALING	1	0.865828527380933	1	-0.735957675060966	1
HALLMARK_IL6_JAK_STAT3_SIGNALING	0.470033109513886	1	-1	-1	-0.41352557800
HALLMARK_DNA_REPAIR	-0.409676246660787	0.810892719475831	0.475499997467221	-1	-1
HALLMARK_G2M_CHECKPOINT	0.793470604401491	1	-1	-1	0.26131038782
HALLMARK_APOPTOSIS	1	1	0.0186317285522308	-1	0.60534965416

Showing 1 to 10 of 50 entries

Previous 1 2 3 4 5 Next

Result display for “Run gene set variation analysis” on TCGA dataset

(5). Run nearest template prediction for TCGA dataset and validation datasets

Parameter Settings: The “Steps” chooses “Run nearest template prediction”, and “Run nearest template prediction on tcga or validation cohort” chooses “TCGA”. First, we should indicate the template from the results of differential expression analysis, here “Indicate the algorithm” chooses “limma” and “Indicate the direction” chooses “up-regulated”. For the processing of expression data, we normalize the data by default (both “Indicate if the expression data should be further scaled” and “Indicate if the expression data should be further centered” choose “Yes”). After that, we indicate 1000 as “the permutations for p-value estimation”, and then choose “cosine” for “the distance measurement”. Additionally, we indicate 123456 as “an integer value for p-value reproducibility” and

utilize default settings for “The width of output figure”, “The height of output figure” and “The name of the nearest template prediction heatmap”. Finally, we click the “Process” button to run nearest template prediction. For two validation datasets, “Run nearest template prediction on tcga or validation cohort” chooses “Validation”, keep other parameters unchanged, and then click the “Process” button to run nearest template prediction in turn.

The screenshot shows a software interface with a blue header bar and a white main area. The header bar contains the text "Run nearest template prediction".

Steps

- Run differential expression analysis
- Run biomarker identification procedure
- Run gene set enrichment analysis
- Run gene set variation analysis
- Run nearest template prediction
- Run partition around medoids classifier
- Run consistency evaluation using Kappa statistics

Run nearest template prediction

This step aims to assign potential subtype labels on tcga or validation cohort using Nearest Template Prediction (NTP) based on predefined templates derived from current identified subtypes.

Run nearest template prediction on tcga or validation cohort

TCGA Validation

Choose template from 'Run biomarker identification procedure':

Indicate the algorithm

deseq2 edger limma

Indicate the direction

up-regulated down-regulated

Indicate if the expression data should be further scaled

Yes No

Indicate if the expression data should be further centered

Yes No

Indicate the permutations for p-value estimation

1000

Indicate the distance measurement

cosine pearson spearman kendall

Input an integer value for p-value reproducibility

123456

The width of output figure

5

The height of output figure

5

The name of the nearest template prediction heatmap

ntpheatmap(TCGA)

Process

Run nearest template prediction on tcga or validation cohort

TCGA Validation

Parameter settings for “Run nearest template prediction”

Result Display:

The software will generate tables to show the subtype prediction results of each sample in TCGA dataset as well as two validation datasets using “Nearest Template Prediction” (NTP). Besides, consistency between prediction results and templates in TCGA dataset as well as two validation datasets will be evaluated through heatmaps, and higher degree of consistency indicates better prediction performance.

[Download pdf of the figure](#)

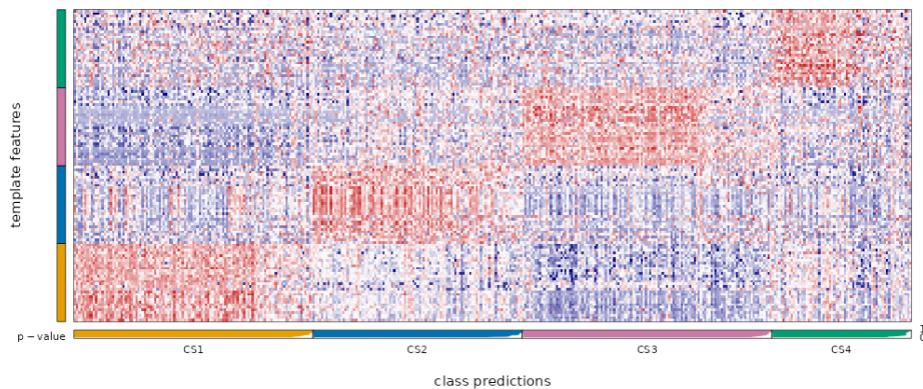


Table: the results of nearest template prediction (TCGA)

Copy	CSV	Excel	Print	Show 10 entries	Search:					
				prediction	d.CS1	d.CS2	d.CS3	d.CS4	p.value	FDR
TCGA-HQ-A5NE-01A		CS1		0.5624	0.6853	0.5774	0.5985	0.2527	0.2613	
TCGA-FD-A5BT-01A		CS1		0.5187	0.6355	0.8644	0.8229	0.001	0.0015	
TCGA-DK-A3IQ-01A		CS2		0.7343	0.3889	0.7571	0.7477	0.001	0.0015	

Result display for “Run nearest template prediction” on TCGA dataset

[Download pdf of the figure](#)

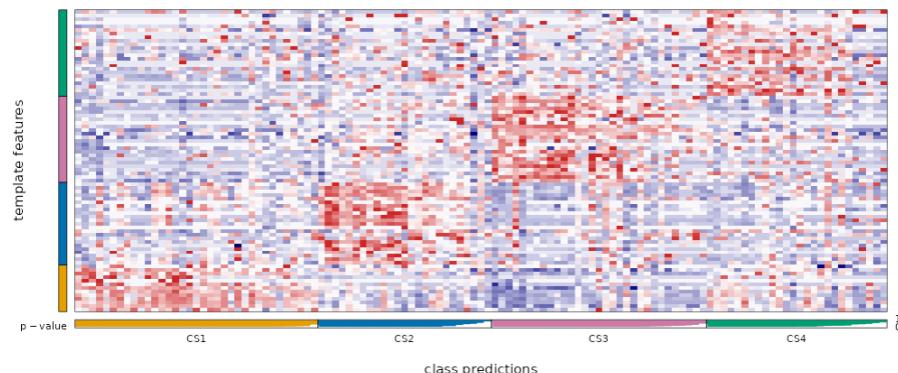


Table: the results of nearest template prediction (Validation)

Copy	CSV	Excel	Print	Show 10 entries	Search:					
				prediction	d.CS1	d.CS2	d.CS3	d.CS4	p.value	FDR
FR_1_U133_2.CEL		CS3		0.7556	0.7056	0.5346	0.692	0.002	0.0045	
FR_103_U133_2.CEL		CS3		0.8266	0.7992	0.4677	0.6782	0.001	0.0027	
FR_106_U133_2.CEL		CS2		0.7434	0.572	0.8391	0.8381	0.001	0.0027	

Result display for “Run nearest template prediction” on validation dataset (affy)

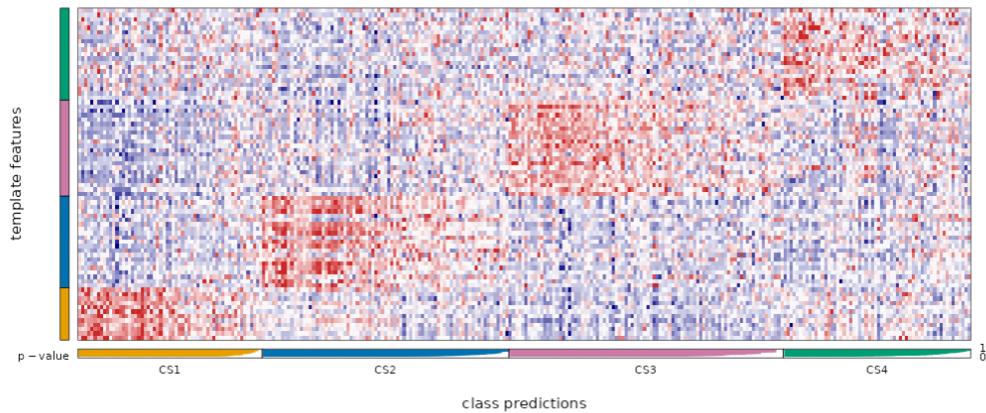
[Download pdf of the figure](#)

Table: the results of nearest template prediction (Validation)

	prediction	d.CS1	d.CS2	d.CS3	d.CS4	p.value	FDR
GSM806803	CS3	0.8511	0.7957	0.4802	0.7163	0.001	0.0029
GSM806804	CS2	0.7392	0.3537	0.6758	0.7841	0.001	0.0029
GSM806805	CS3	0.8066	0.7669	0.7664	0.7794	0.8312	0.837

Result display for “Run nearest template prediction” on validation dataset (illumina)

(6). Run partition around medoids classifier for TCGA dataset and validation datasets

Parameter Settings: The “Steps” chooses “Run partition around medoids classifier”, and “Run partition around medoids classifier on tcga or validation cohort” chooses “TCGA”. For the processing of the “NA” values in “normalized expression training data” and “normalized expression testing data”, we choose “Remove directly”. Here, the “training data” refers to TCGA dataset. If we want to predict the subtype of each sample in TCGA dataset, the “testing data” refers to TCGA dataset. If we want to predict the subtype of each sample in validation dataset, the “testing data”

refers to validation dataset. After that, we choose “No” for “Whether indicate a subset of genes to be used”. Finally, we click the “Process” button to run partition around medoids classifier. For two validation datasets, “Run partition around medoids classifier on tcga or validation cohort” chooses “Validation”, keep other parameters unchanged, and then click the “Process” button to run partition around medoids classifier in turn.

