Manual of SSP

Content

1. How to start?

A live version of SSP is hosted at **http://web.biotcm.net/SSP/** and mirror site **http://www.biotcm.net/SSP/**.

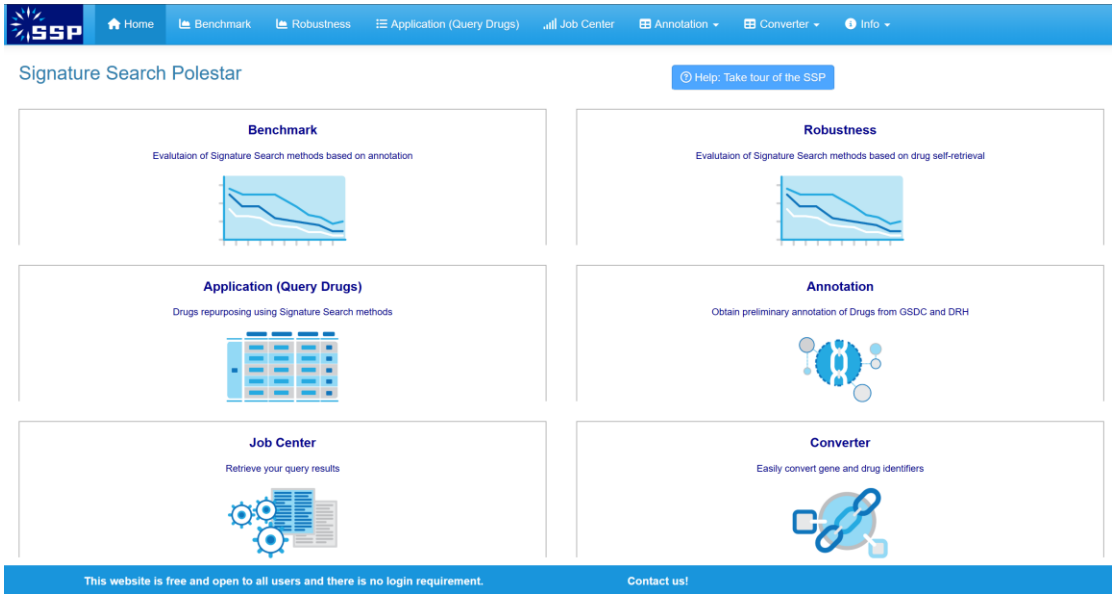


Figure S1 Homepage of SSP

If you want to deploy SSP on your own server, please visit SSP website (**http://web.biotcm.net/SSP/** or mirror site **http://www.biotcm.net/SSP/**), then visit the info-help page to get a full installation of SSP (~3G) (**Figure S2A**). In addition, we also provide source code on https://gitee.com/auptz/benchmark-ss (Chinese) or https://github.com/AuPtZ/BenchmarkSS (English) and download all files and run the "app.R" in RStudio (**Figure S2B**). Notably, essential packages are need to be installed before you run.

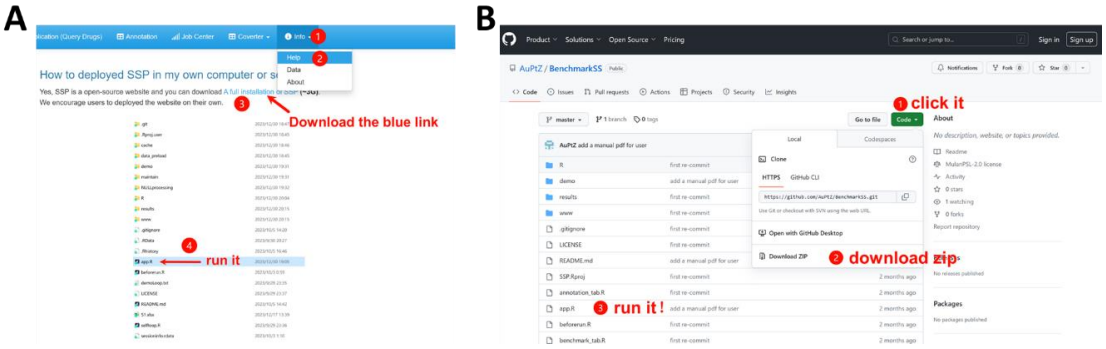


Figure S2 Two ways to deploy SSP. (A) a full installation on SSP website. (B) a mirror repository (only source code) on GitHub.

Important notice for new visit: As SSP is hosted on the Shiny Server, there is a latency period while the server initializes a session for new users. During this initialization phase, server-side packages are gradually loaded and the majority of reactive widgets may become unresponsive, a common occurrence that can impact the user experience. To mitigate this issue, a pop-up window has been implemented to indicate when the server initialization is complete. It is recommended that

users await until the window closes. Certainly, if a user is currently accessing the SSP, and a new user attempts to access it, the SSP will directly invoke the previous initialization without displaying a pop-up window.

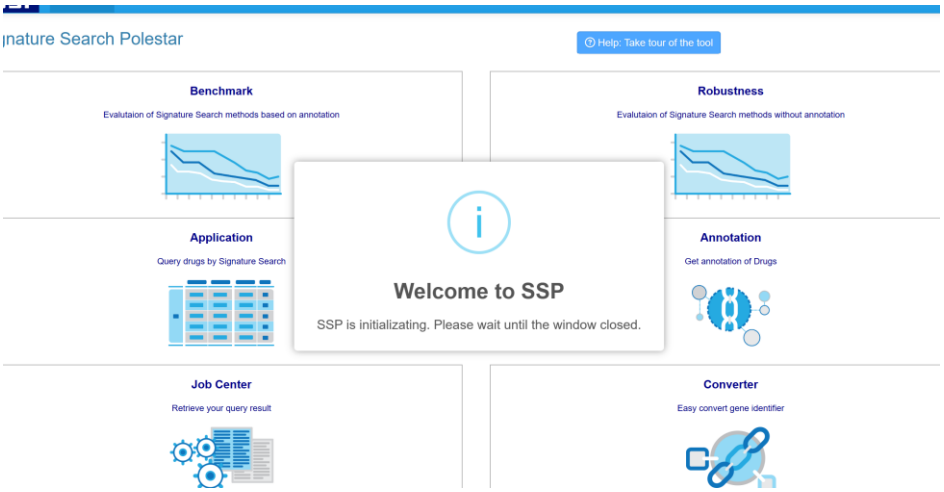


Figure 3 Pop-up window of initialization when new user visits.

2. Prepare input data

SSP require Two types of data and selection of dataset:

- An oncogenic signature (OGS), header with Gene and log2FC (**Figure S4A**). It typically consists of differentially expressed genes (DEGs) derived from sequencing samples of cell or animal experiments, or patient cohorts, such as GEO, TCGA and ICGC.

  The OGS should comprise a minimum of 20 genes, with at least 10 genes exhibiting a positive log2 fold change (log2FC > 0) and 10 genes showing a negative log2 fold change (log2FC < 0). Notably, SSP accepts the genes in the format of gene symbol and assumes that input OGS are statistically significant (adjust $p < 0.05$), ensuring the significance of further analysis. Should your OGS contain genes formatted with alternative identifiers (such as EntrezID, Ensembl, UniProt, Gene name, etc.), proceed to the Converter module (for Gene) for the necessary conversion.

- Drug annotations for AUC and ES (**Figure S4B** and **Figure S4C**) and users must use at least one method to assess the performance of SSM in Benchmark.

  The Drug annotations for AUC should comprise a minimum of 50 drugs, and for ES should comprise a minimum of 10 drugs.

