



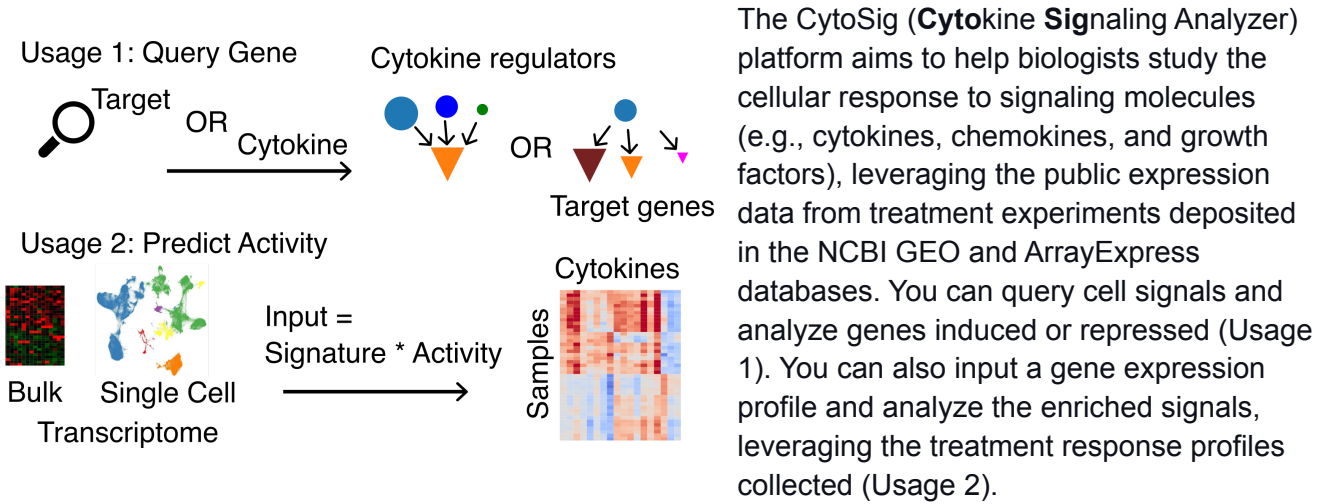
Cytokine Signaling Analyzer

Peng Jiang (peng.jiang@nih.gov)

Table of Content

Introduction	1
Register New User	1
Usage 1: Query genes or cytokines	2
Usage 1.1, Query target genes of a cytokine.	2
Usage 1.2, Query cytokine regulators of a target gene.	5
Usage 2: Predict cytokine activities	7
Input data format	7
Result format	8

Introduction



Register New User

Click the “REGISTER NEW” button in the menu and fill in all fields. After submission, you will receive a confirmation email to activate your account. Commercial users can get a free trial account first before apply a full license.