The screenshot shows a software interface with a blue header bar containing the text "Run differential expression analysis", "Run biomarker identification procedure", "Run gene set enrichment analysis", "Run gene set variation analysis", "Run nearest template prediction", "Run partition around medoids classifier" (which is highlighted in blue), and "Run consistency evaluation using Kappa statistics".

Run partition around medoids classifier

This step aims to use partition around medoids (PAM) classifier to predict potential subtype labels on tcga or validation cohort and calculate in-group proportions (IGP) statistics.

Run partition around medoids classifier on tcga or validation cohort

TCGA Validation

Indicate action for NA values in normalized expression training data

Remove directly KNN imputation

Indicate action for NA values in normalized expression testing data

Remove directly KNN imputation

Whether indicate a subset of genes to be used

Yes No

Process

Run partition around medoids classifier on tcga or validation cohort

TCGA Validation

Parameter settings for “Run partition around medoids classifier”

Result Display:

The software will generate tables to show the subtype prediction results of each sample in TCGA dataset as well as two validation datasets using “Partition around Medoids” (PAM) classifier. Besides, the values of IGP statistics for each subtype will be also displayed through tables, which reflects the similarity and reproducibility between the training data and testing data. Larger value of IGP indicates higher consistency.

DEA	Biomarkers Identification	GSEA	GSVA	NTP	PAM	Kappa Statistics
Table 1: evaluation of similarity and reproducibility of the acquired subtypes between discovery and validation cohorts using in-group proportion (IGP) statistic (TCGA)						
Copy	CSV	Excel	Print	CS1	CS2	CS3
0.897959183673469	0.762886597938144	0.970802919708029	0.671875	CS4		

Table 2: the results of partition around medoids classifier (TCGA)

Copy	CSV	Excel	Print	Show 10 entries	Search: <input type="text"/>
sampleID					prediction
TCGA-HQ-A5NE-01A					CS4
TCGA-FD-A5BT-01A					CS1
TCGA-DK-A3IQ-01A					CS2
TCGA-ZF-AA4X-01A					CS3
TCGA-GD-A3OQ-01A					CS4
TCGA-ZF-AA51-01A					CS2

Result display for “Run partition around medoids classifier” on TCGA dataset

[DEA](#) [Biomarkers Identification](#) [GSEA](#) [GSVA](#) [NTP](#) [PAM](#) [Kappa Statistics](#)

Table 1: evaluation of similarity and reproducibility of the acquired subtypes between discovery and validation cohorts using in-group proportion (IGP) statistic (Validation)

Copy	CSV	Excel	Print	CS1	CS2	CS3	CS4
				0.941176470588235	0.727272727272727	0.846153846153846	0.818181818181818

Table 2: the results of partition around medoids classifier (Validation)

Copy	CSV	Excel	Print	Show 10 entries	Search:
sampleID					prediction
FR_1_U133_2.CEL					CS3
FR_103_U133_2.CEL					CS3
FR_106_U133_2.CEL					CS2
FR_12_U133_2.CEL					CS4
FR_120_U133_2.CEL					CS1
FR_127_U133_2_2.CEL					CS3

Result display for “Run partition around medoids classifier” on validation dataset (affy)

[DEA](#) [Biomarkers Identification](#) [GSEA](#) [GSVA](#) [NTP](#) [PAM](#) [Kappa Statistics](#)

Table 1: evaluation of similarity and reproducibility of the acquired subtypes between discovery and validation cohorts using in-group proportion (IGP) statistic (Validation)

Copy	CSV	Excel	Print	CS1	CS2	CS3	CS4
				0.757142857142857	0.671875	0.905982905982906	0.628571428571429

Table 2: the results of partition around medoids classifier (Validation)

Copy	CSV	Excel	Print	Show 10 entries	Search:
sampleID					prediction
GSM806803					CS3
GSM806804					CS2
GSM806805					CS3
GSM806810					CS1
GSM806814					CS1
GSM806822					CS2

Result display for “Run partition around medoids classifier” on validation dataset (illumina)

(7). Run consistency evaluation using Kappa statistics for TCGA dataset and validation datasets

Parameter Settings: The “Steps” chooses “Run consistency evaluation using Kappa statistics”, and “Choose the results you have obtained in 'Run nearest template prediction' and 'Run partition around medoids classifier' procedures” chooses all the options. Then, we set the parameters containing “Indicate the label of the first subtype”, “Indicate the label of the second subtype”, “The width of output figure”, “The height of output figure” and “The name of the consistency heatmap” for “Run Kappa on tcga cohort (CMOIC vs NTP)”, “Run Kappa on tcga cohort (CMOIC vs PAM)”, “Run Kappa on tcga cohort (NTP vs PAM)” and “Run Kappa on validation cohort (NTP vs PAM)” in turn. Here, “CMOIC” represents the clustering results, “NTP” represents the prediction results using NTP method, and “PAM” represents the prediction results using PAM method. After that, we click the “Process” buttons to run consistency evaluation using Kappa statistics sequentially. Finally, we click the “Finish” button to finish the whole multi-omics analysis procedure.

Run Kappa on tcga cohort (CMOIC vs NTP)

Indicate the label of the first subtype

CMOIC_TCGA

Indicate the label of the second subtype

NTP_TCGA

The width of output figure

5

The height of output figure

5

The name of the consistency heatmap

constheatmap(CMOIC_VS_NTP_TCGA)

Process

Run Kappa on tcga cohort (CMOIC vs PAM)

Indicate the label of the first subtype

CMOIC_TCGA

Indicate the label of the second subtype

PAM_TCGA

The width of output figure

5

The height of output figure

5

The name of the consistency heatmap

constheatmap(CMOIC_VS_PAM_TCGA)

Process

Run Kappa on tcga cohort (NTP vs PAM)

Indicate the label of the first subtype

NTP_TCGA

Indicate the label of the second subtype

PAM_TCGA

The width of output figure

5

The height of output figure

5

The name of the consistency heatmap

constheatmap(NTP_VS_PAM_TCGA)

Process

Run Kappa on validation cohort (NTP vs PAM)

Indicate the label of the first subtype

NTP_Validation

Indicate the label of the second subtype

PAM_Validation

The width of output figure

5

The height of output figure

5

The name of the consistency heatmap

constheatmap(NTP_VS_PAM_Validation)

Process

Finish

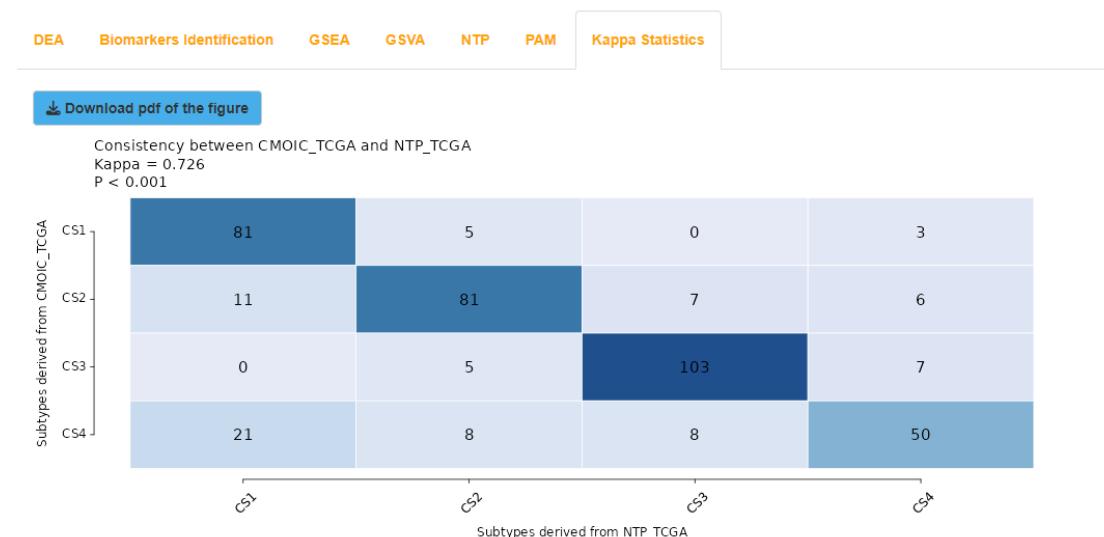
All steps in MOVICS have been finished!

Finish

Parameter settings for “Run consistency evaluation using Kappa statistics”

Result Display:

The software will calculate Kappa statistics, and then generate heatmaps to evaluate the consistency between clustering results and prediction results or the consistency between prediction results derived from NTP and PAM method. The greater the Kappa statistic is, the higher the degree of consistency is. Generally speaking, Kappa statistic that is greater than 0.4 indicates a high degree of consistency, and greater than 0.7 indicates an extremely high degree of consistency.



The process of running consistency evaluation using Kappa statistics between current subtypes derived from multi-omics clustering and NTP-predicted subtypes on tca cohort has been finished, you can download and check the figure. Now keep on the next part of 'Run consistency evaluation using Kappa statistics'.