  Drug annotations are commonly sourced from databases and resources such as ChEMBL, PubChem, scientific literature, clinical trials, and DrugBank. Users have two options: ① Download a blank annotation table and label it manually (**Figure S4D**), or ② Independently compile annotations from various sources and upload them into the Converter module to get a format-compatible annotation file (**Figure S4E**). Converter module (for Drug) could convert drugs with other identifiers or try to correct the drug name (capitalization or the presence of spaces, hyphen, as illustrated in demo1 button) into the acceptable format (drug name in LINCS L1000). In addition, Converter module both accept the input in one-column tab-separated file (for ES) or two-column tab-separated file (for AUC) and will keep the second column in output.

It is impractical to annotate all drugs; however, the more annotations obtained, the more accurate the results will be.

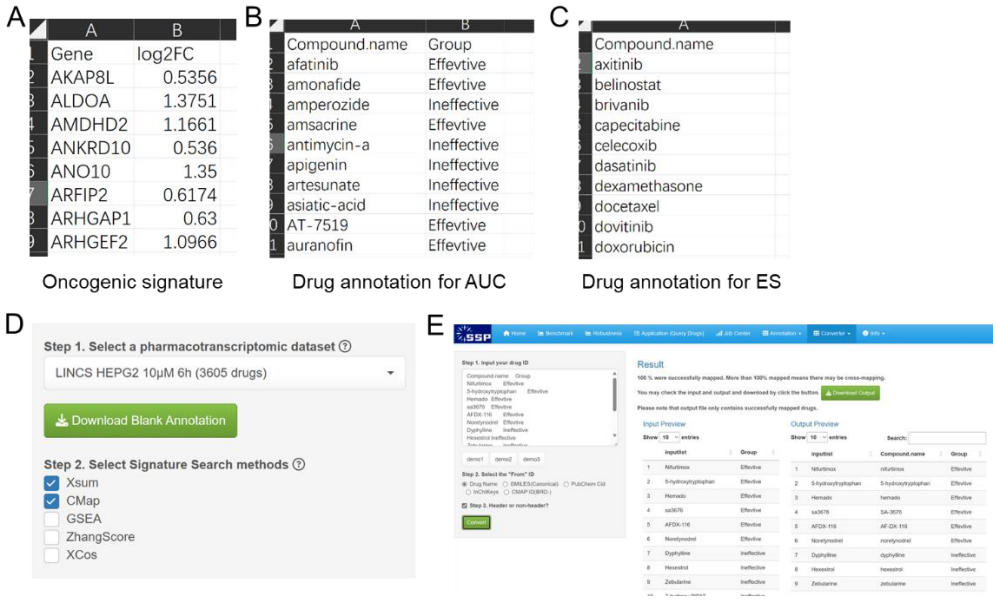


Figure S4 Demo input in SSP

- In addition, SSP also require user to select a built-in pharmacotranscriptomic dataset (PTD) to perform a job. PTDs are originally sourced from LINCS L1000 (GSE92742), is presented in an n × m matrix where n represents the drug names and m denotes a list of gene symbols along with their corresponding log2FC values.

	A	B	C	D	E
1	erastin	promethazine	IKK3-inhibitor-IX	orantinib	fluspirilene
2	0.856790721	0.305442989	-0.192084193	0.445083737	-2.408152103
3	0.518512964	0.16569677	0.476561785	1.013044238	-1.655670762
4	-0.621897042	-0.802020311	1.514186263	0.472688138	-0.367278069
5	0.458046645	-0.336790383	-1.435213923	0.201365829	-0.325466812
6	-0.225423694	-0.693069398	1.570700049	0.428453147	-0.548683226
7	0.251141697	-0.58618319	-1.025687814	-0.475418508	1.19861722
8	-0.104918249	0.494013131	1.033460736	0.747183084	0.791781902
9	0.58414495	-0.077427194	-0.82226193	-0.177413836	-0.263801754
10	-0.575057149	0.003155112	-0.059363365	0.101829395	0.171995655
11	-0.523953676	-0.097454175	0.016299017	-0.561868548	1.767373562
12	-0.327870011	-0.066550262	-0.554158986	0.201053411	1.295616388
13	0.291723728	-0.421341121	-0.957800329	0.031723335	-0.361877382
14	0.06516014	-0.062972084	-0.395871103	-0.870026231	1.442822456
15	0.223884076	0	0.156281605	0.206369132	-0.787587404
16	-0.448647708	-0.68056798	0.180772841	0.657884181	0.921201766
17	-0.033641316	0.378573984	-0.036469594	0.271739364	-0.785263956

Figure S5 The matrix of pharmacotranscriptomic dataset.

3. Operation in Benchmark module

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

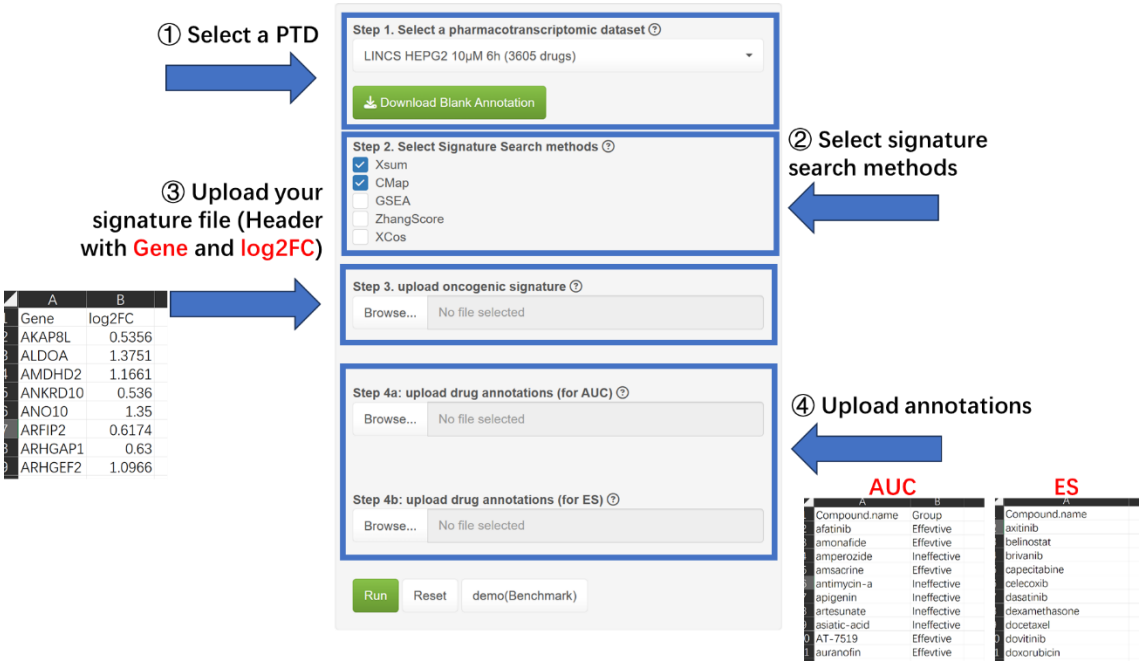


Figure S6 Workflow in Benchmark module

The Benchmark module, as shown in **Figure S6**, requires the following steps:

- Select pharmacotranscriptomic dataset (LINCS L1000)
- Select SSMs to test (at least two)
- A signature (header with gene and log2FC) 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 GSEA-like enrichment score (ES), the lower ES indicates better performance. (Yang *et al.*, 2022)

Finally, click the Run button, and you will obtain a job ID jobid starting with "BEN" (**Figure S7**). 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.

