L1000CDS2 Help Page

Summary

L1000 Data

Signature Download

Tutorial

***Input***

Search

Examples and Signatures

Configuration

Metadata

Recent Searches

***Result***

Table

Header

Table of Combinations

Substructure Enrichment

API

Contact

**Summary**

L1000CDS² is a LINCS L1000 characteristic direction signature search engine to find consensus signatures that match users’ up and down input gene lists or input signature vectors. The underlying dataset is part of the [LINCS L1000](http://www.lincscloud.org/) small molecule expression profiles generated at the Broad Institute by the [Connectivity Map team](http://lincscloud.org/about/cmap-team/). The differentially expressed (DE) genes of these profiles were calculated using the [characteristic direction](http://maayanlab.net/CD/) method. Depending on the user’s input, L1000CDS² uses either a gene-set method or cosine distance method to run the search. When up/down gene lists are submitted to L1000CDS², the search engine compares the input gene lists to the DE genes computed from the LINCS L1000 data and descriptive information of the top 50 matched signatures is returned. When a signature is submitted to L1000CDS² in the format of "gene symbol, expression value", the search engine calculates a cosine distance between the input signature and every characteristic direction signature in the database, and the top 50 signatures of either the largest (reverse mode) or the smallest (mimic mode) cosine distances are returned. L1000CDS² leverages the efficiency of matrix operations to perform the search and finishes a query against more than 20,000 signatures in less than a decisecond using the gene-set method, or less than 3 seconds using the cosine distance method. The app is developed by the [Ma'ayan Laboratory](http://icahn.mssm.edu/research/labs/maayan-laboratory) at the [Icahn School of Medicine at Mount Sinai](http://icahn.mssm.edu/) for the [BD2K-LINCS-DCIC](http://www.lincs-dcic.org/) and the [KMC-IDG](http://commonfund.nih.gov/idg/fundedresearch) NIH funded projects.

**L1000 mRNA Expression Data**

The L1000 mRNA gene-expression dataset is generated as part of the [Library of Integrated Network-based Cellular Signatures (LINCS) project](http://www.lincsproject.org/), a [Common Fund](https://commonfund.nih.gov/) program run by NIH. LINCS aims to systematically profile the molecular and phenotypical outcomes of agent perturbed human cells. The variety of agents includes chemical compounds, different micro-environments, endogenous ligands, gene knockdown and overexpression. The L1000 dataset comprises of over a million gene expression profiles of chemically or genetically perturbed human cell-lines. The L1000 technology developed at Broad Institute measures only ~1000 genes in each sample and estimates the expression of the rest of the transcriptome using an empirical model. In this way, the technology significantly reduces the cost and makes large-scale gene expression profiling possible. L1000CDS² currently includes a subset of the chemically perturbed gene expression profiles, specifically the profiles in the CPC and CPD peturbagen groups. After averaging replicates and removing insignificant signatures, we remained with 22,926 signatures in the L1000CDS² database.

