A case study using SSP

Content

1. How to start?

A live version of SSP is hosted at http://web.biotcm.net/SSP/.

Figure S1 Homepage of SSP

If you want to deploy SSP on your own server, please visit https://gitee.com/auptz/benchmark-ss and download all files and run the "app.R" in Rstudio. Notably, essential packages and software listed in **Table S2** are need to be installed before you run.

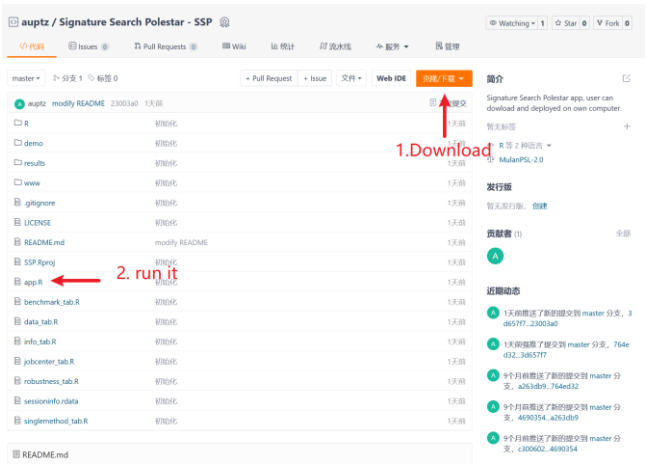


Figure S2 Source code of SSP in Gitee

2. Prepare input data

SSP require Two types of data:

A disease signature, header with Gene and log2FC (**Figure S3A**) and Drug annotations for AUC and ES evaluation (**Figure S3B** and **Figure S3C**). Notably, we provide two types of drug annotation methods (AUC and ES), and users must use at least one method to assess the performance of SS. It is impractical to annotate all drugs; however, the more annotations obtained, the more accurate the results will be. SSP provide demo files in Data Page (**Figure S3D**).

Figure S3 Demo input in SSP

3. Operation in Benchmark module

In this module, you can evaluate signature search (SS) methods based on signature and well-annotated drugs in LINCS L1000.

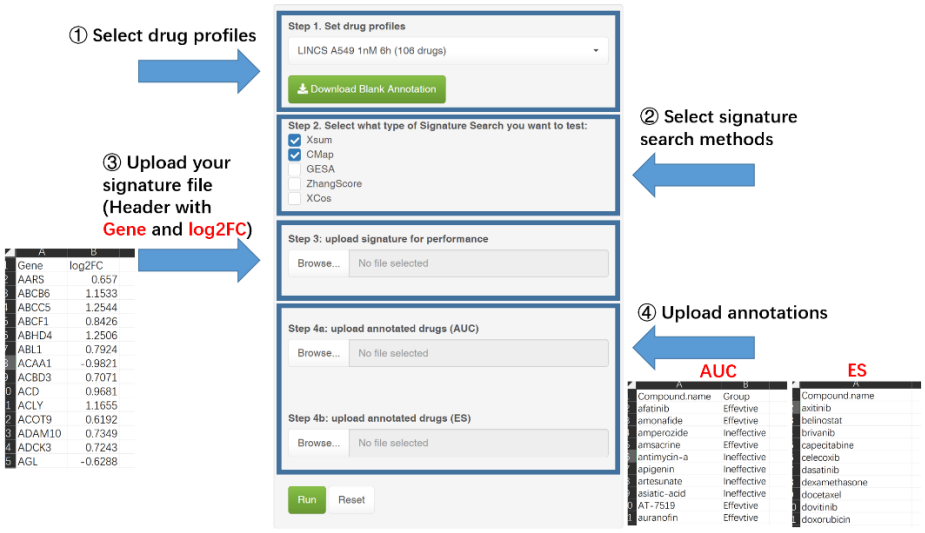


Figure S4 Workflow in Benchmark module

The Benchmark module, as shown in **Figure S4**, requires the following steps:

- Select Drug profiles (LINCS L1000)
- Select SS methods to test (at least two)
- A signature (header with gene and logFC) to perform test
- Drug annotations which user can download blank annotation table of drugs by click the download button

If you have annotations for effective or ineffective LINCS L1000 drugs (generally based on whether IC50 < 10μM), you can upload them into step 4a. We will then calculate the drug scores and rank them based on the confusion matrix using the Area Under Curve (AUC), the higher AUC indicates better performance. If you have annotations for effective LINCS L1000 drugs (generally

based on Clinical info, such as FDA-approved drugs), you can upload them into step 4b. We will then calculate the drug scores and perform drug set enrichment score (ES), the lower ES indicates better performance. [1]
In addition, you can also down load blank annotation and prepare drug on yourself.