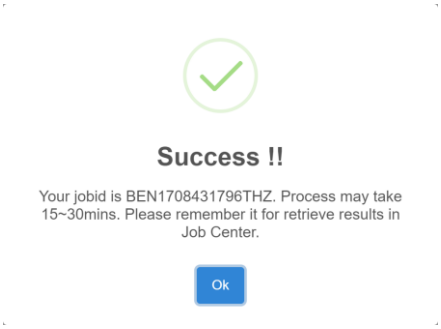
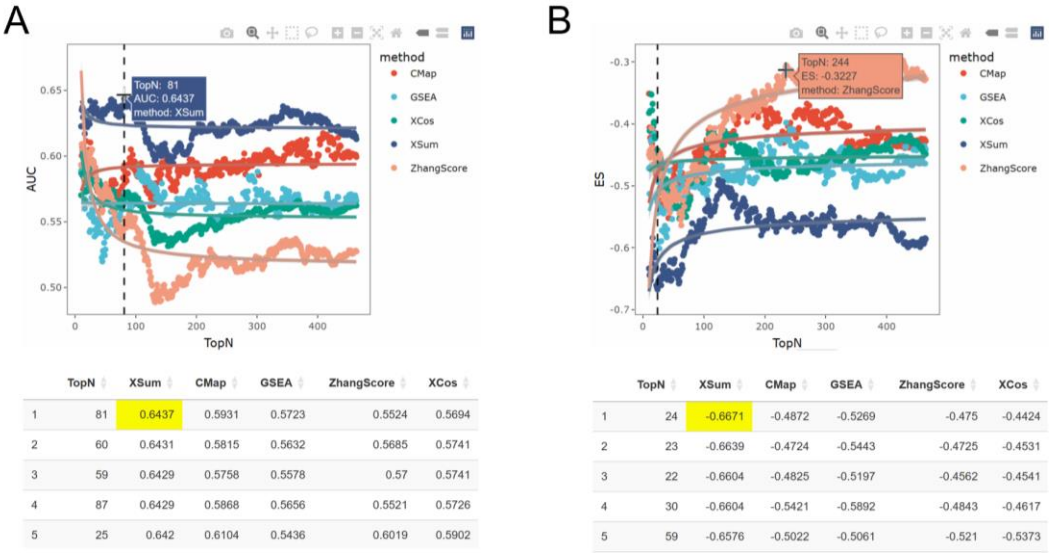


Figure S7 Pop-up window after successful submission in SSP Benchmark module

The results are displayed in scatter plot, as shown in **Figure S8**, you can hover over each point to view the specific evaluation score and the corresponding topN parameter. Each plot is following with a table and the best SSM and topN are presented in yellow. As seen in the demo results, the XSum method has a higher score in AUC and a lower score in ES. Scores for AUC performs well at around topN 80 and scores for ES performs well at around topN 25.



	TopN	XSum	CMap	GSEA	ZhangScore	XCos
1	81	0.6437	0.5931	0.5723	0.5524	0.5694
2	60	0.6431	0.5815	0.5632	0.5685	0.5741
3	59	0.6429	0.5758	0.5578	0.57	0.5741
4	87	0.6429	0.5868	0.5656	0.5521	0.5726
5	25	0.642	0.6104	0.5436	0.6019	0.5902

	TopN	XSum	CMap	GSEA	ZhangScore	XCos
1	24	-0.6671	-0.4872	-0.5269	-0.475	-0.4424
2	23	-0.6639	-0.4724	-0.5443	-0.4725	-0.4531
3	22	-0.6604	-0.4825	-0.5197	-0.4562	-0.4541
4	30	-0.6604	-0.5421	-0.5892	-0.4843	-0.4617
5	59	-0.6576	-0.5022	-0.5061	-0.521	-0.5373

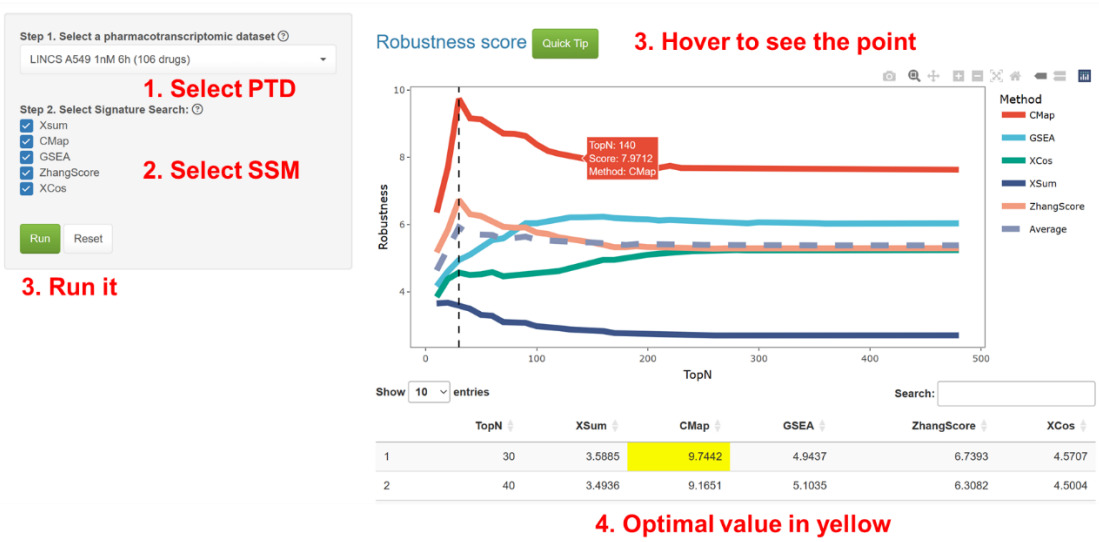
Figure S8 The result of AUC and ES result in Benchmark module

3. Operation in Robustness module

In this module, you can evaluate the performance of signature search method (SSMs). In the Benchmark module, we tested SSMs based on drug annotations. However, it may not be appropriate when there is insufficient annotation for drugs in PTD, such as subtype cancer or rare cancer. Taking into consideration that the size of signatures plays a crucial role in the performance of all SSMs, we have proposed a rigorous and robust analysis approach to determine the optimal number of genes in oncogenic signatures(Tian *et al.*, 2023).

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

- Select a PTD (LINCS L1000)
- Select pharmacotranscriptomic dataset (LINCS L1000)
- Click “run” button and the result are presented in the right panel



	TopN	XSum	CMap	GSEA	ZhangScore	XCos
1	30	3.5885	9.7442	4.9437	6.7393	4.5707
2	40	3.4936	9.1651	5.1035	6.3082	4.5004

Figure S9 Demo operation in Robustness module

As seen in the demo results, the CMap method has a higher Rscore at around topN 30, 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 Robustness module may differ from that in the Benchmark module. Researchers are strongly encouraged to utilize the Benchmark module, in alignment with their specific fields of study. Should the number of topN genes from the Robustness module exceed the length of the OGS, it is recommended to assess whether the scores obtained from the Robustness module at the corresponding length are close to the optimal values. If not, consideration should be given to replacing the OGS.

4. Operation in Application module

In this module, you can apply signature search method (SSMs) and topN to query drugs based on the oncogenic signature input.

An oncogenic signature (OGS), header with Gene and log2FC. It typically consists of differentially expressed genes (DEGs) derived from sequencing samples of cell or animal experiments, or patient cohorts, such as GEO, TCGA and ICGC.