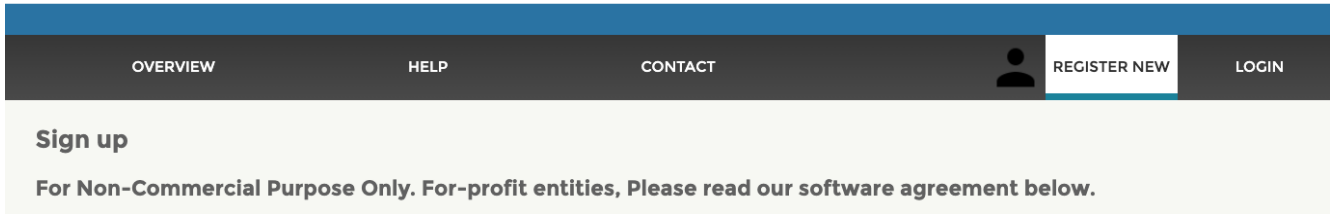
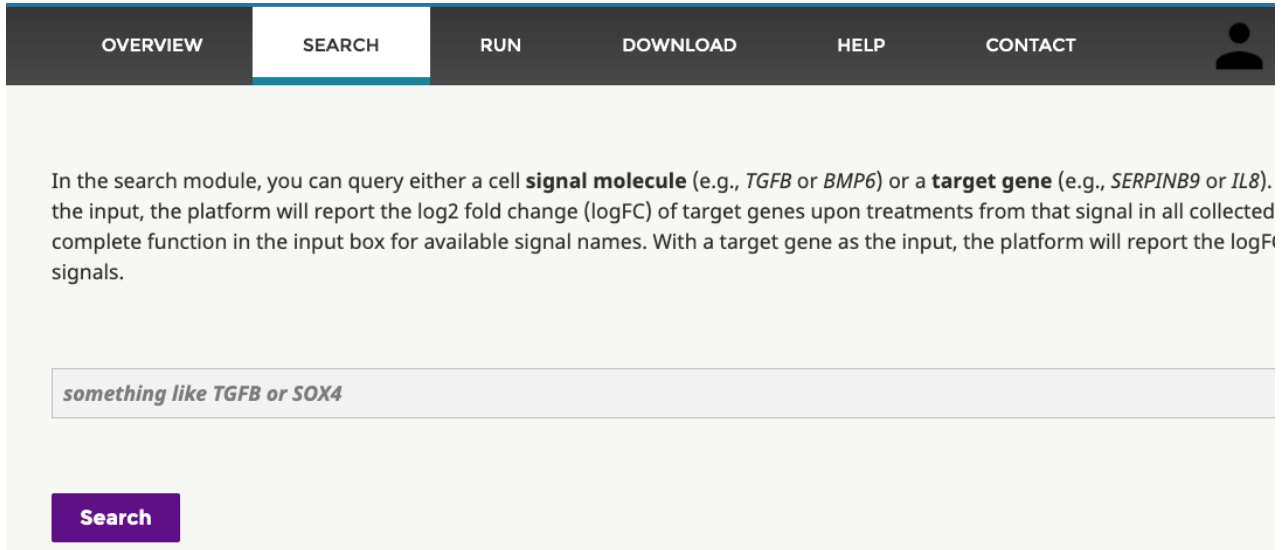


Figure 1. Register new user.

Usage 1: Query genes or cytokines

A demo video is [available through this link](#).

Please click the “SEARCH” menu to enter this module. You can either query target genes of a cytokine or cytokine regulators that regulate many target gene.



OVERVIEW SEARCH RUN DOWNLOAD HELP CONTACT

In the search module, you can query either a cell **signal molecule** (e.g., *TGFB* or *BMP6*) or a **target gene** (e.g., *SERPINB9* or *IL8*). the input, the platform will report the log2 fold change (logFC) of target genes upon treatments from that signal in all collected complete function in the input box for available signal names. With a target gene as the input, the platform will report the logFC signals.