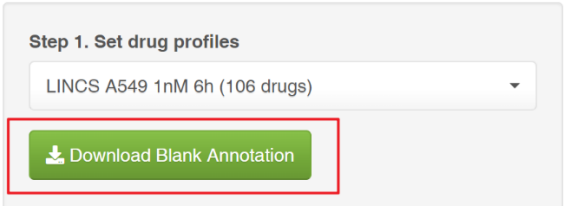


Figure S5 Blank annotation

Finally, click the Run button, and you will obtain a job ID jobid starting with "BEN" (**Figure S6**). It may take approximately 15 minutes to obtain the results, but you can close the page and input the job ID in the job center for later result inquiry.

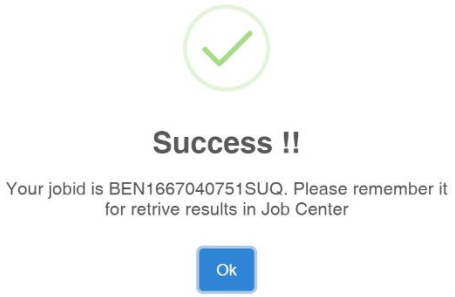


Figure S6 Successful submission window in SSP Benchmark module

The results are displayed in scatter plot, as shown in Figure 7, You can hover over each point to view the specific evaluation score and the corresponding topN parameter.

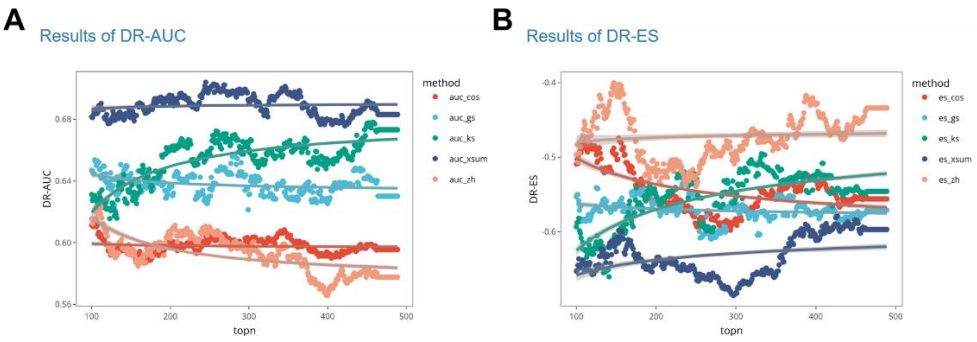


Figure S7 The result of AUC and ES result in Benchmark module

As seen in the demo results, the Xcos method has a higher score (auc_cos) in AUC and a lower score (es_cos) in ES. Both of these scores perform well at around topN 300. The CMap method follows closely behind.

3. Operation in Robustness module

In this module, you can evaluate the performance of signature search (SS) methods. In the Benchmark module, we tested SS methods based on drug annotations. However, it may not be appropriate when there is insufficient annotation for the profiles. Taking into consideration that the size of signatures plays a crucial role in the performance of all signature search methods, we have proposed a rigorous and robust analysis approach to determine the optimal number of genes in disease signatures[2].