The OGS should comprise a minimum of 20 genes, with at least 10 genes exhibiting a positive log2 fold change (log2FC > 0) and 10 genes showing a negative log2 fold change (log2FC < 0). Notably, SSP accepts the genes in the format of gene symbol and assumes that input OGS are statistically significant (adjust $p < 0.05$), ensuring the significance of further analysis. Should your OGS contain genes formatted with alternative identifiers (such as EntrezID, Ensembl, UniProt, Gene name, etc.), proceed to the Converter module (for Gene) for the necessary conversion.

Optimal SSM and topN are determined in Benchmark and Application, of note, if you use two OGSs in SS_cross, please make sure both OGSs share the near or same topN and optimal SSM.

Application module provide three approaches to drug repurposing (**Figure S9**).

- **Single method**: Query drugs by one of SSMs, as the traditional drug repurposing way. Typically, abs(logFC) > ±1 is used for filter differential expression genes.
- **SS_cross**: Query drugs by two OGSs, and rank them by overall scores. SS_cross aims to found drugs sharing consensus between multiple signatures.

- **SS_all**: Query drugs in multiple SSMs (with superiority in Benchmark and Application module) and rank them with same direction (up or down) by robust rank aggregation (RRA), SS_all take all selected SSM into account and found the "greatest drugs".

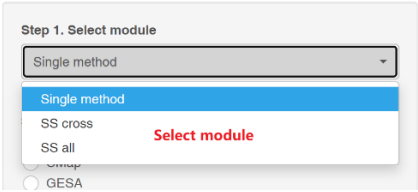


Figure S10 Method selection in Application module

Different way requires different steps:

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

① Select a desired SSMs,

② Select one pharmacotranscriptomic dataset,

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

④ Set how many topN genes (up and down) used, it may be hinted from the Benchmark module or Robustness module.

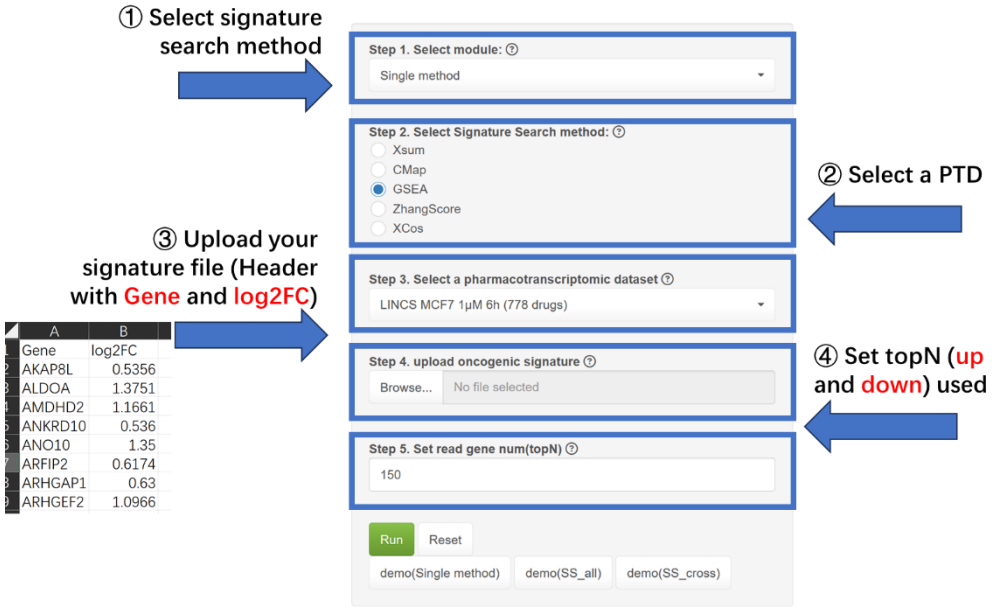


Figure S11 Usage of single method in Application module

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

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

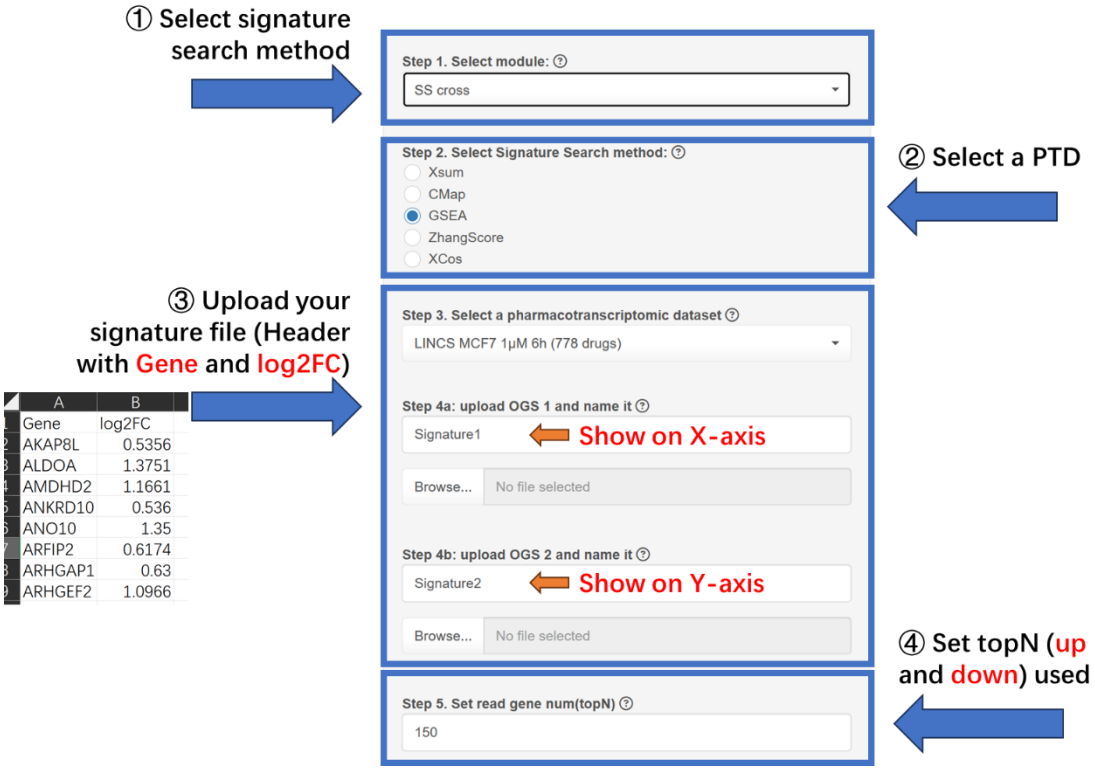


Figure S12 Usage of SS_cross in Application module

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

We can select some methods and direction to rank the drugs, generally, as SSP use oncogenic signature, please choose "**down**".

Figure S13 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.

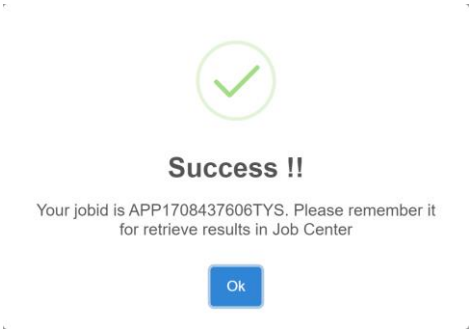


Figure S14 Successful submission window in SSP Application module

The results of Application module are presented in **Figure S14**.