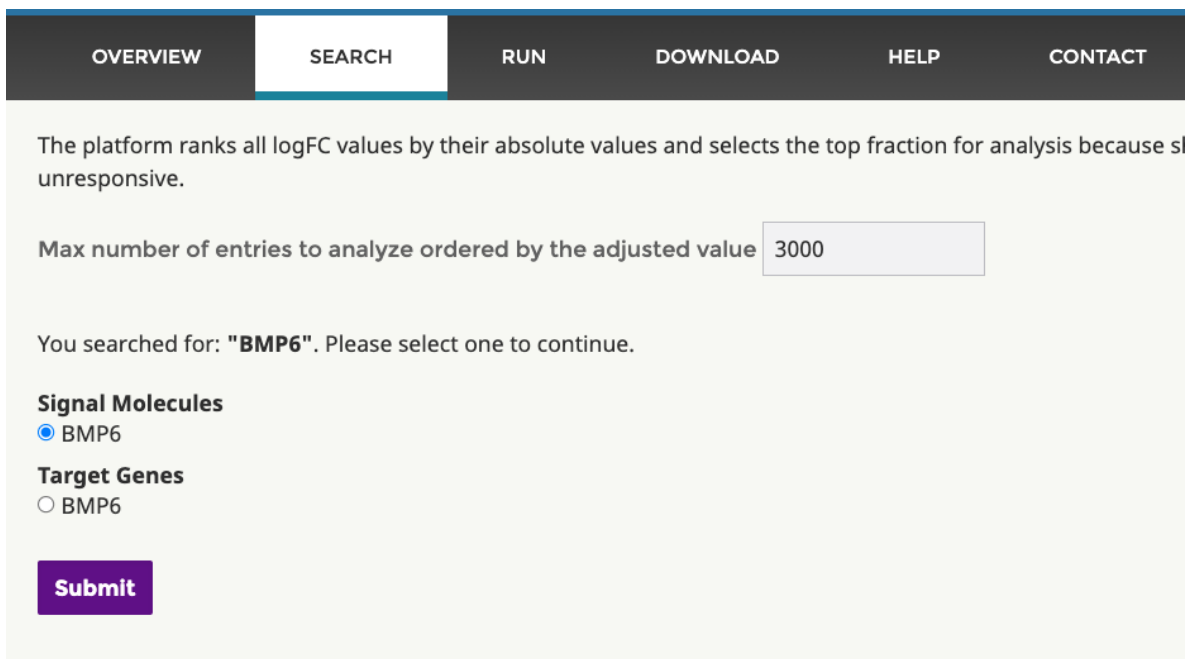
something like *TGFB* or *SOX4*

Search

Figure 2. Search interface.

Usage 1.1, Query target genes of a cytokine.

For example, you can input BMP6 in the search box and click Search. The CytoSig database only included cytokines with data curated from public repositories; thus, the result may not be available for under-studied cytokines. After submitting, you will get a selection interface like this:



OVERVIEW SEARCH RUN DOWNLOAD HELP CONTACT

The platform ranks all logFC values by their absolute values and selects the top fraction for analysis because some are unresponsive.

Max number of entries to analyze ordered by the adjusted value 3000

You searched for: "**BMP6**". Please select one to continue.

Signal Molecules
☒ BMP6

Target Genes
☐ BMP6

Submit

Figure 3. Search result parameter interface.

There are two parameters on this page:

Max number of entries to analyze ordered by the adjusted value: Each cytokine treatment profile may contain logFC values for about 20 thousand human genes. Also, a cytokine may have many treatment profiles from various studies and cell models. Thus, enumerating all results on the web browser will take a significant amount of time. Therefore, the system will rank all target genes by their absolute logFC values and only show top results.

Signaling Molecules or Target Genes: For a cytokine name, they could either serve as a signaling molecule regulating many targets or a target gene with upstream cytokine regulators. Such ambiguity will not exist if the input is only a target gene, for example *SOX4*. For the *BMP6* example, we will select the first option and click Submit. The number indices in the figure are explained in detail below.

