

# IW-SRAHC User Manual

*Version: 1.0*

*Date: February 10, 2021*

## 1 How to run the web server

This web-tool is mainly finished in Python and HTML, a Linux operating system server is recommended. This deploy instruction is finished on CentOS 7. In other environment and distribution versions of Linux, commands may be a little different. To run the web server, Python3 and MongoDB must be installed correctly.

Make sure you have Python3 installed. The minimum version of Python is 3.7.x, you can confirm the version on your machine using this command:

```
python3 --version
```

then download the source code of IW-SRAHC from GitHub. Choose a directory where you wish to run the server, and clone source code by using this command:

```
git clone https://github.com/wylapp/iwsrahc
```

and change into the directory iwsrahc.

```
cd iwsrahc
```

Create a virtual environment for the server and activate it.

```
python3 -m virtualenv iwenv  
source iwenv/bin/activate
```

If an error occurs saying that python do not have virtualenv module, that's very normal, Install one by using this command:

```
pip3 install virtualenv
```

and the whole source code tree of IW-SRAHC is shown in Figure S1. Then install all required packages.



**Figure S1:** Source code tree of IW-SRAHC

```
pip3 install -r requirements.txt
```

Our embedded TCGA dataset was placed in `app/data`, however, it is too large and cannot be uploaded to GitHub. So you need to download it from our server and extract it into the `app/data` directory. You can download the dataset from <https://generalapi.top/dataset>

To run the server, there's still one step, the MongoDB server needs more configurations. A GUI management software for MongoDB will simplify this work flow. First, you need to create a new database called `cluster_record`, then, create two collections called `tcga_list` and `cancers`. This two collections are the index of the embedded TCGA dataset. Download the compressed json files from <https://generalapi.top/mongoindex> and extract it. Import two json files into the collection `cancers` and collection `tcga_list`.

In `utilities/dburi.py`, there's an URI (universal resource index) of the MongoDB like this:

```
pymongo.MongoClient("mongodb://server:port/")
```

you need to edit this URI to fit your DB server's configuration before running the web server. To run web server, use the following command:

```
python3 run.py
```

the default URL of the web-tool is `http://localhost:5000`, this can be changed in `run.py`.