[Download pdf of the figure](#)

Consistency between CMOIC_TCGA and PAM_TCGA
Kappa = 0.786
 $P < 0.001$

Subtypes derived from CMOIC_TCGA	CS1	CS2	CS3	CS4
CS1	77	2	0	10
CS2	6	90	7	2
CS3	0	1	114	0
CS4	15	4	16	52

Subtypes derived from PAM_TCGA

The process of running consistency evaluation using Kappa statistics between current subtypes derived from multi-omics clustering and PAM-predicted subtypes on tcga cohort has been finished, you can download and check the figure. Now keep on the next part of 'Run consistency evaluation using Kappa statistics'.

[Download pdf of the figure](#)

Consistency between NTP_TCGA and PAM_TCGA
Kappa = 0.764
 $P < 0.001$

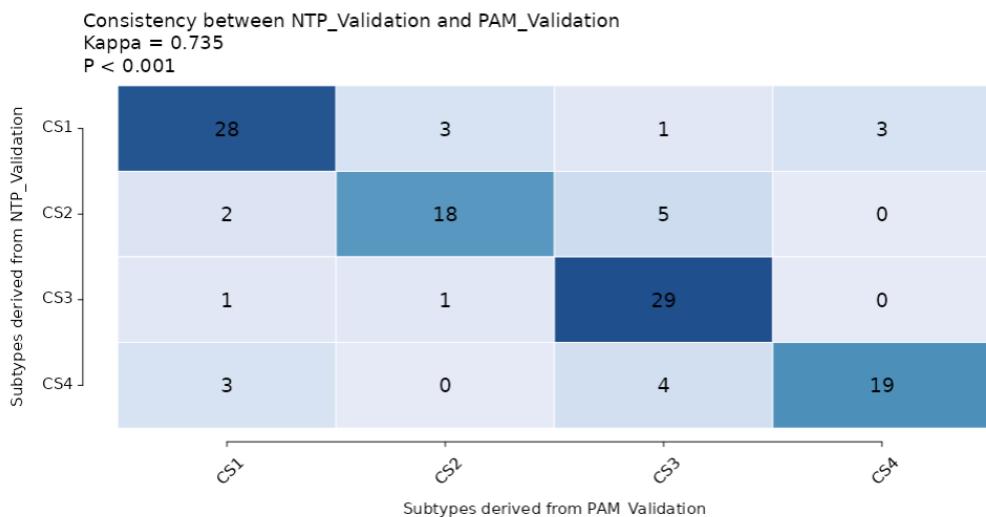
Subtypes derived from NTP_TCGA	CS1	CS2	CS3	CS4
CS1	86	8	2	17
CS2	6	83	9	1
CS3	0	4	113	1
CS4	6	2	13	45

Subtypes derived from PAM_TCGA

The process of running consistency evaluation using Kappa statistics between NTP-predicted subtypes and PAM-predicted subtypes on tcga cohort has been finished, you can download and check the figure. Now keep on the next part of 'Run consistency evaluation using Kappa statistics'.

Result display for "Run consistency evaluation using Kappa statistics" on TCGA dataset

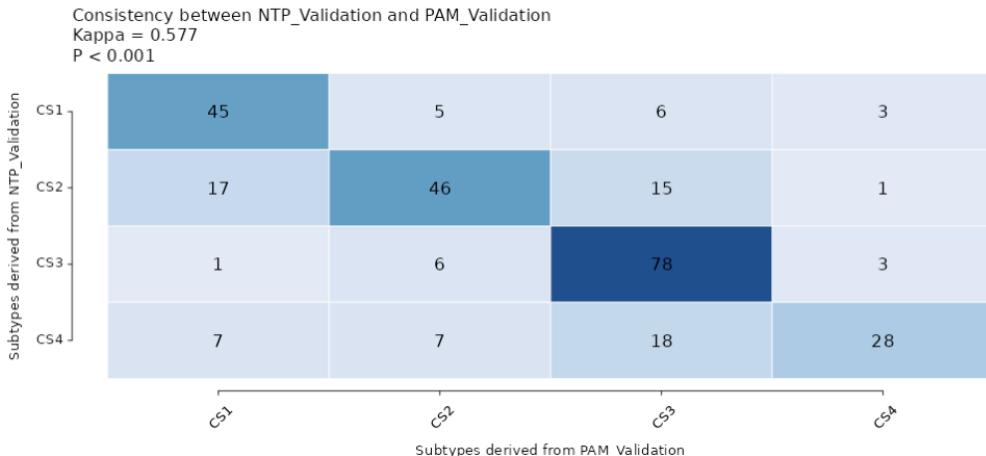
[Download pdf of the figure](#)



The process of running consistency evaluation using Kappa statistics between NTP-predicted subtypes and PAM-predicted subtypes on validation cohort has been finished, you can download and check the figure. Now all steps in 'RUN Module' have been finished!

Result display for “Run consistency evaluation using Kappa statistics” on validation dataset (affy)

[Download pdf of the figure](#)



The process of running consistency evaluation using Kappa statistics between NTP-predicted subtypes and PAM-predicted subtypes on validation cohort has been finished, you can download and check the figure. Now all steps in 'RUN Module' have been finished!

Now all steps in MOVICS have been finished, you can download and check all the tables and figures, as well as the '.RData' files. Then you can further process the obtained results, thanks for using MOVICS RShiny and we are looking forward to your valuable feedback and another visit!

Result display for “Run consistency evaluation using Kappa statistics” on validation dataset (illumina)

The fifth step: COMP Module (Validation dataset)

In order to further validate the accuracy and reliability of the clustering results, we also complete partial analyses in “COMP Module” on two external validation datasets in turn. Since the two validation datasets only include mRNA as well as clinical and survival data, we only finish “Compare survival outcome”, “Compare clinical features”, “Compare drug sensitivity” and “Compare agreement with other subtypes”.

(1). Compare survival outcome for validation dataset

Parameter Settings: “Module switching options” on the upper left of the software chooses “COMP Module”, and then the “Steps” chooses “Compare survival outcome”, and “Compare survival outcome on TCGA datasets or Validation datasets” chooses “Validation”. After that, “Model-free approaches for subtype prediction in validation cohort” chooses “NTP” and “PAM” in turn, and other parameters keep unchanged. Finally, we click the “Process” button to compare survival outcomes among subtypes derived from NTP and PAM methods for two validation datasets in turn.

Compare survival outcome

In this step, we will compare the prognosis of different subtypes based on the clustering results from 'GET Module' by Kaplan-Meier survival curve.

Pay attention: the format of survival time should be days and the values of survival status should be 0 or 1 (0: censoring; 1: event). Please make sure you provide the correct survival information first.

Compare survival outcome on TCGA datasets or Validation datasets

TCGA Validation

Model-free approaches for subtype prediction in validation cohort

NTP PAM

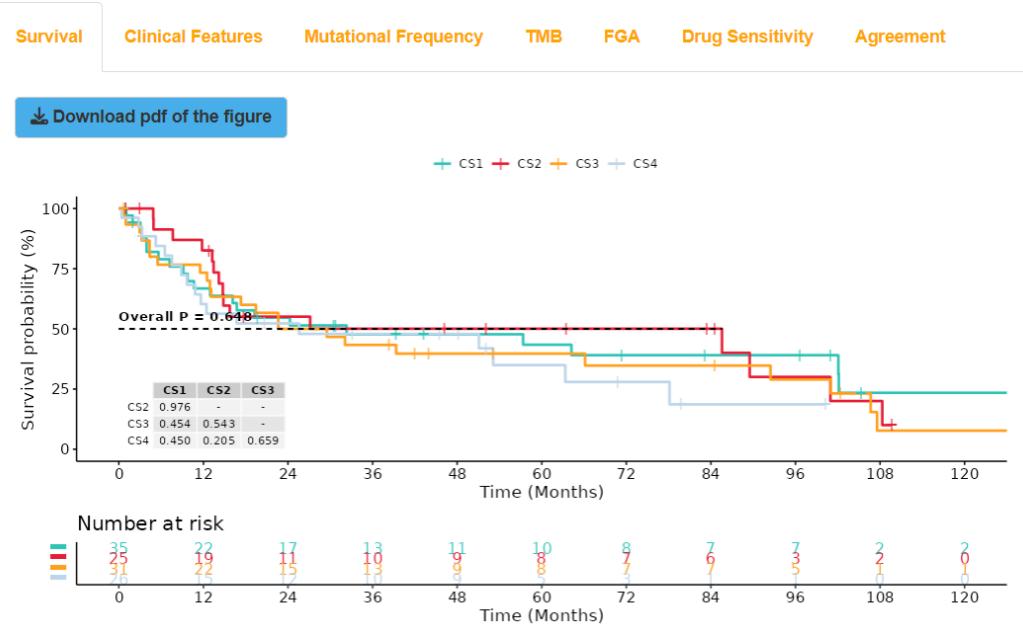
Model-free approaches for subtype prediction in validation cohort