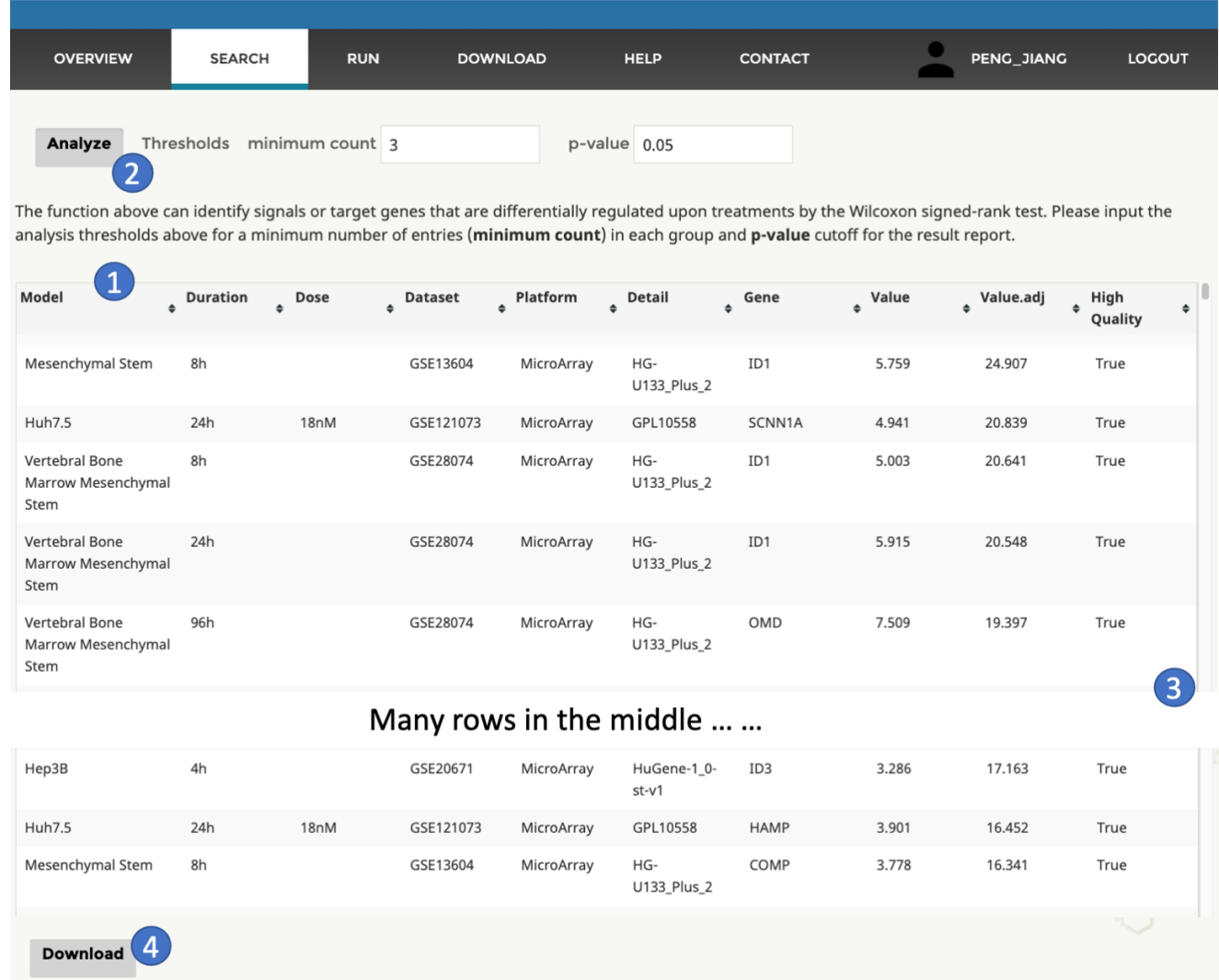
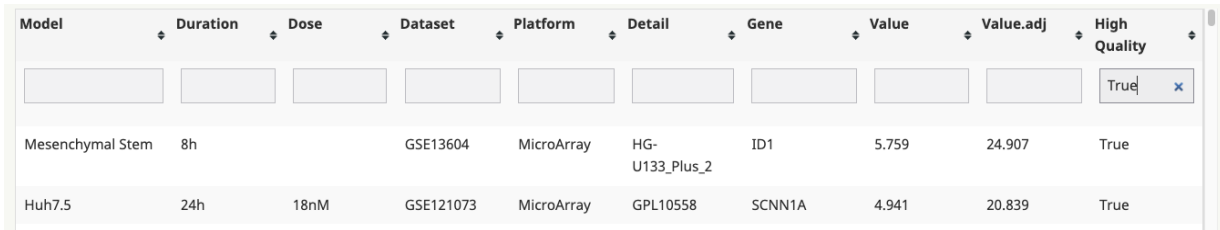


Figure 4. Search results.