The Robustness module, as shown in **Figure S8**, requires the following steps:

- Select Drug profiles (LINCS L1000)
- Select SS methods to test (at least two)
- Click run button and the result are presented in the right

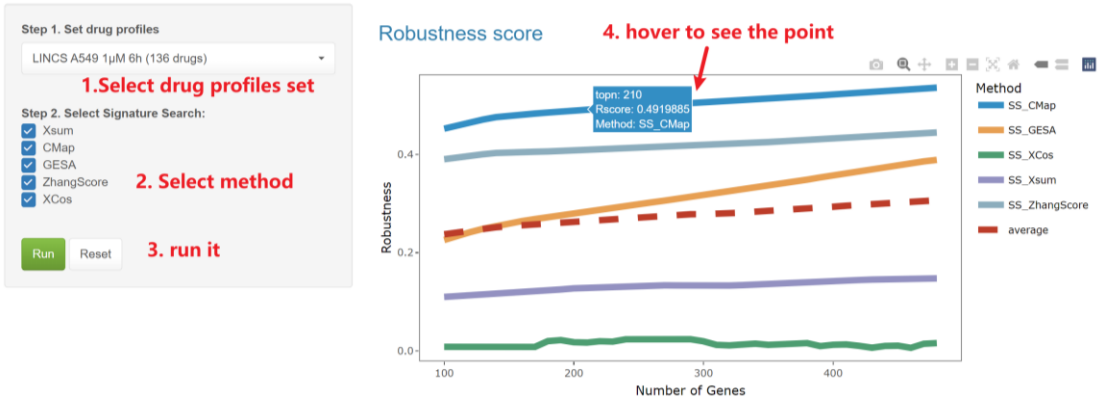


Figure S8 The result of Rscore in Robustness module

As seen in the demo results, the CMap method has a higher Rscore at around topN 480, and CMap, GSEA, and ZhangScore are above the average line, which means they are better and can be combined in Application module. Please note that as there use a drug-drug metrics, rather than disease-drug metrics, so it may not be identical to Benchmark module which we highly recommend for drug repurposing.

4. Operation in Application module

In this module, you can apply signature search (SS) methods to query drugs based on the disease signature input. Here we provide three ways to find promising drugs:

- **Single method**: Query drugs by one of signature search methods, as the traditional way. Typically, abs(logFC) > ±1 is used for filter differential expression genes.
- **SS_cross**: Query drugs by two signatures, and rank them by overall scores. SS_cross aims to found poly-pharmacological drugs.
- **SS_all**: Query drugs in multi SS methods and rank them with same direction (up or down) by robust rank aggregation (RRA),SS_all take all SS into account and found the "greatest common drugs".

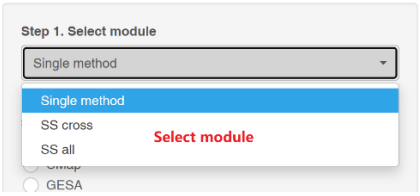


Figure S9 Method selection in Application module

Different way requires different steps:

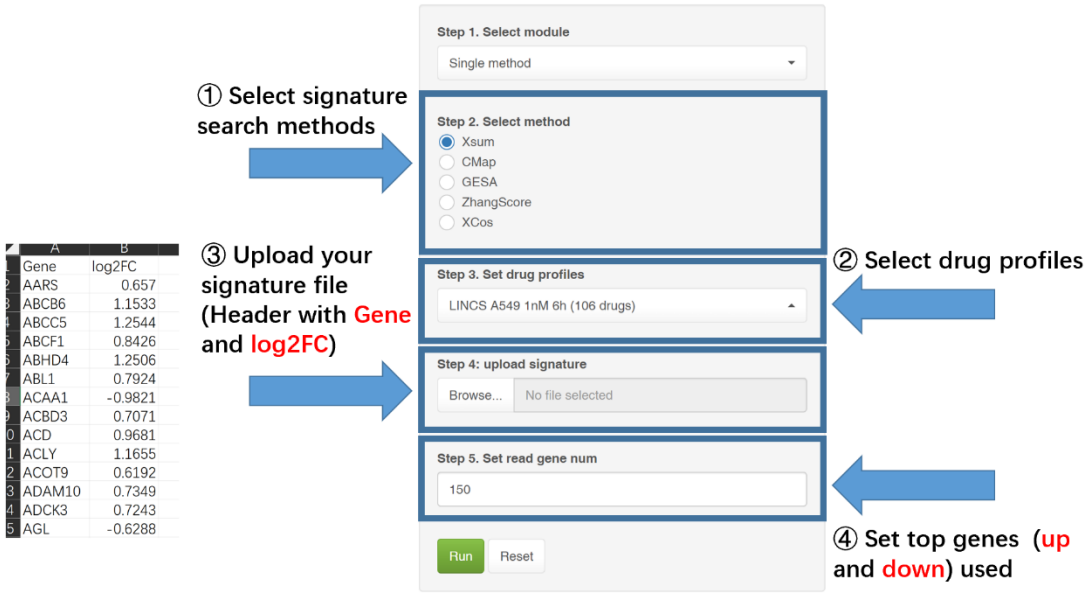
For **Single method**, we need four steps:

① Select a desired signature search method,

② Select one drug profiles,

③ Upload your signature file (Header with Gene and log2FC), and

④ Set how many top genes (up and down) used, it may be hinted from the benchmark module or robustness module.

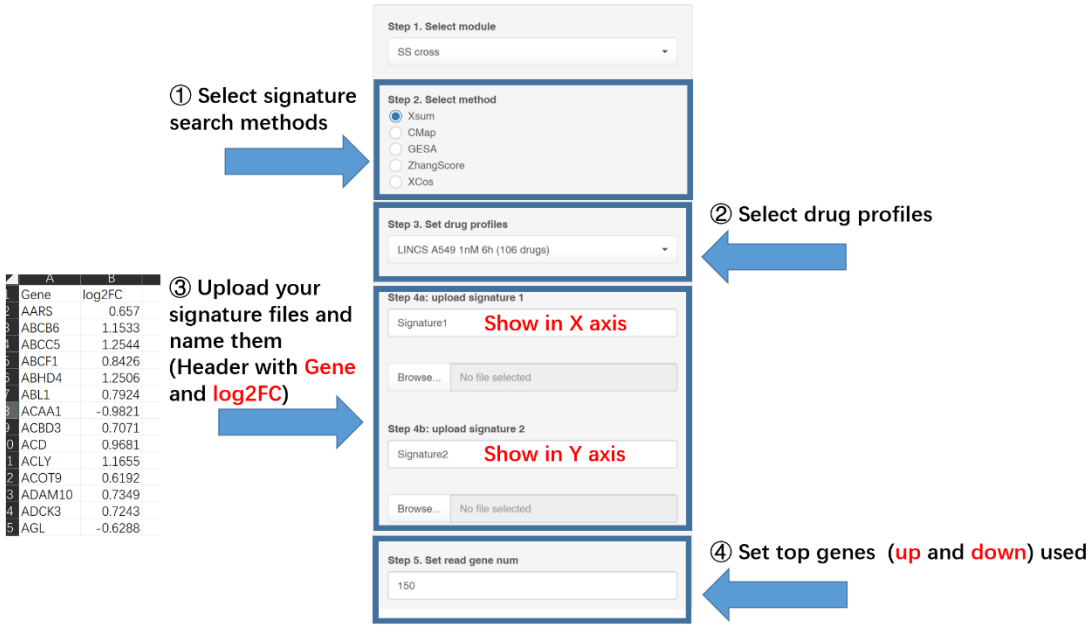


	A	B
1	Gene	log2FC
2	AARS	0.657
3	ABCB6	1.1533
4	ABCC5	1.2544
5	ABCF1	0.8426
6	ABHD4	1.2506
7	ABL1	0.7924
8	ACAA1	-0.9821
9	ACBD3	0.7071
10	ACD	0.9681
11	ACLY	1.1655
12	ACOT9	0.6192
13	ADAM10	0.7349
14	ADCK3	0.7243
15	AGL	-0.6288

Figure S10 Usage of single method in Application module

For **SS cross** , steps ③ is different:

two signatures' files and their names are required, name of the first signature represents X axis and the second Y axis in result figure.



	A	B
1	Gene	log2FC
2	AARS	0.657
3	ABCB6	1.1533
4	ABCC5	1.2544
5	ABCF1	0.8426
6	ABHD4	1.2506
7	ABL1	0.7924
8	ACAA1	-0.9821
9	ACBD3	0.7071
10	ACD	0.9681
11	ACLY	1.1655
12	ACOT9	0.6192
13	ADAM10	0.7349
14	ADCK3	0.7243
15	AGL	-0.6288

Figure S11 Usage of SS_cross in Application module

For **SS all** , steps ① is different:

We can select some methods and direction to rank the drugs, generally, if we upload a disease signature, chose " **down** ", else, chose " **up** ".