NTP PAM

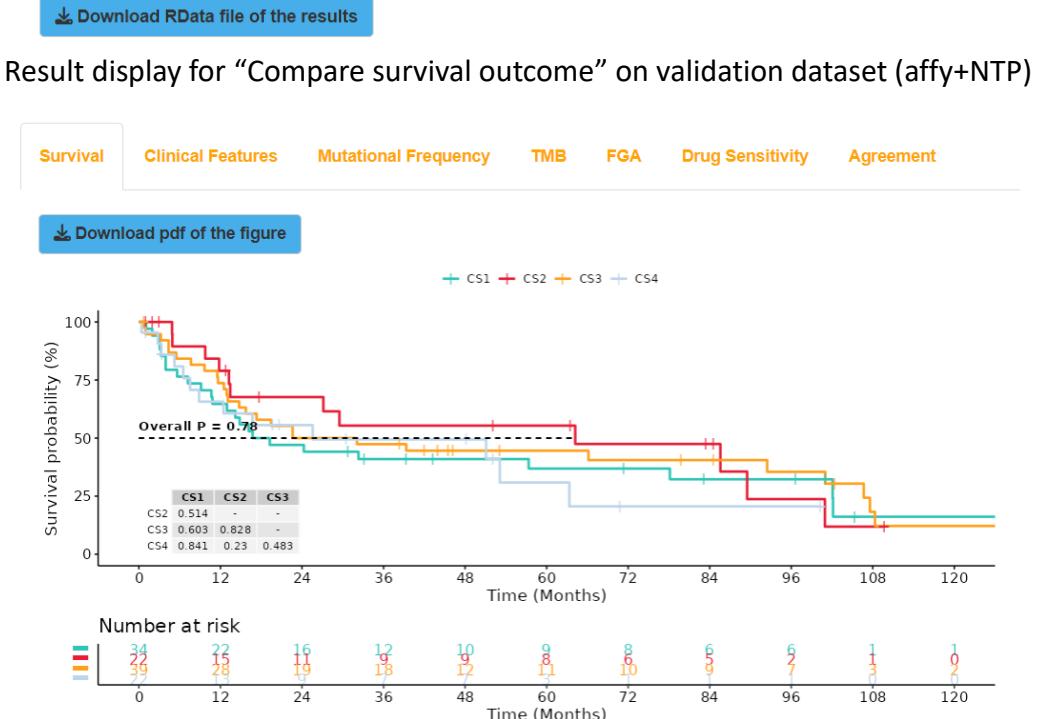
Parameter settings for "Compare survival outcome" on validation dataset

Result Display:

Like the TCGA dataset, the software will also generate Kaplan-Meier curve for each subtype derived from NTP and PAM methods for two validation datasets, and compare the survival differences among each subtype at the same time.



The process of comparing survival outcome on validation datasets has been finished, you can download and check the figure as well as the '.RData' file. Now let's turn to the next step--'Compare clinical features'.

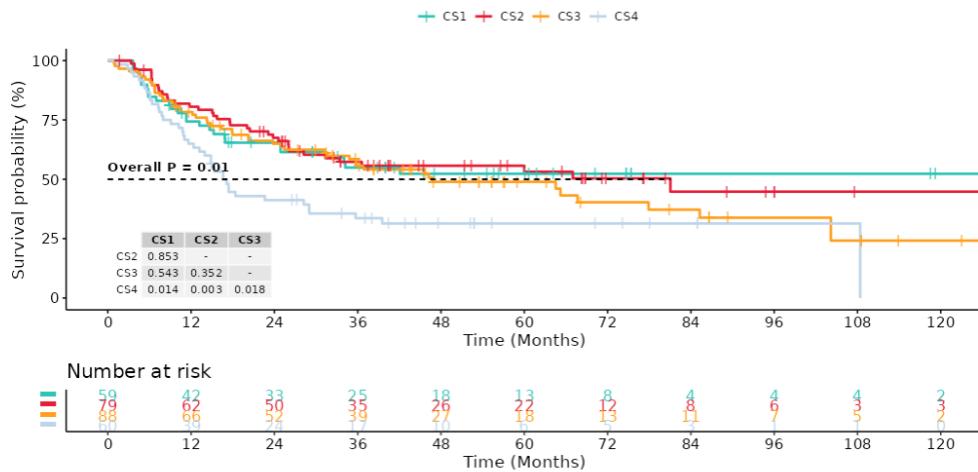


The process of comparing survival outcome on validation datasets has been finished, you can download and check the figure as well as the '.RData' file. Now let's turn to the next step--'Compare clinical features'.

Result display for "Compare survival outcome" on validation dataset (affy+PAM)

Survival Clinical Features Mutational Frequency TMB FGA Drug Sensitivity Agreement

[Download pdf of the figure](#)



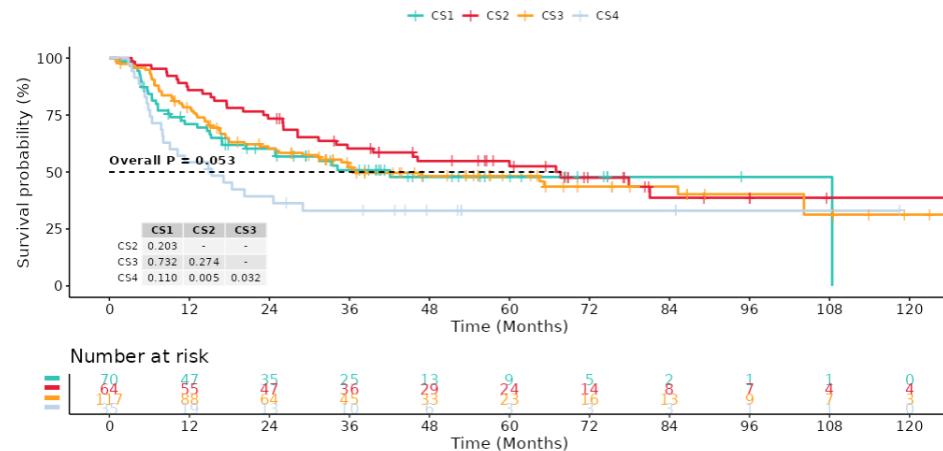
The process of comparing survival outcome on validation datasets has been finished, you can download and check the figure as well as the '.RData' file. Now let's turn to the next step--'Compare clinical features'.

[Download RData file of the results](#)

Result display for “Compare survival outcome” on validation dataset (illumina+NTP)

Survival Clinical Features Mutational Frequency TMB FGA Drug Sensitivity Agreement

[Download pdf of the figure](#)



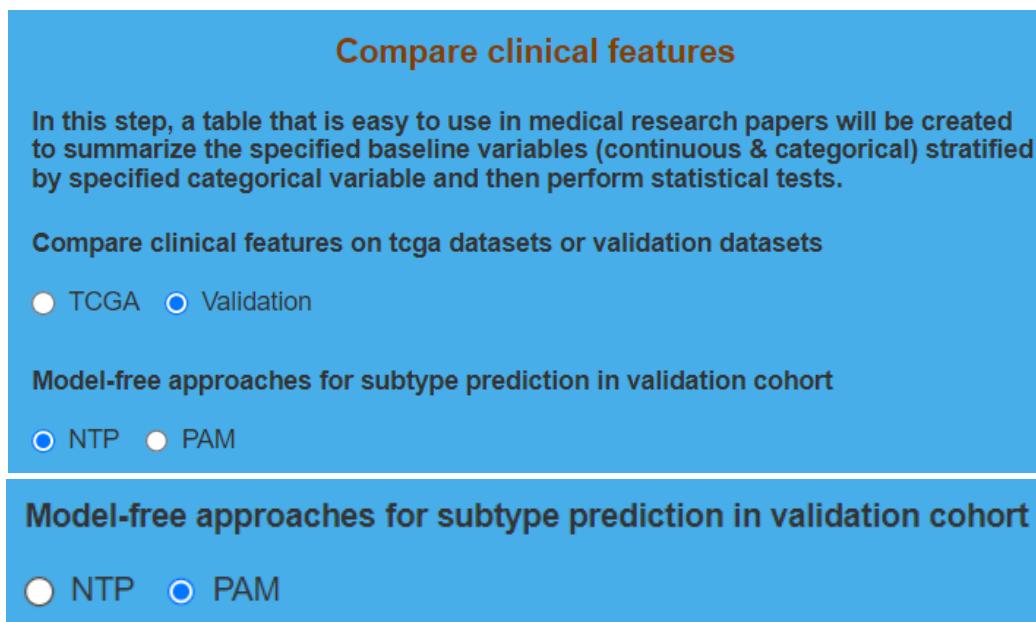
The process of comparing survival outcome on validation datasets has been finished, you can download and check the figure as well as the '.RData' file. Now let's turn to the next step--'Compare clinical features'.

[Download RData file of the results](#)

Result display for “Compare survival outcome” on validation dataset (illumina+PAM)

(2). Compare clinical features for validation dataset

Parameter Settings: The “Steps” chooses “Compare clinical features”, and “Compare clinical features on tcga datasets or validation datasets” chooses “Validation”. Other parameters keep unchanged, and then we click the “Process” button to compare clinical features among subtypes derived from NTP and PAM methods for two validation datasets in turn.



Parameter settings for “Compare clinical features” on validation dataset

Result Display:

Like the TCGA dataset, the software will also generate tables to display the comparison results of clinical features among subtypes derived from NTP and PAM methods for two validation datasets.