1, Header with multiple functions listed below.

- Mouse over each column name will trigger the explanation of each field.
- Mouse click on each column name will sort rows in either ascending or descending (click again) order by column values.

- Row selector will pop up once you put the mouse cursor on the space beneath the header. You may input a keyword to select a subset of rows using the column value. For example, you can input “True” in the box beneath the column “High Quality” to only analyze treatment profiles that passed our quality controls. The row selector will affect the analysis module introduced later (Figure 6).



Model	Duration	Dose	Dataset	Platform	Detail	Gene	Value	Value.adj	High Quality
Mesenchymal Stem	8h		GSE13604	MicroArray	HG-U133_Plus_2	ID1	5.759	24.907	True
Huh7.5	24h	18nM	GSE121073	MicroArray	GPL10558	SCNN1A	4.941	20.839	True

Figure 5. Row selectors.

The meaning of each column is available below.

- **Model:** Cell model used in the cytokine treatment experiment
- **Duration:** Duration of treatment experiment, if available
- **Dose:** Dose of treatment experiment, if available
- **Dataset:** The database ID for the data in a row
- **Platform:** Transcriptomic profiling platform
- **Detail:** Further details of the platform
- **Gene:** Target gene
- **Value:** Log2 fold change (logFC) value between treatment and control conditions.
- **Value.adj:** Adjusted value through normalizing the original logFC by the standard deviation across all logFC values in each treatment profile.
- **High Quality:** This column indicates whether the current dataset has passed our quality controls, testing gene expression correlations between cytokine ligand (or receptor) and target genes across many human bulk-tissue datasets.

2, Significance analysis. This function module will identify differentially regulated target genes of the current cytokine.

For each target gene, the module compares logFC values from different studies with zero through the two-sided Wilcoxon signed-rank test. If the user applied row selectors (Figure 5), this module will only analyze the selected rows. The “minimum count” represents the threshold for the minimal number of studies required for a significant result. The “p-value” represents the threshold for statistical significance to include in the result table, which includes columns: target gene ID, Mean and Standard deviation (Std) of logFC across studies, and p-values (Figure 6). Clicking on the column name will sort rows. For example, BMP6 significantly downregulated CCL2 levels and may upregulate a few other genes across studies.

ID	Mean	Std	p-value
CCL2	-7.93	2.13	0.0431
EVI2A	-6.29	1.43	0.0431
ARNT2	-6.27	1.72	0.0431
CCL7	-3.47	0.405	0.0431
SNAI2	4.21	1.60	0.0431
KAZALD1	5.04	1.11	0.0431
ATOH8	5.38	2.94	0.0431
SAMD11	6.15	1.52	0.0277

Figure 6. Analysis results.

Besides the statistics table above, an interactive summary plot is also available. When you hover the mouse cursor over each dot, the plot will show the cell model and treatment conditions.