### L1000 CD signature download

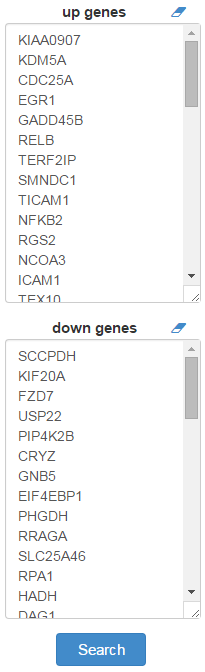
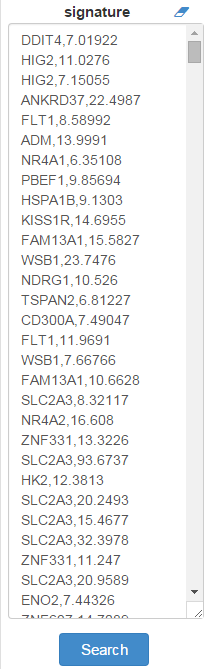
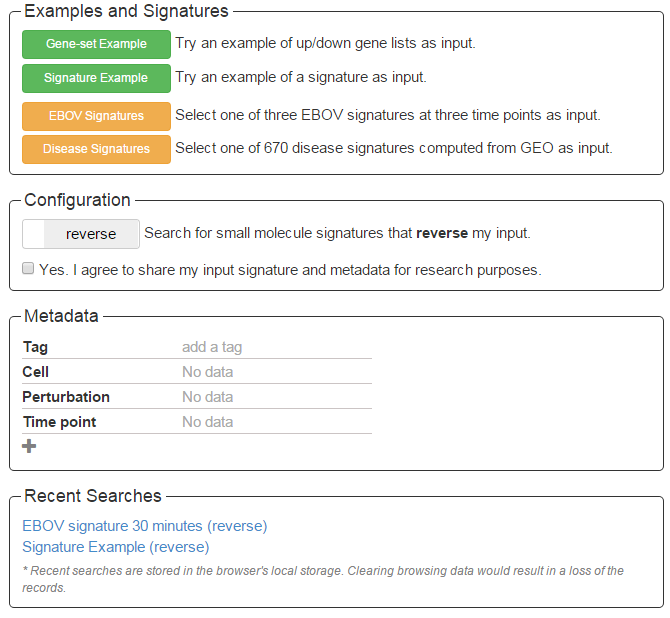
38,9031 L1000 characteristic direction signatures including the aforementioned CPC and CPD subset can be downloaded [here](http://amp.pharm.mssm.edu/public/L1000CDS_download/).

**Tutorial**

*Input*

**Search**

The entry point into L1000CDS² is to paste up/down gene lists into the up/down gene text boxes (Fig. 1) or paste a signature (Fig. 2) into the up gene text box. A signature is a list of genes and their differential expression values separated by a comma. The search button will only become enabled when both up/down genes text boxes are filled by gene lists, or when the up gene textbox is filled with a signature. Clicking the Search button, the information for the top 50 signatures will be displayed in a [table](http://localhost:8182/L1000CDS2/help/#table) in a new page.

*Fig3.* *Screenshot of the four input form sections: Examples and Signatures, Configuration, Metadata and Recent Searches*

*Fig2. Screenshot of the input up gene text box filled with a signature.*

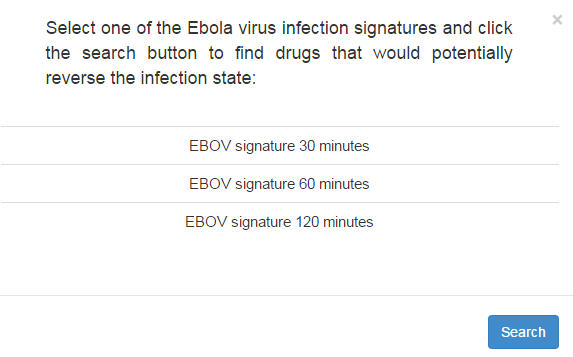
*Fig1. Screenshot of the input text boxes filled with up/down gene lists.*