Survival Clinical Features Mutational Frequency TMB FGA Drug Sensitivity Agreement

Table: Summarization of clinical variables stratified by current subtypes (Validation)

		level	CS1	CS2	CS3	CS4	p	test
1	n		35	25	31	26		
2	Age (%)	<=70	21 (60.0)	11 (44.0)	23 (74.2)	10 (38.5)	0.028	exact
3		>70	14 (40.0)	14 (56.0)	8 (25.8)	16 (61.5)		
4	Gender (%)	FEMALE	11 (31.4)	7 (28.0)	6 (19.4)	3 (11.5)	0.272	exact
5		MALE	24 (68.6)	18 (72.0)	25 (80.6)	23 (88.5)		
6	T_Stage (%)	T2	8 (22.9)	7 (28.0)	12 (38.7)	5 (19.2)	0.739	exact
7		T3	19 (54.3)	13 (52.0)	12 (38.7)	15 (57.7)		
8		T4	8 (22.9)	5 (20.0)	7 (22.6)	6 (23.1)		
9	futime (median [IQR])		600.00 [195.65, 2067.35]	480.00 [390.00, 2541.93]	690.00 [259.71, 1818.79]	570.00 [239.74, 1582.07]	0.879	nonnorm
10	fustat (%)	0	13 (37.1)	10 (40.0)	7 (22.6)	9 (34.6)	0.507	exact
11		1	22 (62.9)	15 (60.0)	24 (77.4)	17 (65.4)		

Result display for “Compare clinical features” on validation dataset (affy+NTP)

Survival Clinical Features Mutational Frequency TMB FGA Drug Sensitivity Agreement

Table: Summarization of clinical variables stratified by current subtypes (Validation)

		level	CS1	CS2	CS3	CS4	p	test
1	n		34	22	39	22		
2	Age (%)	<=70	21 (61.8)	11 (50.0)	24 (61.5)	9 (40.9)	0.367	exact
3		>70	13 (38.2)	11 (50.0)	15 (38.5)	13 (59.1)		
4	Gender (%)	FEMALE	10 (29.4)	8 (36.4)	6 (15.4)	3 (13.6)	0.150	exact
5		MALE	24 (70.6)	14 (63.6)	33 (84.6)	19 (86.4)		
6	T_Stage (%)	T2	6 (17.6)	6 (27.3)	16 (41.0)	4 (18.2)	0.136	exact
7		T3	17 (50.0)	14 (63.6)	15 (38.5)	13 (59.1)		
8		T4	11 (32.4)	2 (9.1)	8 (20.5)	5 (22.7)		
9	futime (median [IQR])		548.99 [234.58, 2069.57]	683.96 [313.24, 2396.06]	690.00 [353.35, 2225.54]	555.00 [169.53, 1536.21]	0.565	nonnorm
10	fustat (%)	0	9 (26.5)	10 (45.5)	11 (28.2)	9 (40.9)	0.371	exact
11		1	25 (73.5)	12 (54.5)	28 (71.8)	13 (59.1)		

Result display for “Compare clinical features” on validation dataset (affy+PAM)

Survival	Clinical Features	Mutational Frequency	TMB	FGA	Drug Sensitivity	Agreement

Table: Summarization of clinical variables stratified by current subtypes (Validation)

		Copy	CSV	Excel	Print													
				level		CS1		CS2		CS3		CS4		p		test		
1	n			59			79			88			60					
2	Age (%)	<=70	42 (71.2)		52 (65.8)		57 (64.8)		30 (50.0)		0.087		exact					
3		>70	17 (28.8)		27 (34.2)		31 (35.2)		30 (50.0)									
4	Gender (%)	FEMALE	10 (23.3)		13 (20.6)		12 (17.6)		12 (30.0)		0.490		exact					
5		MALE	33 (76.7)		50 (79.4)		56 (82.4)		28 (70.0)									
6	T_Stage (%)	T2	38 (64.4)		52 (65.8)		45 (51.1)		31 (51.7)		0.236		exact					
7		T3	17 (28.8)		20 (25.3)		29 (33.0)		24 (40.0)									
8		T4	4 (6.8)		7 (8.9)		14 (15.9)		5 (8.3)									
9	futime (median [IQR])		900.00 [335.55, 1706.55]		992.10 [472.05, 1995.00]		987.14 [367.50, 1686.14]		507.00 [268.42, 1172.25]		0.043		nonnorm					
10	fustat (%)	0	33 (55.9)		43 (54.4)		40 (45.5)		19 (31.7)		0.020		exact					
11		1	26 (44.1)		36 (45.6)		48 (54.5)		41 (68.3)									

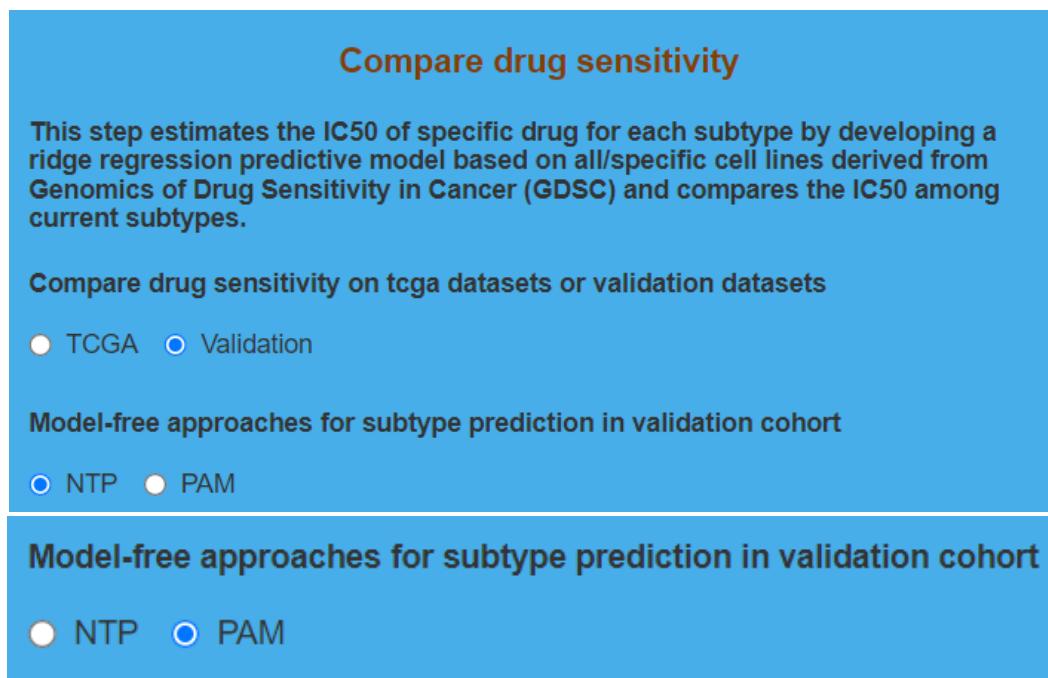
Result display for “Compare clinical features” on validation dataset (illumina+NTP)

Survival	Clinical Features	Mutational Frequency	TMB	FGA	Drug Sensitivity	Agreement												
		Copy	CSV	Excel	Print													
				level		CS1		CS2		CS3		CS4		p		test		
1	n			70			64			117			35					
2	Age (%)	<=70	42 (60.0)		42 (65.6)		75 (64.1)		22 (62.9)		0.921		exact					
3		>70	28 (40.0)		22 (34.4)		42 (35.9)		13 (37.1)									
4	Gender (%)	FEMALE	12 (23.1)		11 (23.9)		13 (14.4)		11 (42.3)		0.031		exact					
5		MALE	40 (76.9)		35 (76.1)		77 (85.6)		15 (57.7)									
6	T_Stage (%)	T2	41 (58.6)		34 (53.1)		70 (59.8)		21 (60.0)		0.325		exact					
7		T3	24 (34.3)		26 (40.6)		31 (26.5)		9 (25.7)									
8		T4	5 (7.1)		4 (6.2)		16 (13.7)		5 (14.3)									
9	futime (median [IQR])		713.55 [270.00, 1255.50]		1309.50 [720.69, 2116.40]		912.00 [372.00, 1632.00]		453.00 [190.50, 1233.00]		<0.001		nonnorm					
10	fustat (%)	0	36 (51.4)		31 (48.4)		56 (47.9)		12 (34.3)		0.411		exact					
11		1	34 (48.6)		33 (51.6)		61 (52.1)		23 (65.7)									

Result display for “Compare clinical features” on validation dataset (illumina+PAM)

(3). Compare drug sensitivity for validation dataset

Parameter Settings: The “Steps” chooses “Compare drug sensitivity”, and “Compare drug sensitivity on tcga datasets or validation datasets” chooses “Validation”. Other parameters keep unchanged, and then we click the “Process” button to compare drug sensitivity among subtypes derived from NTP and PAM methods for two validation datasets in turn.



Parameter settings for “Compare drug sensitivity” on validation dataset

Result Display:

Like the TCGA dataset, the software will also generate box-violin plots to show the IC₅₀ comparison results for drugs among subtypes derived from NTP and PAM methods for two validation datasets, and the estimated IC₅₀ of each sample in two validation datasets for drugs will be also displayed through tables.