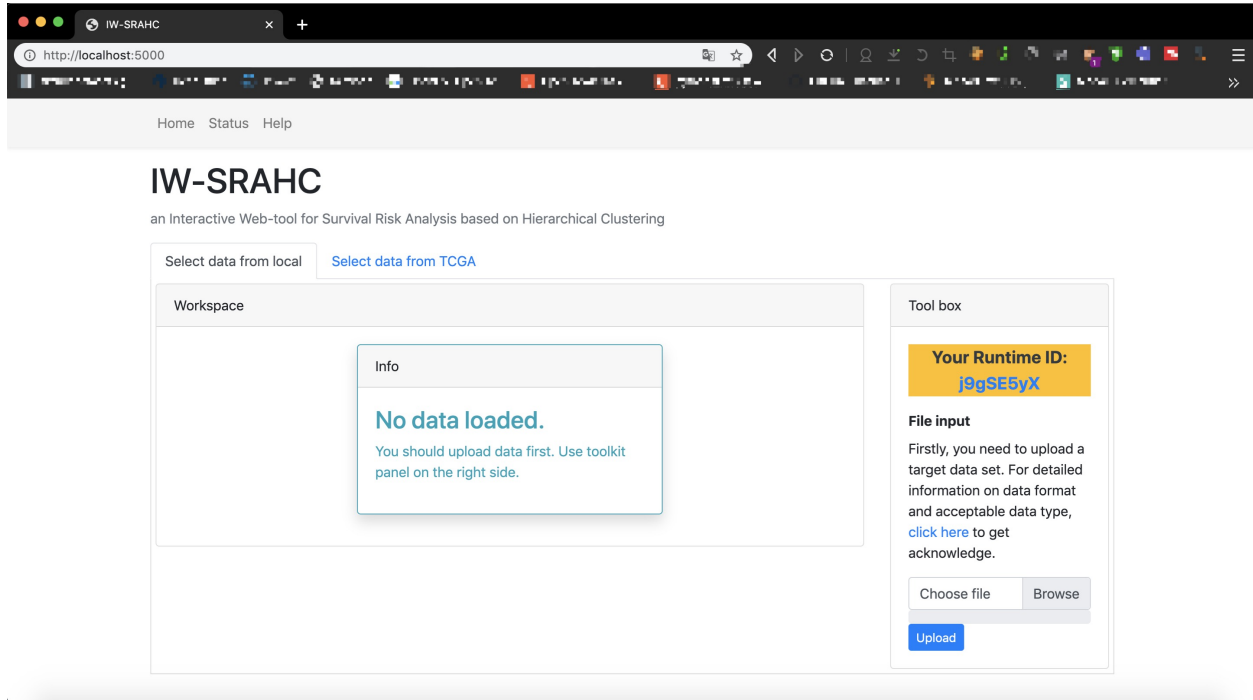
## 2 Web tool usage instructions

IW-SRAHC is a web tool, which means all your operations can be done just in the web browser. Figure S2 is the overall look of IW-SRAHC. There are three items in the navigation bar:

- **Home** The homepage of IW-SRAHC
- **Status** Includes a history of all runtime's status

- **Help** The user manual of IW-SRAHC

And a main panel contains a workspace on the left and a toolbox on the right laying below the headline.



**Figure S2:** An overall look of IW-SRAHC

## 2.1 Dataset preparation

For users who use TCGA dataset which has been embedded in this tool, this part is not important. For users who want to upload their own dataset from local, please notice the following tips:

- Currently, this tool only supports excel datasheet, you can upload a file with .xlsx extension only.
- Rows denote samples and columns denote samples' status and features.
- First column is the name of each sample, and the first row is the name of status, lifetime and features. **Leave the cell (row1, col1) as blank, samples and features must have their names.**
- Cells from (row2, col2) and behind are treated as data areas. The second column is the sample's survival status, where 0 indicates censored and 1 indicates dead. The third column is the sample's survival time, and the suggested units are days and weeks. The fourth and

	A	B	C	D	E	F
1		status	lifetime	feature1	feature2	feature3
2	sample1	0	112	1.887		
3	sample2	1	19	0.116		
4	sample3	1	21			
5	sample4	0				
6	sample5					
7	sample6					
8	sample7					
9	sample8					
10	sample9					
11	...					
12						

**Figure S3:** Format of the dataset

subsequent columns are the gene expression values or other feature values.

- To change a dataset, you need to refresh the web page.

Figure S3 gives an instant view of what the dataset looks like.

## 2.2 Feature selection

Once your data is uploaded to the server or a TCGA dataset is chosen, a list of all available features will be shown in a table, you can select features for Cox regression.

- At least one feature is selected. Click the row or the checkbox works equally.
- The column "item name" is the search field, you can use the search box at the top of the table.
- "Confirm" button will submit your selections; "show selections" button will show your selections; to clear selected items, use "clear all" button.
- To view results of projective clustering on different features, just reselect features and click the "confirm" button. **This operation will erase the previous results, so please make sure that you've saved all important results.**

The whole procedure of feature selection is shown in Figure S4.