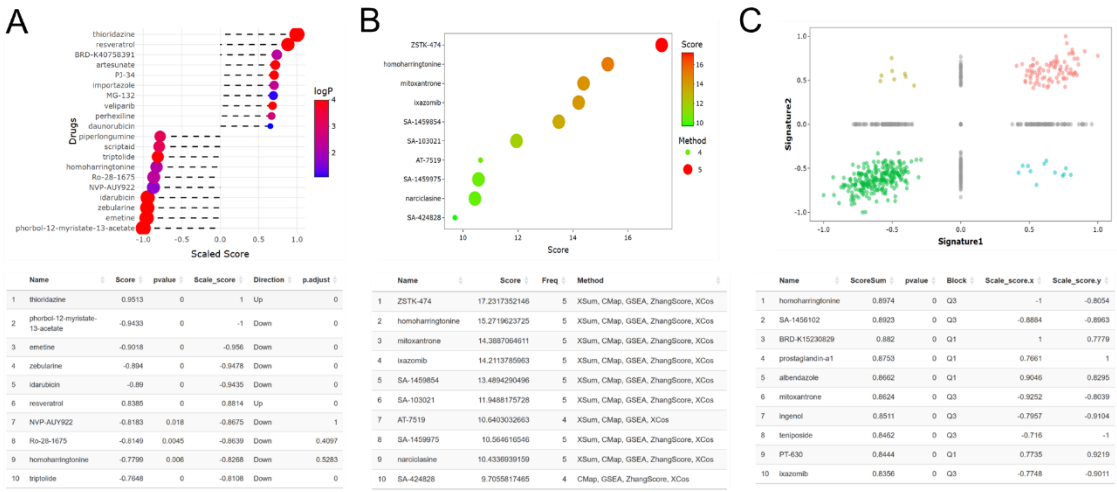


Figure S15 Results in Application module

Figure S15A 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 a potentially activating effect on the disease, while negative values indicate drugs with a potentially therapeutic effect. The color represents the p-value, and the size of the bubbles is proportional to the absolute value of the scores.

Figure S15B demonstrates the results of SS_all, displaying the top 10 drugs with significant scores computed by RRA 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 SSMs. The color represents the number of methods enriched, and the size of the bubbles is correlated with the negative logarithm of the scores.

Figure S15C 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 3 (the lower left corner) indicates that the drug is therapeutic for both diseases.

In addition, a “Quick Tip” button is provided above the plot and user can click to see a detailed explanation of plot and column of table.

5. Operation in other modules

5.1 Annotation

SSP provides the preliminary drug annotation to facilitate user's manual drug annotation.

For AUC, this module integrates annotation data for 286 drugs in 30 cancers sourced from the GDSC database (Yang *et al.*, 2013). A threshold of 10μM is applied, classifying drugs below this value as effective and those above as ineffective. To address IC50 redundancy, the median IC50 value for duplicate drugs is utilized to represent their activity. Every cancer type is covered with a minimum of 271 drugs (Supplementary Table S1).

For ES, the module includes indication-based annotation data for 163 FDA-approved drugs across 15 cancer types, curated manually from the DRH database, with a minimum of five drugs annotated per cancer type (Supplementary Table S2). Users can directly download annotation files for the Benchmark module corresponding to ES and AUC.

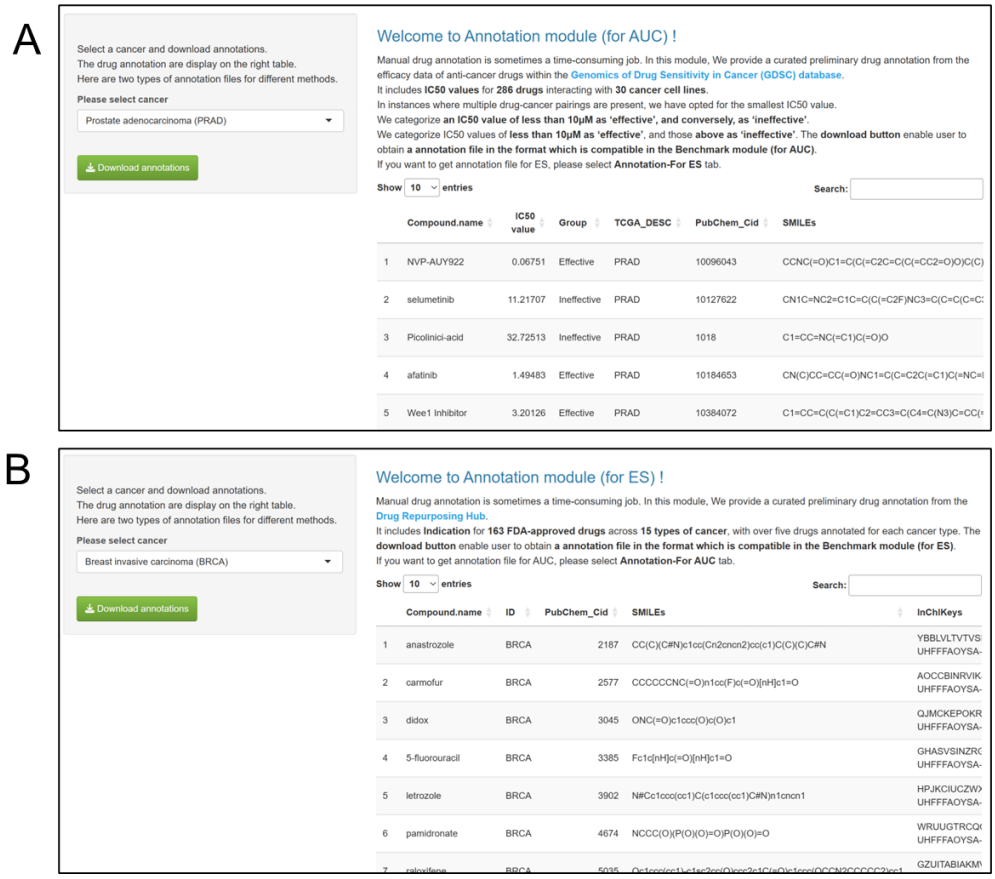


Figure S16 Webpage of Annotation module.

5.2 Job center

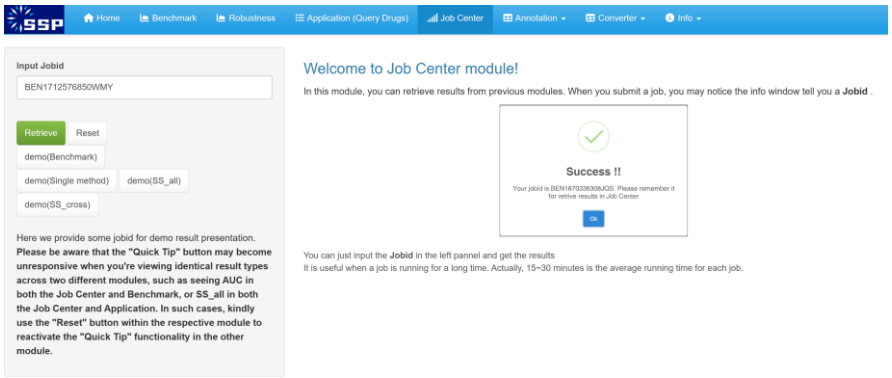


Figure S17 Webpage of Job center

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.