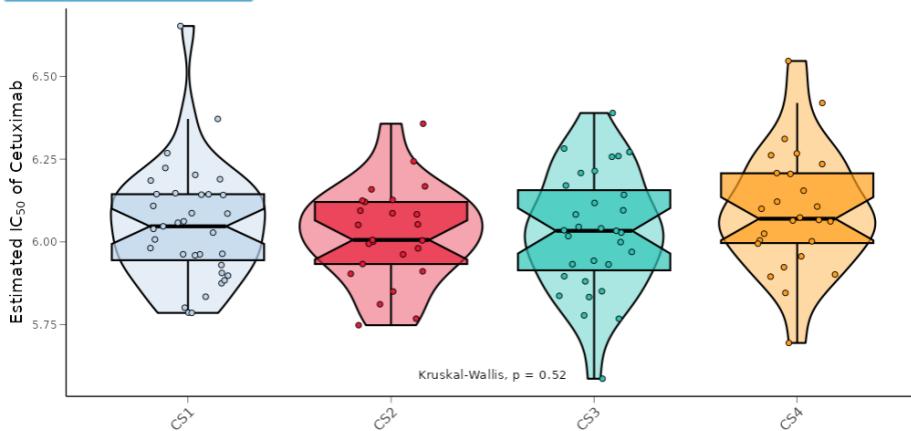
[Download pdf of the figure](#)


Table: Comparison of estimated IC50 (Validation) for Cetuximab among identified subtypes

[Copy](#) [CSV](#) [Excel](#) [Print](#) Show 10 entries

 Search:

	Est.IC50	Subtype
FR_120_U133_2.CEL	5.78698094205484	CS1
FR_14_U133_2.CEL	6.05795247622992	CS1
FR_181_U133_2.CEL	5.96264550272376	CS1

Result display for “Compare drug sensitivity” on validation dataset (affy+NTP+ Cetuximab)

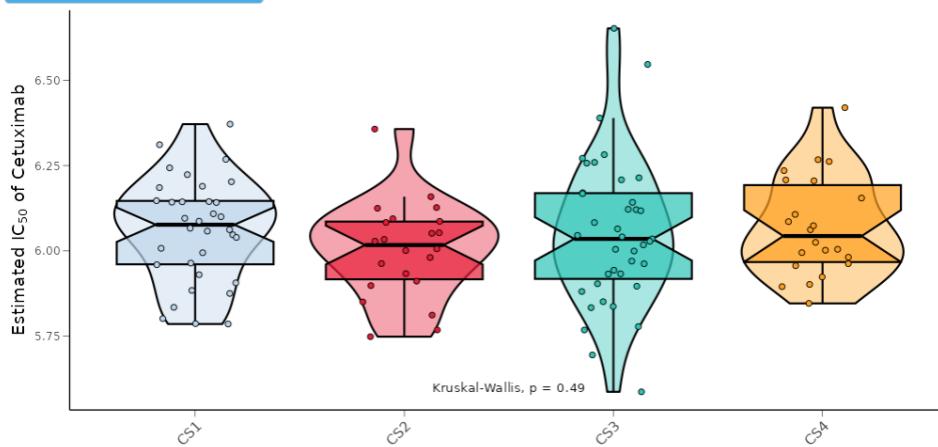
[Download pdf of the figure](#)


Table: Comparison of estimated IC50 (Validation) for Cetuximab among identified subtypes

[Copy](#) [CSV](#) [Excel](#) [Print](#) Show 10 entries

 Search:

	Est.IC50	Subtype
FR_120_U133_2.CEL	1.87491573880593	CS1
FR_14_U133_2.CEL	2.02567634852706	CS1
FR_181_U133_2.CEL	1.99193585977775	CS1

Result display for “Compare drug sensitivity” on validation dataset (affy+PAM+ Cetuximab)

Survival	Clinical Features	Mutational Frequency	TMB	FGA	Drug Sensitivity	Agreement
----------	-------------------	----------------------	-----	-----	------------------	-----------

[Download pdf of the figure](#)

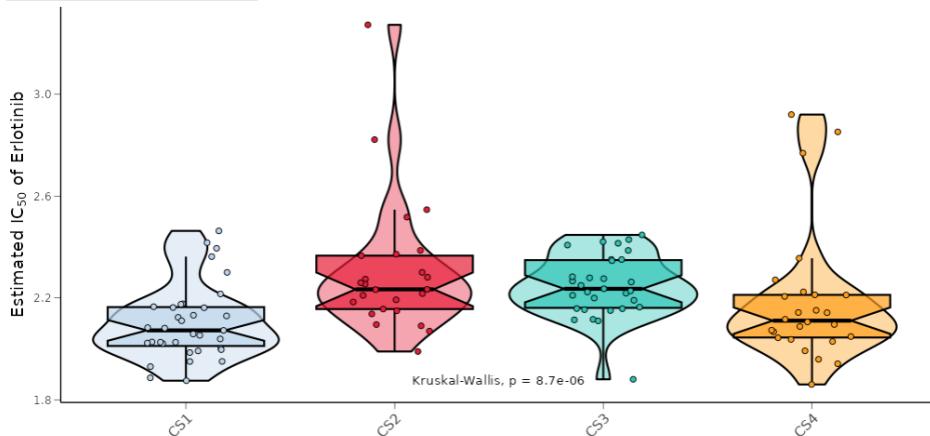


Table: Comparison of estimated IC₅₀ (Validation) for Erlotinib among identified subtypes

Copy CSV Excel Print Show 10 entries Search:

	Est.IC50	Subtype
FR_120_U133_2.CEL	1.87491573880593	CS1
FR_14_U133_2.CEL	2.02567634852706	CS1
FR_181_U133_2.CEL	1.99193585977775	CS1

Result display for “Compare drug sensitivity” on validation dataset (affy+NTP+ Erlotinib)

Survival	Clinical Features	Mutational Frequency	TMB	FGA	Drug Sensitivity	Agreement
----------	-------------------	----------------------	-----	-----	------------------	-----------

[Download pdf of the figure](#)

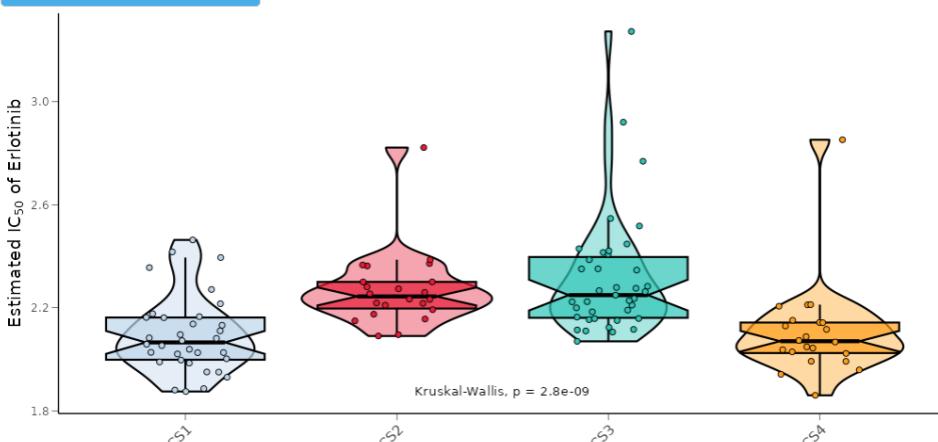


Table: Comparison of estimated IC₅₀ (Validation) for Erlotinib among identified subtypes

Copy CSV Excel Print Show 10 entries Search:

	Est.IC50	Subtype
FR_120_U133_2.CEL	1.87491573880593	CS1
FR_14_U133_2.CEL	2.02567634852706	CS1
FR_181_U133_2.CEL	1.99193585977775	CS1

Result display for “Compare drug sensitivity” on validation dataset (affy+PAM+ Erlotinib)

Survival Clinical Features Mutational Frequency TMB FGA Drug Sensitivity Agreement

[Download pdf of the figure](#)

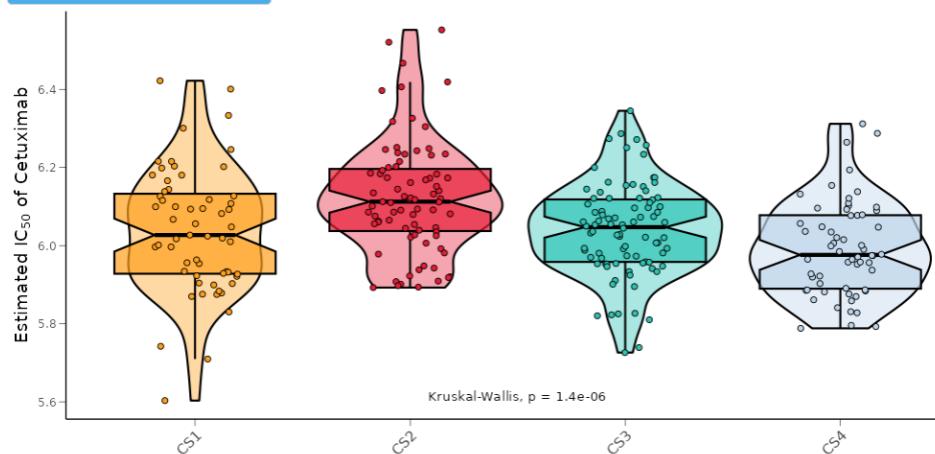


Table: Comparison of estimated IC50 (Validation) for Cetuximab among identified subtypes