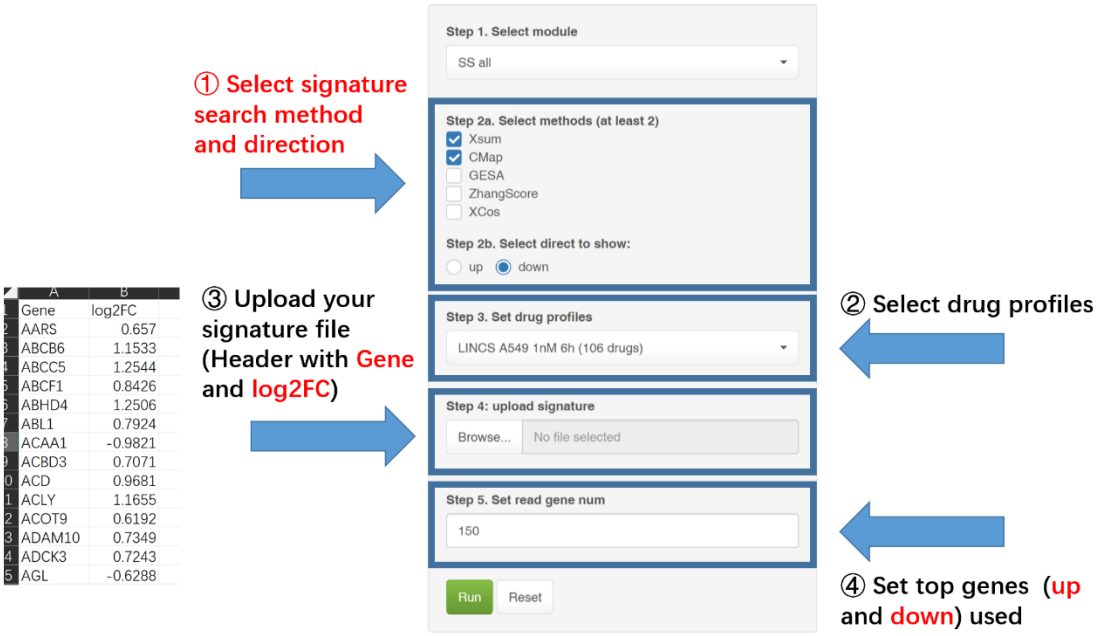


Figure S12 Usage of SS_all in Application module

Finally, click the Run button and you will get a **jobid**. It may take 15 mins to get results, but don't worry, you can close the page and input **jobid** in job center for result inquiry later.