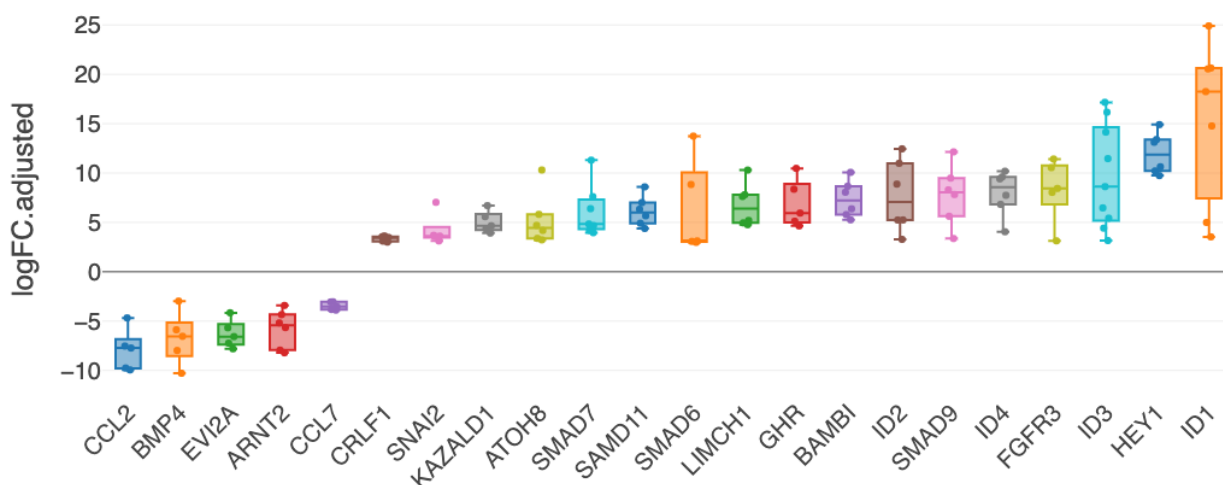


Figure 7. Interactive result plot.

3, Scroll bar. Scroll the big table up and down.

4, Download. Download the table in CSV format. The entire dataset of cytokine treatment response is also available on the [Download page](#).

Usage 1.2, Query cytokine regulators of a target gene.

The second query mode is to find upstream cytokine regulators of a query gene. For example, you can input *IL8* in the search box and click Search. Then, you will see an interface below with all fields explained in Figure 3. First, the system used the standard gene symbol of *IL8*, which is *CXCL8*. Second, we want to see upstream cytokine regulators of *CXCL8*, but not its downstream target. Thus, we select the query as "Target Genes". If you query a target gene that is not cytokine, such as *SOX4*, you will get the interface with only one option of "Target Genes"

OVERVIEW	SEARCH	RUN	DOWNLOAD	HELP	CONTACT
<p>The platform ranks all logFC values by their absolute values and selects the top fraction for analysis because si unresponsive.</p> <p>Max number of entries to analyze ordered by the adjusted value <input type="text" value="3000"/></p> <p>You searched for: "IL8". Please select one to continue.</p> <p>Signal Molecules</p> <p><input type="radio"/> CXCL8</p> <p>Target Genes</p> <p><input checked="" type="radio"/> CXCL8</p> <p><input type="button" value="Submit"/></p>					

Figure 8. Search result parameter interface.

After clicking the submit button, you will get a result table very similar to the one explained in Figure 4 except that the Gene column (representing target genes) in Figure 4 will be replaced with the Signal column (representing cytokine signaling molecules) in Figure 9.