Copy CSV Excel Print Show 10 entries Search:

	Est.IC50	Subtype
GSM806810	6.05597392190824	CS1
GSM806864	5.9341718070258	CS1
GSM806868	6.02420552926831	CS1

Result display for “Compare drug sensitivity” on validation dataset (illumina+NTP+Cetuximab)

Survival Clinical Features Mutational Frequency TMB FGA Drug Sensitivity Agreement

[Download pdf of the figure](#)

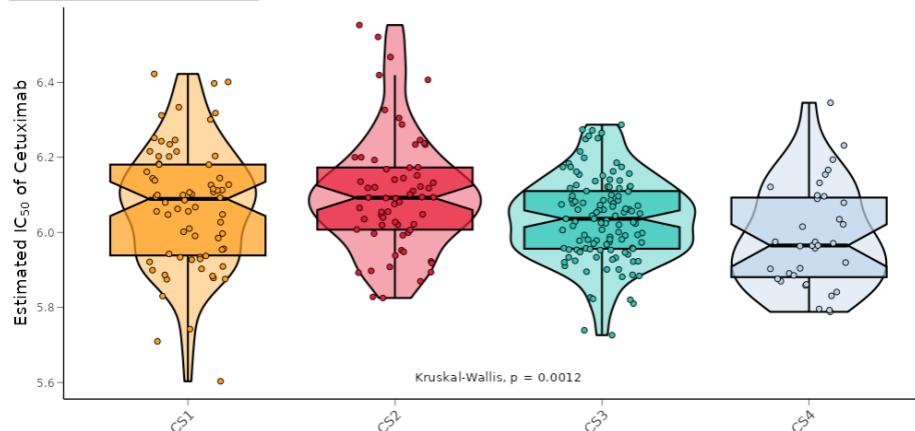


Table: Comparison of estimated IC50 (Validation) for Cetuximab among identified subtypes

Copy CSV Excel Print Show 10 entries Search:

	Est.IC50	Subtype
GSM806810	6.05597392190822	CS1
GSM806814	6.06247363049068	CS1
GSM806864	5.9341718070258	CS1

Result display for “Compare drug sensitivity” on validation dataset (illumina+PAM+Cetuximab)

Survival Clinical Features Mutational Frequency TMB FGA Drug Sensitivity Agreement

[Download pdf of the figure](#)

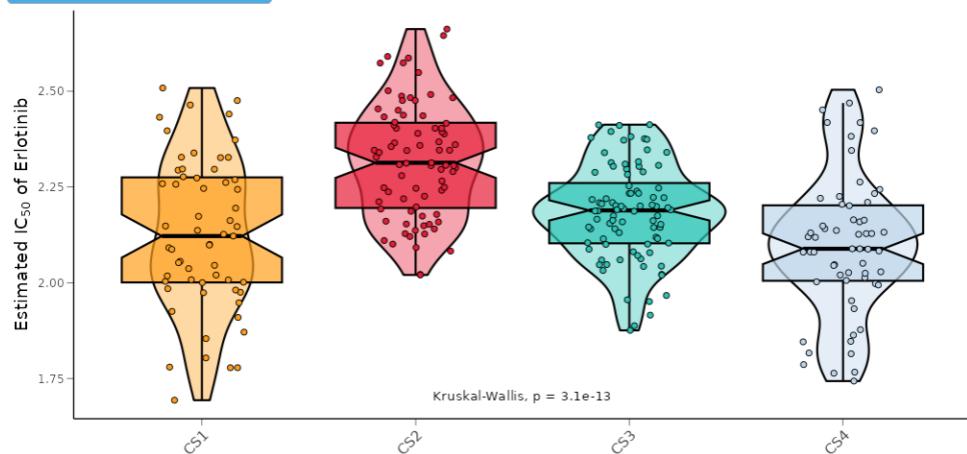


Table: Comparison of estimated IC50 (Validation) for Erlotinib among identified subtypes

Copy CSV Excel Print Show 10 entries Search:

	Est.IC50	Subtype
GSM806810	2.00030403109302	CS1
GSM806864	2.007819835814	CS1
GSM806868	2.29641779049512	CS1

Result display for “Compare drug sensitivity” on validation dataset (illumina+NTP+ Erlotinib)

Survival Clinical Features Mutational Frequency TMB FGA Drug Sensitivity Agreement

[Download pdf of the figure](#)

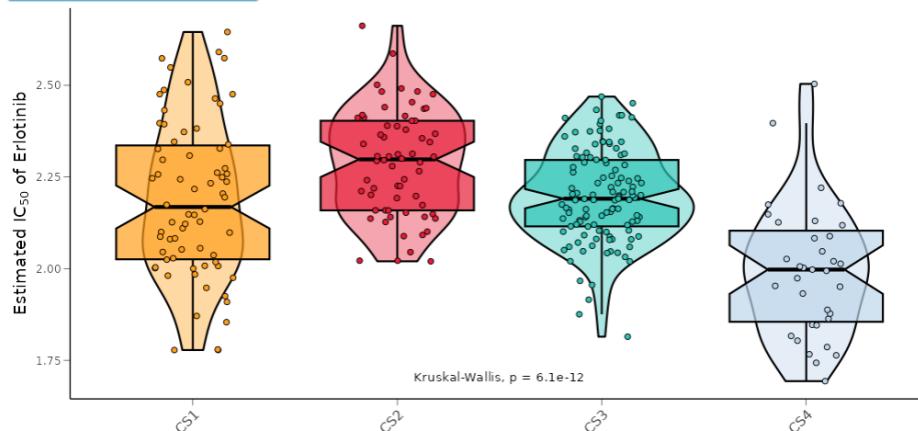


Table: Comparison of estimated IC50 (Validation) for Erlotinib among identified subtypes

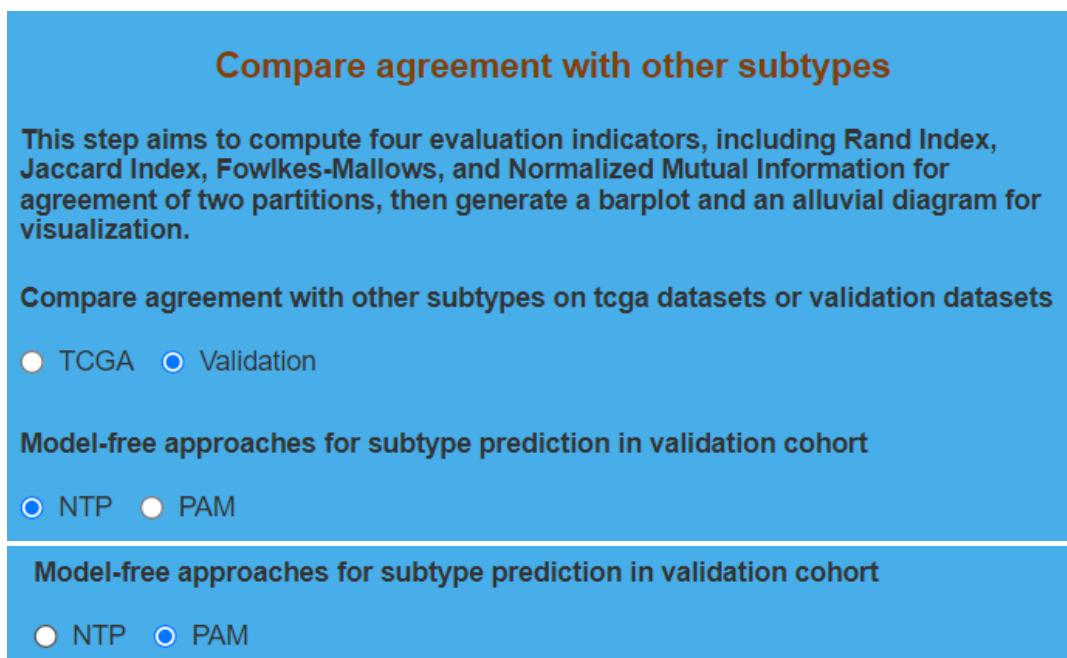
Copy CSV Excel Print Show 10 entries Search:

	Est.IC50	Subtype
GSM806810	2.00030403109301	CS1
GSM806814	2.11024789954329	CS1
GSM806864	2.00781983581398	CS1

Result display for “Compare drug sensitivity” on validation dataset (illumina+PAM+ Erlotinib)

(4). Compare agreement with other subtypes for validation dataset

Parameter Settings: The “Steps” chooses “Compare agreement with other subtypes”, and “Compare agreement with other subtypes on tcga datasets or validation datasets” chooses “Validation”. Other parameters keep unchanged, and then we click the “Process” button to compare agreement between subtypes derived from NTP as well as PAM methods and other traditional subtypes for two validation datasets in turn.