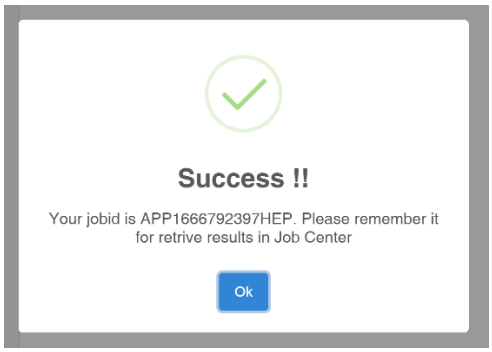


Figure S13 Successful submission window in SSP Application module

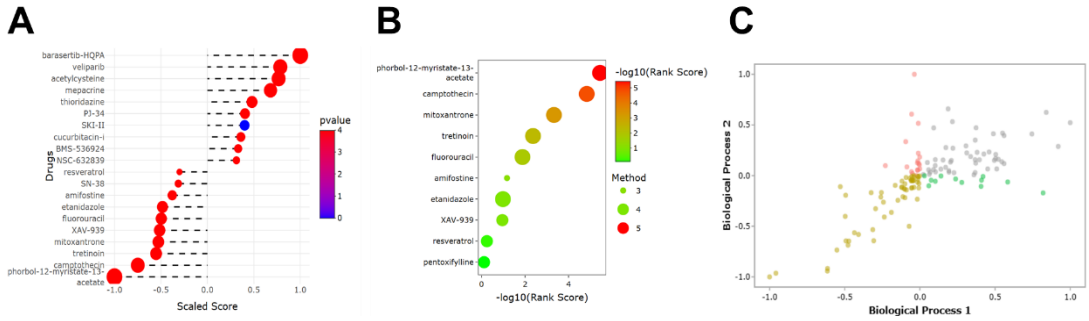


Figure S14 Results in Application module

Figure S14A illustrates the results of the single method using a lollipop chart, showing the top 5 drugs with the highest positive scores and the top 5 drugs with highest negative scores. Positive values indicate drugs with an activating effect on the disease, while negative values indicate drugs with a therapeutic effect. The color represents the p-value, and the size of the bubbles is proportional to the absolute value of the scores.

Figure S14B demonstrates the results of SS_all, displaying the top 10 drugs with significant Rank scores in the same direction (Up or Down). The top drug is more likely to be promising as it is also the highest-ranked drug in most SS methods. The color represents the number of methods enriched, and the size of the bubbles is correlated with the negative logarithm of the scores.

Figure S14C showcases the results of SS_cross, where all drugs are plotted in a scatter plot based on their scores on the x-y plane. The scores are normalized from -1 to 1. Different quarters represent different effects of the drug on the disease. For example, quarter 4 indicates that the drug is therapeutic for both diseases.

5. Operation in other modules

Annotation

SSP integrates annotation data for 286 drugs in 30 cancers extracted from the GDSC database[8]. Users can directly download the corresponding annotation files for Benchmark module.

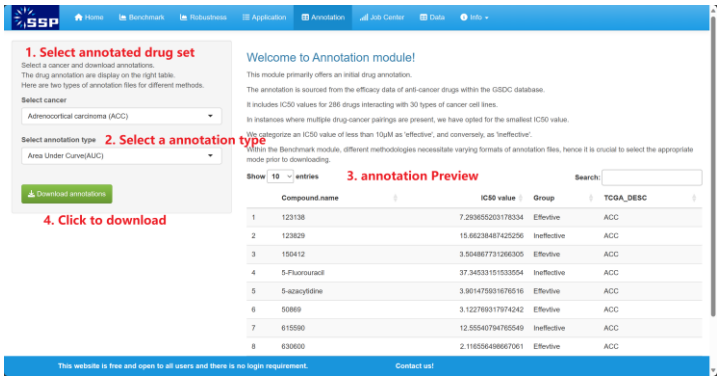


Figure S15 Webpage of Annotation module.

Job center

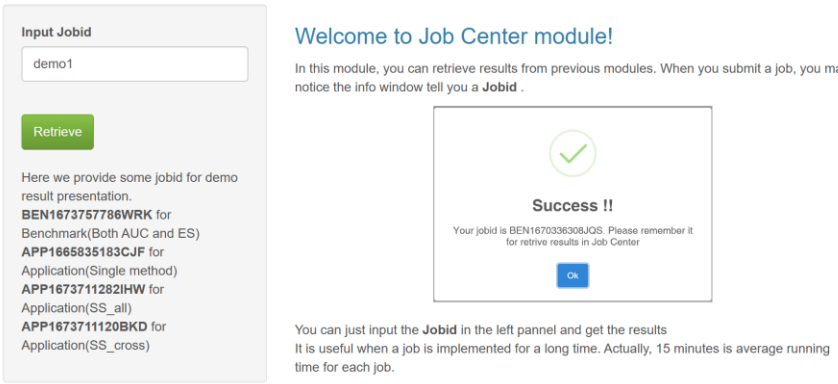


Figure S16 Webpage of Job center

Considering the large computational workload of the SSP website, we provide a jobid for each task execution, which allows users to retrieve the computational results of their submitted jobs. Users can enter the job id in the job center to view the results of their previous job submissions.

Data page

Figure S17 Webpage of Data

Data page provide the demo file , scripts and curated drug profile based on concentration and cell line. User can download by click the corresponding button.