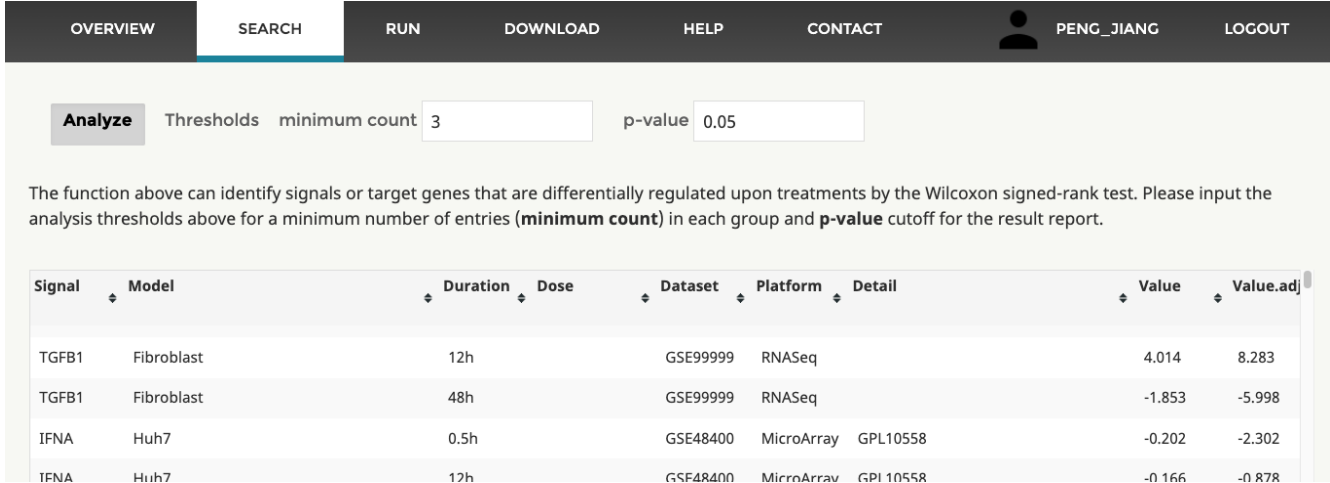


Figure 9. Search results.

ID	Mean	Std	p-value
IL4	-3.85	3.37	4.44e-16
BMP6	-1.64	2.28	0.0469
IFNL	-1.30	2.81	0.0347
IFNB	-0.810	3.15	0.000962
TGFA	1.59	2.86	0.00836
FGF2	2.22	2.73	0.0157
PDGFB	2.27	3.17	0.0409
IL3	2.71	3.77	0.0258
EGF	3.29	5.38	0.000115

After clicking the Analyze button, you will get cytokines that differentially regulate your query gene, with parameters and results similar to Figure 6 and Figure 7, except that the ID column, representing target genes before, will represent cytokine IDs in the current analysis (Figure 11). In the current analysis, the two-sided Wilcoxon signed-rank test is comparing logFC of the same cytokine across different experiments again zero, instead of the same target gene across different experiments (Figure 6). An interactive plot is also available together with the statistical table (Figure 10). Mouse-hovering over each dot will trigger the pop out of the cell model and experimental conditions.

Figure 11. Analysis results.

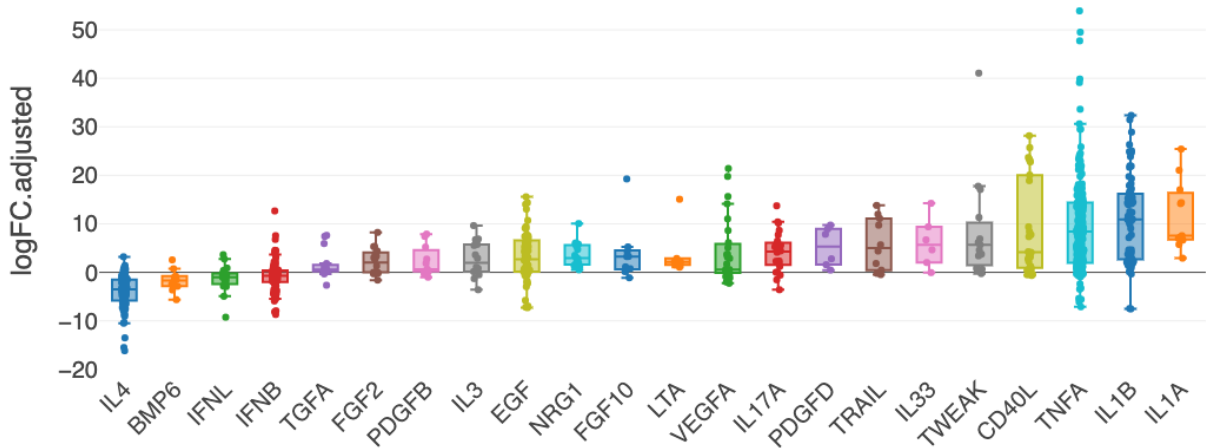


Figure 10. Analysis plots.

Usage 2: Predict cytokine activities

A demo video is [available through this link](#).

The users can predict cytokine signaling activities using the gene expression data of samples through either source code or web application. The source code and usage are available on the [Download page](#). We will introduce the web application below. Please click the “RUN” menu.