5.3 Data page

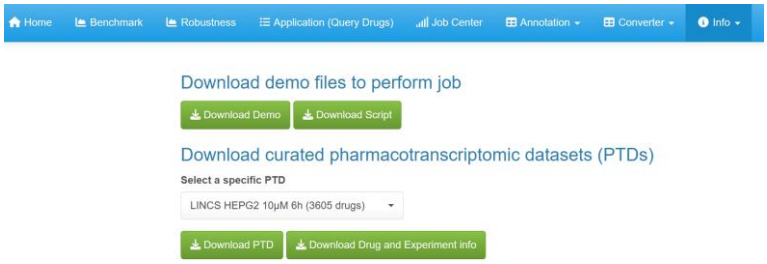


Figure S18 Webpage of Data

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

6. Explanation of methods used in module

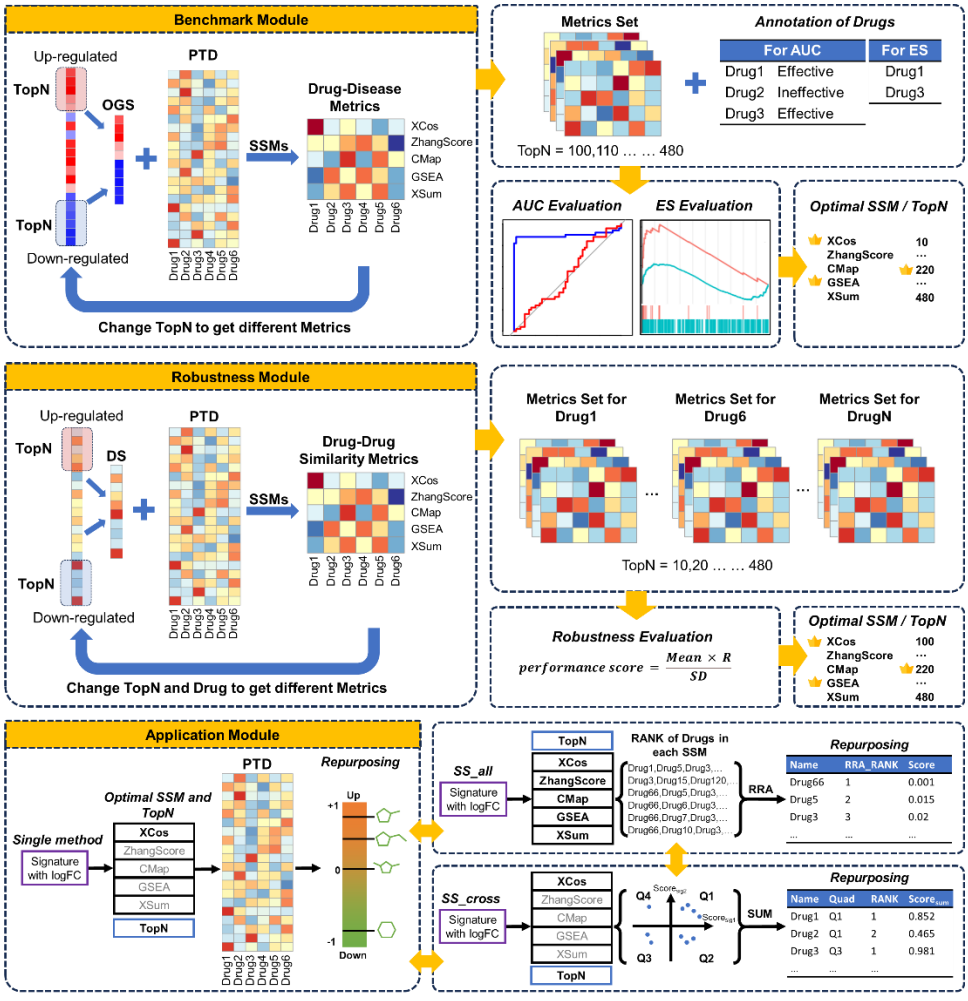


Figure S19 A illustration of methodologies in SSP. DS, drug signature.

A comprehensive illustration of methodologies in SSP is presented for better readability.

6.1 Signature Search methods

XSum (Cheng *et al.*, 2014), CMap (Subramanian *et al.*, 2017), GSEA (Subramanian *et al.*, 2005), ZhangScore (Zhang and Gant, 2008), and XCos (Cheng *et al.*, 2013) are state-of-the-art SSMs, derived from articles with a high number of citations. Users can read these articles to understand the principles behind the methods. These in-house scripts refers to the signatureSearch (Duan *et al.*, 2020) and RCSM (Lin *et al.*, 2020) packages.

The rationale behind our choice of these SSMs is as follows:

1. These methods encompass both highly cited, well-established approaches (e.g., CMap and GSEA) (Subramanian *et al.*, 2017, 2005), and those identified as superior in benchmark analyses (e.g., XCos, XSum, and ZhangScore) (Cheng *et al.*, 2014; Yang *et al.*, 2022; Lin *et al.*, 2020).
2. These methods support identical input and output format, ensuring compatibility and ease of integration into SSP workflow.
3. Due to the high computational demands in SSP, these SSMs are amenable to refactoring for parallel processing, critical for the effective management of extensive datasets.

6.2 Benchmark

As, shown in **Figure S19**, OGS in the benchmark undergoes iterative tailored filtration for assessing the performance of SSMs. Specifically, the genes in OGS are ordered based on logFC, then an equivalent number (topN) of significantly up-regulated and down-regulated genes are selected to form a new OGS. This approach guarantees a comprehensive and balanced representation of the varied gene expression alterations associated with the disease state. Then, five state-of-the-art SSMs, XSum, CMap, GSEA, ZhangScore, and XCos are called to calculate scores. We compute the scores for these SSMs across a range of topN values (10~480) to obtain drug-cancer similarity metrics. In general, a score greater than 0 indicates that the drug is potentially agonistic to the cancer, while a score less than 0 indicates that the drug is potentially therapeutic to the cancer.

Next, based on the drug annotation data, two benchmarking indices, namely area under the curve (AUC) and enrichment score (ES), are generated for evaluating the performance of drug-disease similarity metrics. The methodologies of AUC and ES are derived from commonly used metrics for evaluating drug efficacy (Yang et al., 2022). Although there are some different benchmarking standards (Luo et al., 2022; Fang et al., 2021), they are significantly different in score computation and require more annotation preparation, which may hinder users from easily using.

Within the Benchmark module, these benchmarking standards can be adeptly employed to utilize the scores based on PTDs and annotation information, thereby facilitating an accurate assessment of the precision of SSMs. The AUC uses scores of drugs annotated with “Effective” and “Ineffective” and aims to assess whether the SSM can effectively discriminate between these categories effectively, with a higher score signifying a more effective method. This metric is applicable to drug annotations derived from large-scale experimental screenings, such as determining drug efficacy based on IC50 values. The ES is just like GSEA, aims to evaluate whether the drugs can be enriched at the top of the descending ordered list of all drugs based on scores from SSM, with a negative score indicating a better method. This metric is suitable for drug annotations based on clinical practice, where a minority of drugs are known to be effective, and the efficacy of the majority remains uncertain.

6.3 Robustness

As, shown in **Figure S19**, 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 gene expression profile, creating a signature. This signature was then queried using one of the five signature search method (SSMs) 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:

$$Performance\ score\ = \frac{Mean\ \times\ R}{SD}$$

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). Hence, a higher score means a better performance. In this study, we tested performance scores for the cases of x at 10,20,30.....480, respectively.

6.4 SS_cross and SS_all

As, shown in **Figure S19**, Two methods, The "SS_all" method and "SS_cross" method were designed to find promising drugs with consensus under multiple SSMs or OGSs (Tian et al., 2023). In the SS_all, user is required to select SSMs with superiority in Benchmark or Robustness. Then, drug-cancer metrics are generated by all optimal SSMs against an OGS. The ranks of drugs based on metrics in the same direction (< 0 or > 0) are combined using robust rank aggregation (Kolde et al., 2012). The robust rank aggregation returns a p-value of each drug and we assign log(P-value) as the overall score to each drug. Hence, a drug with top rank in more SSMs results in a higher overall score and indicates greater potential to be promising.