Workbench

Confirm Show selections Clear all

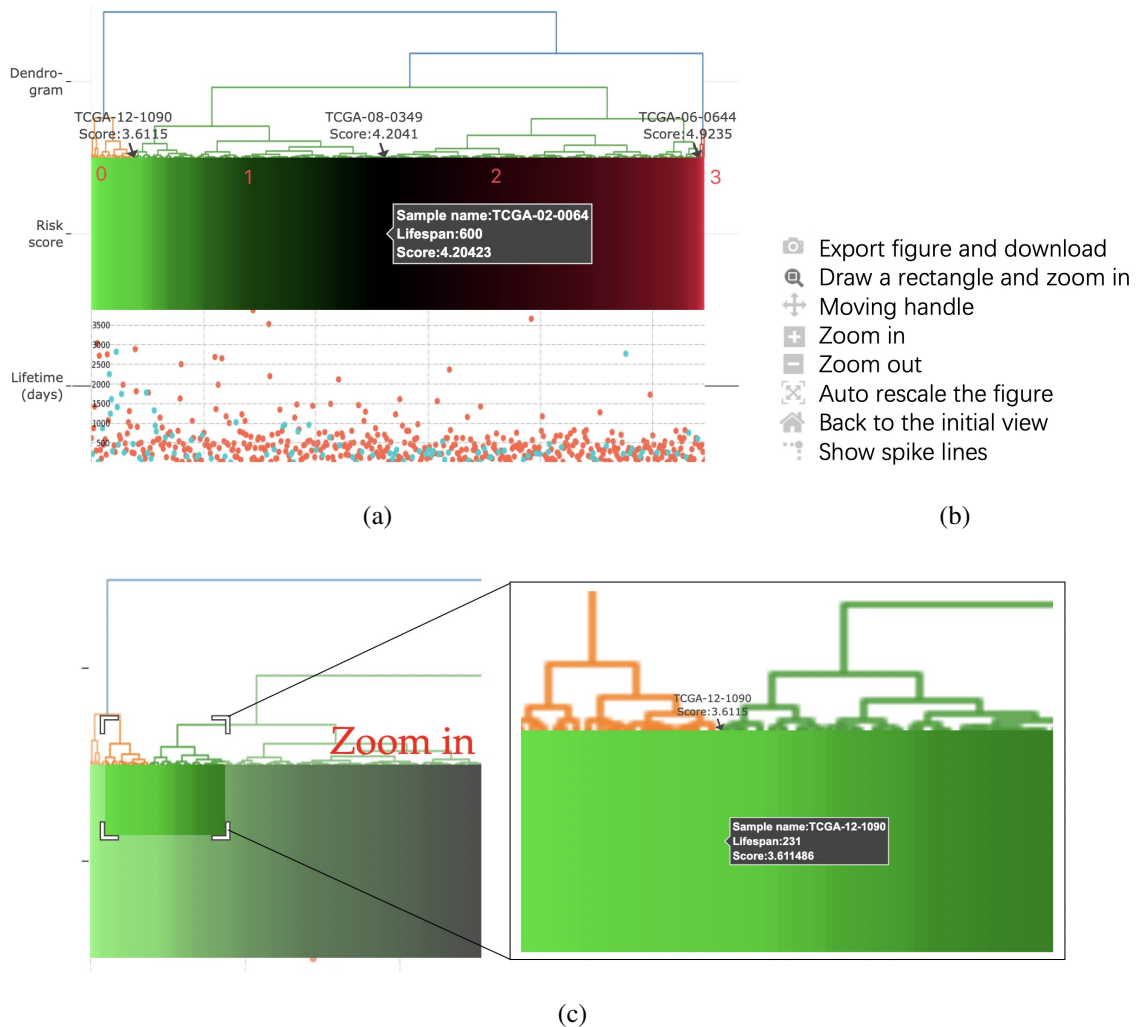
222

	Item ID	Item Name
<input checked="" type="checkbox"/>	125	hsa-miR-222
<input type="checkbox"/>	222	hsa-miR-425-5p

**Figure S4:** How to select target features

## 2.3 Interactive sample division

The result of projective clustering is shown on the webpage in an active figure called Lifetime & riskscore map, users need to interactively divide samples into groups according to not only the dendrogram branches but also the boundaries of different colored blocks in the heatmap and the lifetime scatter plot.



**Figure S5:** Lifetime & riskscore map and cursors

If you hover the mouse pointer over the heat map, a pop-up prompt containing the sample's name, risk score and survival time shows. Sample division can be done by clicking the mouse in the right place, by doing so, you are actually annotating the split line between groups (Figure S5(a)). There are several kinds of cursors and control buttons (Figure S5(b)) to give you a further ability to view this figure and decide where to split. For example, in Figure S5(c), you can draw a

rectangle on the map using the second cursor and zoom into that area to have a closer look of the dendrogram's branches.

After a division of groups is done, click "confirm separations" button. Here's something that you should know:

- **Groups are numbered from left to right, starting from 0.**
- At least split into two groups.
- To view Kaplan-Meier curves of different grouping cases, click "clear all" button first, then split samples into groups in a different way and click "confirm separations" button. **This operation will erase previous results on the web page, so please make sure that you've saved all important results.**

After the re-clustering, a Kaplan-Meier curve will be shown. Click right mouse button on the Kaplan-Meier curve, and choose "save as image" on the context menu to save the figure.