OVERVIEW SEARCH **RUN** DOWNLOAD HELP CONTACT

Please upload a matrix file with columns as samples and rows as genes. The platform will predict the activities of signals. The input value separator should be the tab space, and the gene names should follow the standard gene symbols. The input. Please see a [sample tab file](#). Alternatively, you can upload Excel files ([sample CSV](#) or [sample XLSX](#)). The file size should be smaller than 10MB, and the number of samples should be less than 40. If you need to analyze more samples, please [download the source code](#) and run it locally. The expression values, from either RNASeq or MicroArray, should be transformed by $\log_2(x+1)$. x could be FPKM, RPKM, or TPM for RNASeq. For single-cell RNASeq data, we used $\log_2(\text{TPM}/10 + 1)$. We also recommend quantile-normalization across conditions. Some software automatically include all normalizations. We recommend input differential profiles between the two conditions. If data is from a sample collection without pairs, please mean-centralize the value of each gene across all samples.

Task_file:
 No file chosen

Submit

Figure 12. Web application to run cytokine activity prediction.

Input data format

The input data should be an expression matrix with columns as samples and rows as genes. The gene names should follow the standard gene symbols. The platform accepts text or Gzip files as input. For text files, the input value separator should be the tab space. Alternatively, users can use Excel files. A few sample data files are available on the RUN page (Figure 12).

The file size should be smaller than 10MB, and the number of samples should be less than 40. If you need to analyze more samples, please [download the source code](#) and run it locally. The expression values, from either RNASeq or MicroArray, should be transformed by $\log_2(x+1)$. x could be FPKM, RPKM, or TPM for RNASeq. For single-cell RNASeq data, we used $\log_2(\text{TPM}/10 + 1)$. We recommend inputting differential profiles between the two conditions. If data is from a sample collection without pairs, please mean-centralize the value of each gene across all samples. CytoSig does not work on absolute gene expression values, because explanatory variables in the regression model are differential expression profiles from cytokine treatments.

After submitting the input file, you will see the following result. The first result tab is always a summary heatmap of cytokine target activities predicted across all input samples (Figure 13).



OVERVIEW
SEARCH
RUN
DOWNLOAD
HELP
CONTACT
PENG_JIANG
LOGOUT

Summary

NHBE_SARS-CoV-2@Series1

COVID19Lung@Series15

A549-ACE2_SARS-CoV-2@Series16

A549_SARS-CoV-2@Series2

A549_RSV@Series3

A549_IAV@Series4

A549_SARS-CoV-2@Series5

A549-ACE2_SARS-CoV-2@Series6

Calu3_SARS-CoV-2@Series7

A549_RSV@Series8

A549_HPIV3@Series8

NHBE_IAV@Series9

NHBE_IAVdNS1@Series9

1

2

ID	↕ Coef	↕ StdErr	↕ Zscore	↕ Pvalue
IL1A	3.204e-02	4.025e-03	7.955	0.000e+00
IL22	3.507e-02	4.533e-03	7.719	0.000e+00
IL2	3.513e-02	4.227e-03	8.307	0.000e+00
IL1B	3.296e-02	3.307e-03	9.959	0.000e+00
VEGFA	-1.985e-02	4.301e-03	-4.604	0.000e+00
BMP4	-1.876e-02	3.993e-03	-4.707	0.000e+00
VEGFA	-1.985e-02	4.301e-03	-4.605	0.000e+00
EGF	-3.180e-02	3.838e-03	-8.267	0.000e+00

Download 3

The result page has multiple tabs (Figure 14, Label 1), each of which represents an input sample. Please click the tab name to switch results. Clicking the column name in the header of each result table (Figure 14, Label 2) will sort rows using the column value. Clicking the Download button (Figure 14, Label 3) will download the result in CSV format.

Coef: Coefficient of ridge regression with cytokine treatment profiles as explanatory variables and input sample's gene expression as response variables.

Zscore: (Coefficient - Average coefficient from 1000 randomizations) / StdErr

8