In SS_cross, user is required to prepare two distinct OGSs and then drug-cancer metrics are generated by these OGSs with one SSM. These drugs are then divided into four quadrants based on the sign of these metrics (Chen et al., 2021), representing potentially agonistic response (> 0) or therapeutic response (< 0) (Figure 1F). (Q1: both scores >0, Q2: $Score_{OGS1}$ <0 but $Score_{OGS2}$ >0, Q3: both scores <0, Q4: $Score_{OGS1}$ >0 but $Score_{OGS2}$ <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 drugs that exhibit therapeutic response in both OGSs with higher $Score_{sum}$ show promise for repurposing, particularly those located in lower left corner (Q3).

7. A case study of liver cancer using SSP

Liver cancer, recognized as one of the deadliest malignancies globally, is among the top three causes of cancer death in 46 countries (Rumgay *et al.*, 2022). Majority of liver cancer are Hepatocellular carcinoma (HCC). Despite the advent and approval of numerous drugs, such as Ramucirumab (Zhu

et al., 2019), Pembrolizumab (Finn, Ryoo, *et al.*, 2020), Atezolizumab-Bevacizumab (Finn, Qin, *et al.*, 2020), and Nivolumab-Cabozantinib (Yau *et al.*, 2023). These treatments have generally provided only marginal survival advantages. Consequently, there is an urgent and pressing need for more efficacious therapeutic options to combat HCC.

In this case study, we try to find promising drug for HCC. **Please note all the signature for case study are provided in Info-data page.**

7.1 obtain oncogenic signature

OGS is a gene list with log2FC, which usually generate from experiment or patient cohort. In this study, we obtain a HCC-related OGS from previous publication (Chen *et al.*, 2017).

Drug annotations were then uploaded in converter page to prioritize genes present in the LINCS L1000 dataset and resulted in 54 genes in down and 19 genes in up. (see **CS_OGS.txt**)

In general, we recommend an OGS with at least 20 genes.

7.2 obtain drug annotation

Drug annotation for AUC were obtained from Annotation module (drugs in *Liver hepatocellular carcinoma (LIHC)*), and ChEMBL (version 33) (Zdrazil *et al.*, 2024) and Liver Cancer Model Repository (LIMORE) data sets (Qiu *et al.*, 2019). As we keep the minimum IC_{50} values to reduce the redundancy in the Annotation module, precedence was given to the IC_{50} data treated on HepG2 cell line from ChEMBL and LIMORE during the aggregation process.

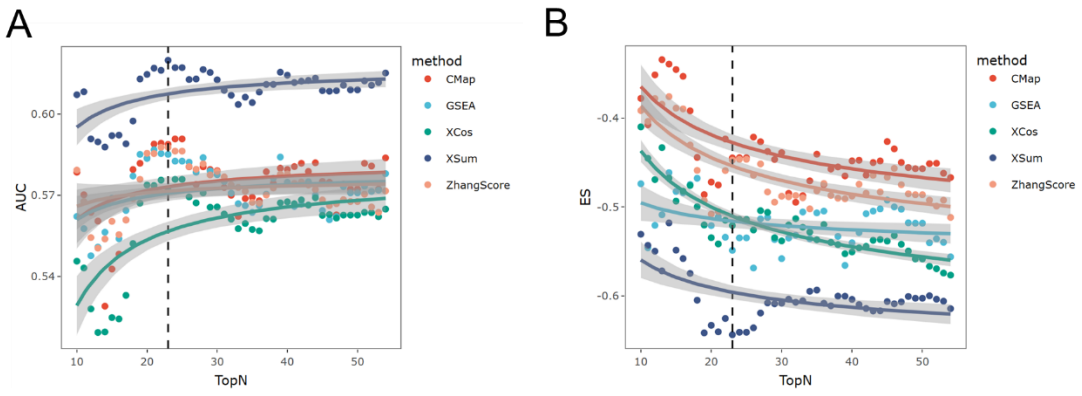
Drug annotation for ES were obtained from Clinicaltrail.gov by identifying HCC-specific clinical trials that involve drug interventions and assess anti-tumor activity.

Drug annotations were then uploaded in converter page to prioritize drugs present in the LINCS L1000 dataset and resulted in 185 drugs in AUC and 29 drugs in ES. (see **CS_ES.txt** and **CS_AUC.txt**). Additionally, within the ES annotation file, only 19 drugs (65.17%) were also present in the AUC annotation file, highlighting the independence of the two annotation sources.

7.3 perform benchmark, find optimal method and TopN

Considering the focus on HCC and the aim to include a broader range of drugs, we selected the pharmacotranscriptomic dataset specific to the HepG2 cell line (treatment with 10μM for 6 hours). The files **CS_OGS.txt**, **CS_ES.txt**, and **CS_AUC.txt** were uploaded to the corresponding fileInput on the benchmark module, selecting "LINCS HEPG2 10μM 6h (3605 drugs)". The task was submitted, and the results (**Figure 20**) indicated that at topN=23, the XSum method achieved optimal performance in both AUC and ES evaluation metrics (maximizing AUC and minimizing ES). Consequently, that the XSum method with topN set to 23 was selected as the parameter and approach for drug repositioning.

It is noteworthy that instances where same topN ranks first in both evaluation metrics are not usual. Typically, topN values that appear in the top 10 for both ES and AUC can be recommended for use in the Application module.



	TopN	XSum	CMap	GSEA	ZhangScore	XCos
1	23	0.6198	0.5888	0.5851	0.5877	0.5718
2	24	0.6171	0.5908	0.5826	0.5863	0.5759
3	25	0.6171	0.5908	0.5826	0.5863	0.5759
4	21	0.6169	0.5889	0.5868	0.5854	0.5736
5	28	0.6164	0.5805	0.5839	0.5816	0.5709

	TopN	XSum	CMap	GSEA	ZhangScore	XCos
1	23	-0.6435	-0.4448	-0.5487	-0.4473	-0.5208
2	19	-0.6414	-0.4861	-0.5203	-0.4926	-0.5202
3	24	-0.6407	-0.4446	-0.5347	-0.4464	-0.5139
4	25	-0.6407	-0.4446	-0.5347	-0.4464	-0.5139
5	21	-0.6402	-0.4753	-0.5111	-0.505	-0.5415

Figure 20 Result in Benchmark module, XSum and topN = 23 are the optimal both in AUC and ES. These parameters are used in Application module.

7.4 drug repurposing by single method in application module

In the Application module, the "single method" option was selected, with CS_OGS.txt provided as the OGS input and topN set to 23. The result, as depicted in the following figure (**Figure 21**), show that among the top 10 drugs in the down-regulation (lower left of the figure), LDN-193189, a selective BMP type I receptor inhibitor, has been reported to exhibit anti-HCC effects (Liang *et al.*, 2021), thereby demonstrating the potential of SSP to identify novel uses for existing drugs.

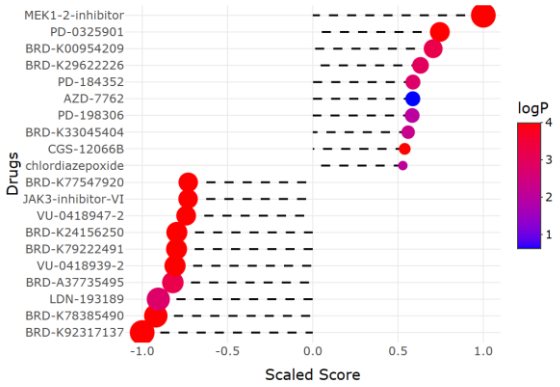


Figure 21 Result of drug repurposing.