6. Explanation of methods used in module

Signature Search methods

XSum[3], CMap[4], GESA[5], ZhangScore[6], and XCos[7] are state-of-the-art SS method, derived from articles with a high number of citations. Users can read these articles to understand the principles behind the methods. These scripts are home-made and has been referenced from the signatureSearch[8] and RCSM[9] packages.

Robustness

For Drugs in LINCS L1000, we assigned labels to drugs in the same group from 1 to n. For each drug, we extracted the top x up-regulated and top x down-regulated DEGs from its profile, creating a signature. This signature was then queried using one of the five signature search (SS) methods to obtain matching scores for all drugs. Subsequently, we ranked the drugs based on these scores. To evaluate the robustness of these methods at different x values, we utilized three parameters, which are:

(1) Correlation (R) of the input and top1 output for all drugs.

(2) Mean of the difference scores between top1 and top2 in outputs.

(3) Standard deviation (SD) of the difference between scores of top1 and top2 in output.

Finally, the robustness score (Rscore) can be expressed by the following formula:

$$Rscore = \frac{Mean \times R}{SD}$$

A method can be considered to have achieved satisfactory performance if it can accurately identify the input active drug (stronger correlation), effectively differentiate between drugs (higher difference score), and demonstrate good stability (lower SD). In this study, we tested performance scores for the cases of x at 100,110,120.....480, respectively.

SS_cross and SS_all

Two methods, the SS_all module and SS_cross was designed to integrate the signature search methods to explore drugs. For SS_all method (Figure 1C), we combined the ranks of TCM active drugs in the same direction (>0 for drug signature or <0 for disease signature) by robust rank aggregation[10]. We assigned an overall score (0~1) to each active drug, with a lower overall score indicating greater significance in all methods.

In addition, it is applicable to screen TCM active drugs with polypharmacological effects based on multiple signatures[11]. For the SS_cross method, we obtained the scores of TCM active drugs for two different pharmacological signatures ($Score_{sig1}$ and $Score_{sig2}$) by the special signature search method, and TCM active drugs were divided into four quadrants based on the scores (Q1: both scores >0, Q2: $Score_{sig1}$ >0 but $Score_{sig2}$ <0, Q3: both scores <0, Q4: $Score_{sig1}$ <0 but $Score_{sig2}$ >0). Then, we calculated a unified score by the square root of absolute values:

$$\mathrm{Score_{sum}} = \sqrt{\mathrm{abs}(\mathrm{Score_{sig1}} \times \mathrm{Score_{sig2}})}$$

The screening strategy was performed by home-made scripts in R 4.0.4.

7. Reference

1. Yang C, Zhang H, Chen M, et al. A survey of optimal strategy for signature-based drug repositioning and an application to liver cancer. eLife 2022; 11:e71880

2. Tian S, Zhang J, Yuan S, et al. Exploring pharmacological active ingredients of traditional Chinese medicine by pharmacotranscriptomic map in ITCM. Brief Bioinform 2023; 24:bbad027

3. Cheng J, Yang L, Kumar V, et al. Systematic evaluation of connectivity map for disease indications. Genome Med 2014; 6:540

4. Subramanian A, Narayan R, Corsello SM, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell 2017; 171:1437-1452.e17

5. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences 2005; 102:15545–15550

6. Zhang S-D, Gant TW. A simple and robust method for connecting small-molecule drugs using gene-expression signatures. BMC Bioinformatics 2008; 9:258

7. Cheng J, Xie Q, Kumar V, et al. Evaluation of analytical methods for connectivity map data. Pac Symp Biocomput 2013; 5–16

8. Duan Y, Evans DS, Miller RA, et al. signatureSearch: environment for gene expression signature searching and functional interpretation. Nucleic Acids Res 2020; 48:e124

9. Lin K, Li L, Dai Y, et al. A comprehensive evaluation of connectivity methods for L1000 data. Briefings in Bioinformatics 2020; 21:2194–2205

10. Kolde R, Laur S, Adler P, et al. Robust rank aggregation for gene list integration and meta-analysis. Bioinformatics 2012; 28:573–580

11. Chen S, Liu X, Peng C, et al. The phytochemical hyperforin triggers thermogenesis in adipose tissue via a Dlat-AMPK signaling axis to curb obesity. Cell Metab 2021; 33:565-580.e7