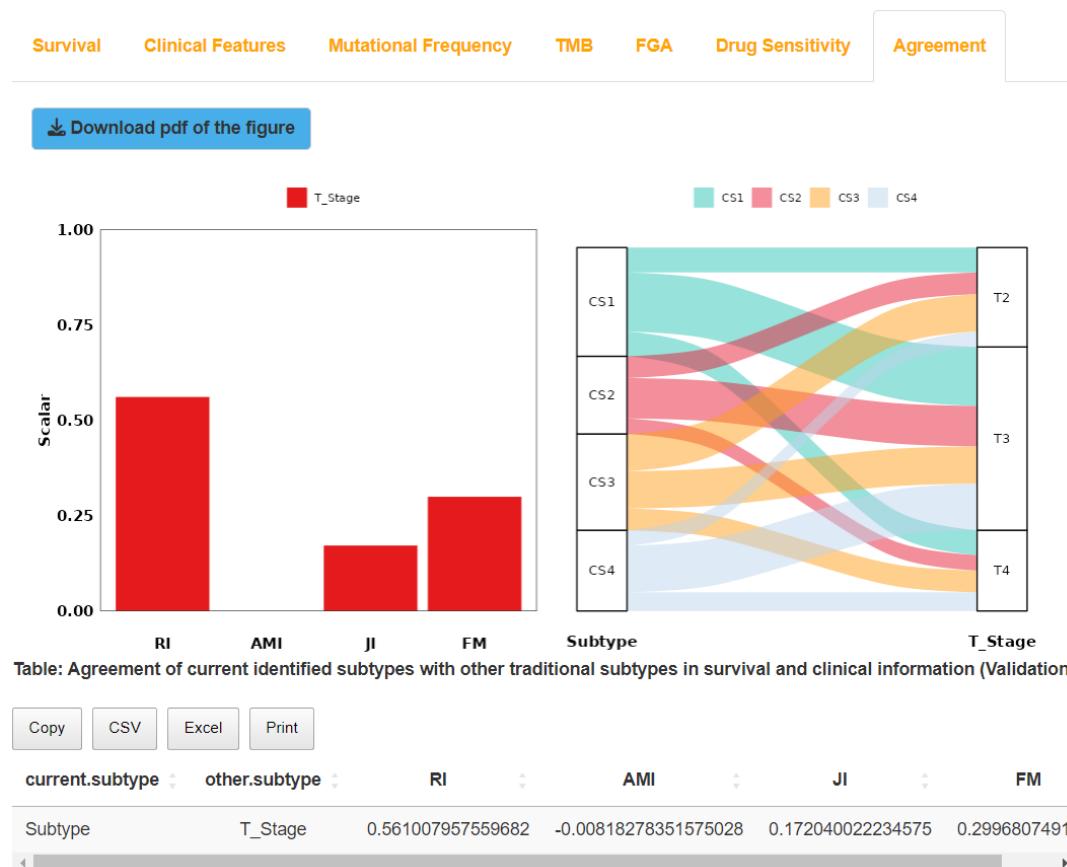


Parameter settings for “Compare agreement with other subtypes” on validation dataset

Result Display:

Like the TCGA dataset, the software will also generate alluvial diagrams to compare agreement between subtypes derived from NTP as well as PAM methods and other traditional subtypes for two validation datasets.

Additionally, bar plots and tables are also generated for two validation datasets to display the values of four statistical indicators which are utilized to evaluate the agreement, including Rand Index (RI), Adjusted Mutual Information (AMI), Jaccard Index (JI), and Fowlkes-Mallows (FM).



Result display for “Compare agreement with other subtypes” on validation dataset (affy+NTP)

Survival Clinical Features Mutational Frequency TMB FGA Drug Sensitivity Agreement

[Download pdf of the figure](#)

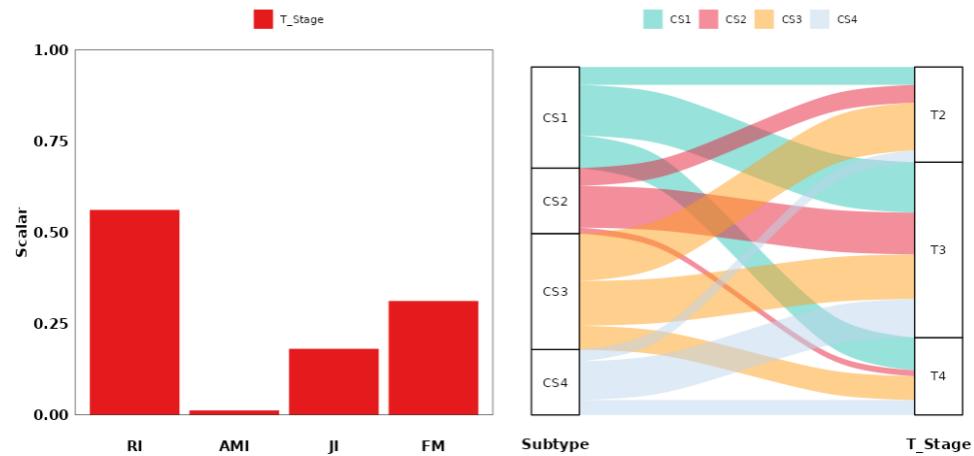


Table: Agreement of current identified subtypes with other traditional subtypes in survival and clinical information (Validation)

[Copy](#) [CSV](#) [Excel](#) [Print](#)

current.subtype	other.subtype	RI	AMI	JI	FM
Subtype	T_Stage	0.561450044208665	0.0129776612333448	0.181518151815182	0.312293391738

Result display for “Compare agreement with other subtypes” on validation dataset
(affy+PAM)

Survival Clinical Features Mutational Frequency TMB FGA Drug Sensitivity Agreement

[Download pdf of the figure](#)

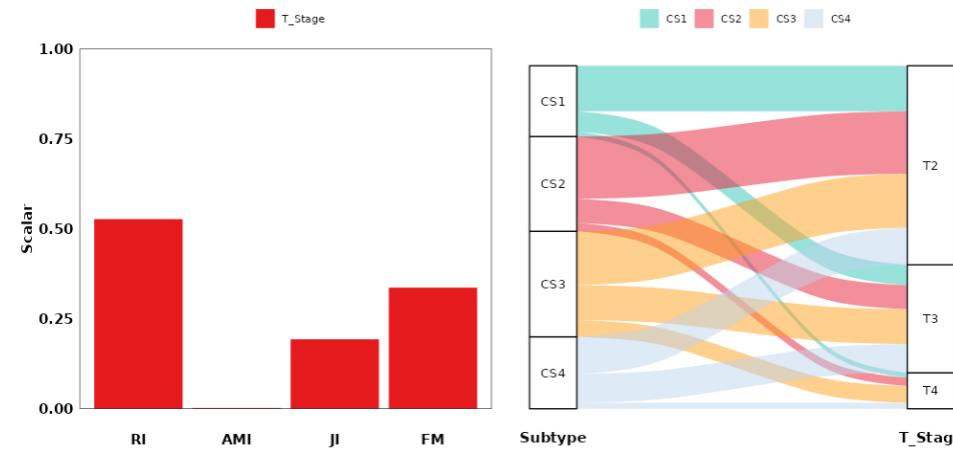


Table: Agreement of current identified subtypes with other traditional subtypes in survival and clinical information (Validation)

[Copy](#) [CSV](#) [Excel](#) [Print](#)

current.subtype	other.subtype	RI	AMI	JI	FM
Subtype	T_Stage	0.526806526806527	0.00288124532722087	0.193265007320644	0.33657001478

Result display for “Compare agreement with other subtypes” on validation dataset
(illumina+NTP)

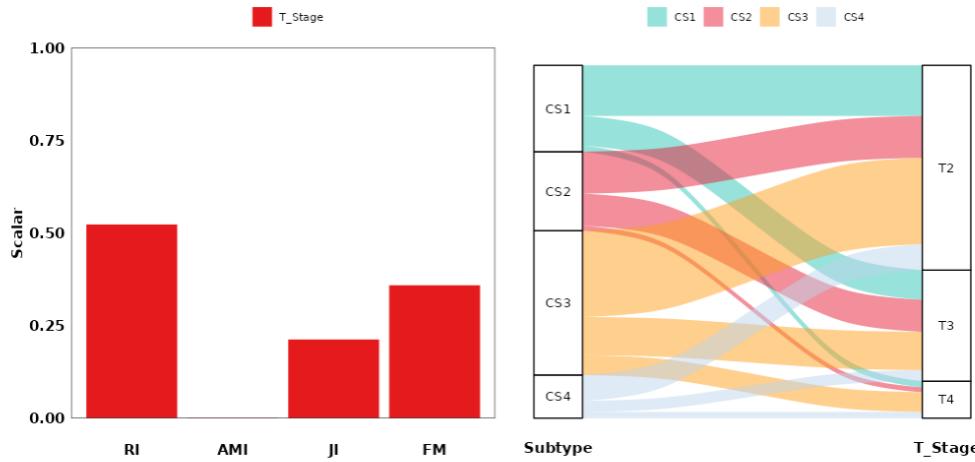
[Survival](#)[Clinical Features](#)[Mutational Frequency](#)[TMB](#)[FGA](#)[Drug Sensitivity](#)[Agreement](#) [Download pdf of the figure](#)

Table: Agreement of current identified subtypes with other traditional subtypes in survival and clinical information (Validation)

Copy	CSV	Excel	Print		
current.subtype	other.subtype	RI	AMI	JI	FM
Subtype	T_Stage	0.523027849343639	0.00119073193240721	0.212772850605435	0.35897943471

Result display for “Compare agreement with other subtypes” on validation dataset
(illumina+PAM)

So far, the use of the software through the example of bladder cancer has been all introduced. Now, let's try to perform multi-omics analysis on other cancers through this software, and we welcome you to point out errors or give valuable advice on our software.