7.5 further exploration of drug

To delve deeper into the properties and experimental context of the drug LDN-193189, it is necessary to access comprehensive drug and experimental information beyond what is provided in the figure and table of the Application module. Hence, the SSP offers an Info-data page that allows users to obtain gene expression data and detailed drug information tailored to a specific pharmacotranscriptomic dataset (PTD). By selecting "LINCS HEPG2 10μM 6h (3605 drugs)" on the info-data page, users can access and download comprehensive drug and experimental information by clicking the respective link (**Figure 22**).

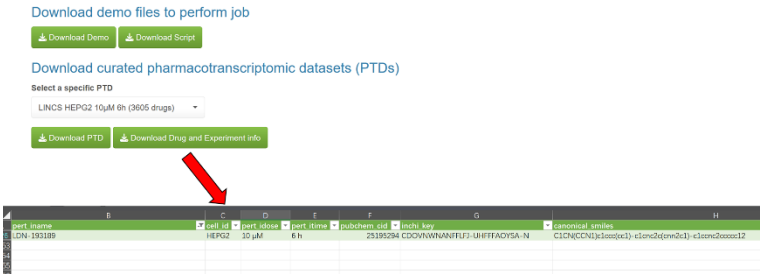


Figure 22 Download information about drug and experiment.

7.6 a simple illustration for other applicable module or methods in these case study

Robustness module: The purpose of the Robustness module is to assess the performance of SSMs in scenarios where drug annotations are limited, such as for subtypes of cancer or rare cancers. In this case study, for example, if we want to study Intrahepatic cholangiocarcinoma, a rare subtype of liver cancer, the drug annotations for the "LINCS HEPG2 10μM 6h (3605 drugs)" dataset are quite insufficient. Therefore, users can proceed to the Robustness module to evaluate the performance of each SSM based on drug self-retrieval and to review the pharmacotranscriptomic dataset (PTD). The optimal SSM and topN values are identified in a manner consistent with the Benchmark module. The optimal method and topN are labelled as in Benchmark module. It should be noted that the performance of SSMs in the Robustness module may differ from that in the Benchmark module. Researchers are strongly encouraged to utilize the Benchmark module, in alignment with their specific fields of study. Should the number of topN genes from the Robustness module exceed the length of the OGS, it is recommended to assess whether the scores obtained from the Robustness module at the corresponding length are close to the optimal values. If not, consideration should be given to replacing the OGS.

SS_cross: When users have two OGSs derived from different sources, they can upload all of them to the SS_cross with properly naming each OGS. Typically, the focus is on the third quadrant of the results, where drugs with negative values (<0) on both the X-axis and Y-axis. It is important to note that we recommend each OGS be evaluated in the Benchmark module. If the optimal topN and SSM for two OGSs are identical or close (with high scores in the same topN or a high ranking in SSM), this indicates a strong match. If not, it is advisable to replace the OGSs. For the plot result, we recommend that users download and annotate figures themselves to identify their drugs of interest. It is worth noting that SSP employs *plotly* package for interactive figures, allowing users to hover their mouse over points to view details of drugs, thus facilitating exploration.

SS_all: When evaluating with the Benchmark module and finding that the performance of SSM is closely aligned, making it difficult to select the most appropriate one, it is advisable to consider all the desirable SSMs that exhibit good performance. In this case study, expect XSum, the performance of other four SSM are very close, so if XSum is the worst, we can combine the four other SSM in SS_all. In addition, if we use Robustness module to determine SSM, SSM over the average line is highly recommend used in SS_all. Please note that SSP with poor performance in the Benchmark and Robustness module is not recommended for inclusion, as it may adversely affect the final outcomes. By uploading the OGS and selecting multiple SSMs that perform well, the results will integrate the outcomes of all considered SSMs and present them accordingly. Typically, the drugs that rank highest in this analysis are often the most promising.

8. Reference

Chen,B. *et al.* (2017) Reversal of cancer gene expression correlates with drug efficacy and reveals therapeutic targets. *Nat Commun*, **8**, 16022.

Chen,S. *et al.* (2021) The phytochemical hyperforin triggers thermogenesis in adipose tissue via a Dlat-AMPK signaling axis to curb obesity. *Cell Metab*, **33**, 565-580.e7.

Cheng,J. *et al.* (2013) Evaluation of analytical methods for connectivity map data. *Pac Symp Biocomput*, 5–16.

Cheng,J. *et al.* (2014) Systematic evaluation of connectivity map for disease indications. *Genome Med*, **6**, 540.

Duan,Y. *et al.* (2020) signatureSearch: environment for gene expression signature searching and functional interpretation. *Nucleic Acids Res*, **48**, e124.

Finn,R.S., Qin,S., *et al.* (2020) Atezolizumab plus Bevacizumab in Unresectable Hepatocellular Carcinoma. *N Engl J Med*, **382**, 1894–1905.

Finn,R.S., Ryoo,B.-Y., *et al.* (2020) Pembrolizumab As Second-Line Therapy in Patients With Advanced Hepatocellular Carcinoma in KEYNOTE-240: A Randomized, Double-Blind, Phase III Trial. *J Clin Oncol*, **38**, 193–202.

Kolde,R. *et al.* (2012) Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics*, **28**, 573–580.

Liang,Z. *et al.* (2021) The binding of LDN193189 to CD133 C-terminus suppresses the tumorigenesis and immune escape of liver tumor-initiating cells. *Cancer Lett*, **513**, 90–100.

Lin,K. *et al.* (2020) A comprehensive evaluation of connectivity methods for L1000 data. *Briefings in Bioinformatics*, **21**, 2194–2205.

Qiu,Z. *et al.* (2019) A Pharmacogenomic Landscape in Human Liver Cancers. *Cancer Cell*, **36**, 179-193.e11.

Rumgay,H. *et al.* (2022) Global burden of primary liver cancer in 2020 and predictions to 2040. *J Hepatol*, **77**, 1598–1606.

Subramanian,A. *et al.* (2017) A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell*, **171**, 1437-1452.e17.

Subramanian,A. *et al.* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences*, **102**, 15545–15550.

Tian,S. *et al.* (2023) Exploring pharmacological active ingredients of traditional Chinese medicine by pharmacotranscriptomic map in ITCM. *Brief Bioinform*, **24**, bbad027.

Yang,C. *et al.* (2022) A survey of optimal strategy for signature-based drug repositioning and an application to liver cancer. *eLife*, **11**, e71880.

Yang,W. *et al.* (2013) Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. *Nucleic Acids Res*, **41**, D955-961.

Yau,T. *et al.* (2023) Nivolumab Plus Cabozantinib With or Without Ipilimumab for Advanced Hepatocellular Carcinoma: Results From Cohort 6 of the CheckMate 040 Trial. *J Clin Oncol*, **41**, 1747–1757.

Zdrazil,B. *et al.* (2024) The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res*, **52**, D1180–D1192.

Zhang,S.-D. and Gant,T.W. (2008) A simple and robust method for connecting small-molecule drugs using gene-expression signatures. *BMC Bioinformatics*, **9**, 258.

Zhu,A.X. *et al.* (2019) Ramucirumab after sorafenib in patients with advanced hepatocellular carcinoma and increased α-fetoprotein concentrations (REACH-2): a randomised, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol*, **20**, 282–296.