**Examples and Signatures**

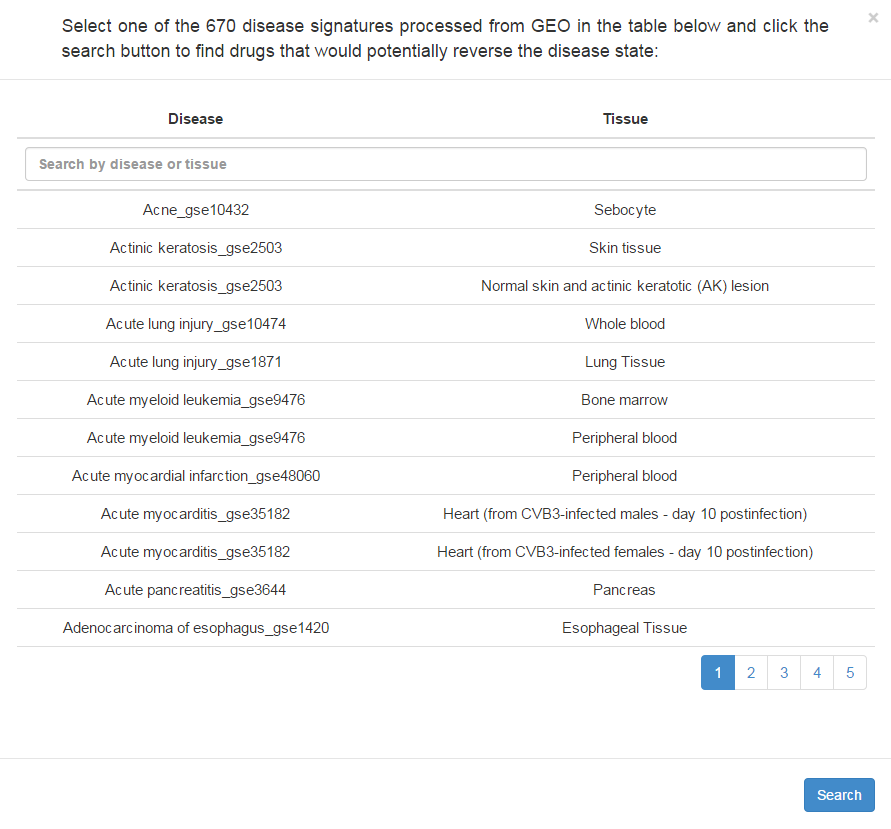
Clicking the Gene-set Example button will fill in an example of up/down gene lists in the text boxes for a demo search using the gene-set method. Clicking the Signature Example button will fill in an example of a signature in the up gene text box for a demo of searching with the cosine distance method.

Clicking the EBOV Signatures button will open a table with three EBOV signatures at three time points (Fig. 4). Selecting a signature by a single click and clicking the Search button can be used to perform a cosine distance search. The selected signature will be automatically filled in the up gene text box and the associated metadata will be filled in the [metadata](file:///C:\Users\Avi-Maayan\AppData\Local\Microsoft\Windows\Temporary%20Internet%20Files\Content.Outlook\6172SJPM\to%20metadata%20section) section.

Clicking on the Diseases Signature button will open a table of disease names and their tissue types including the GEO ID referring to the original study for the disease signature (Fig 5). The table is searchable by disease name or tissue type and can be sorted by either column. Each row represents a differential expression signature of a disease and consists of the differentially expressed genes and their expression values in the disease compared with the relevant normal tissue. These signatures were calculated from gene expression data deposited in GEO. Clicking on a row will select that disease. Clicking on the search button will fill in the signature of the selected disease in the up gene text boxes and the associated metadata in the metadata section and perform a search for small molecules at the same time.



*Fig 4. Screenshot of the EBOV signature table*



*Fig 5. Screenshot of the disease signature table*

**Configuration**

***Mimic/reverse****:* Clicking the mimic/reverse slider can be used to switch between reverse and mimic modes. In the mimic mode for the gene-set search, the input up genes are intersected with the up genes of the gene expression profiles in the L1000CDS² database, and the input down genes are intersected with the down genes for each entry in the database. In the reverse mode for the gene-set search, the input up-genes are intersected with the down-genes, and the input down-genes are intersected with the up-genes. When a cosine distance search is performed, the top 50 signatures will be those of the smallest cosine distances from the input signature in the mimic mode, or those with the largest cosine distances in the reverse mode. **The default mode is reverse.**

***Drug Combinations:* L1000CDS2 also provides the function to search for drug combinations. Check the “Search for drug combinations” checkbox to enable this function. When searching for combinations L1000CDS2 compares every possible pair between the top 50 ranked perturbations and computes a synergy metric for each pair. In a gene-set search, the synergy is simply calculated as the combined overlap of two perturbation signatures with the input. In a cosine distance search, the synergy is calculated as the** orthogonality between two perturbation signatures. The rational is that if two perturbations are orthogonal they probably impart the same effect from two independent pathways.

***Share:*** Users can share their input signatures and metadata so other users can query the signatures and gene sets they submitted. To make submitted input gene sets and signatures “public” for research purposes, a checkbox is available (Fig. 3). The default is set to “No” such that users’ input is made private. Users can still share their input lists, signatures, metadata and results using the share icon on the result page [(Fig. 6).](http://10.91.53.50/L1000CDS2/help/#tableFig) Clicking on the share icon produced a permanent URL that can be shared through e-mail, publications and other documentation.

**Metadata**

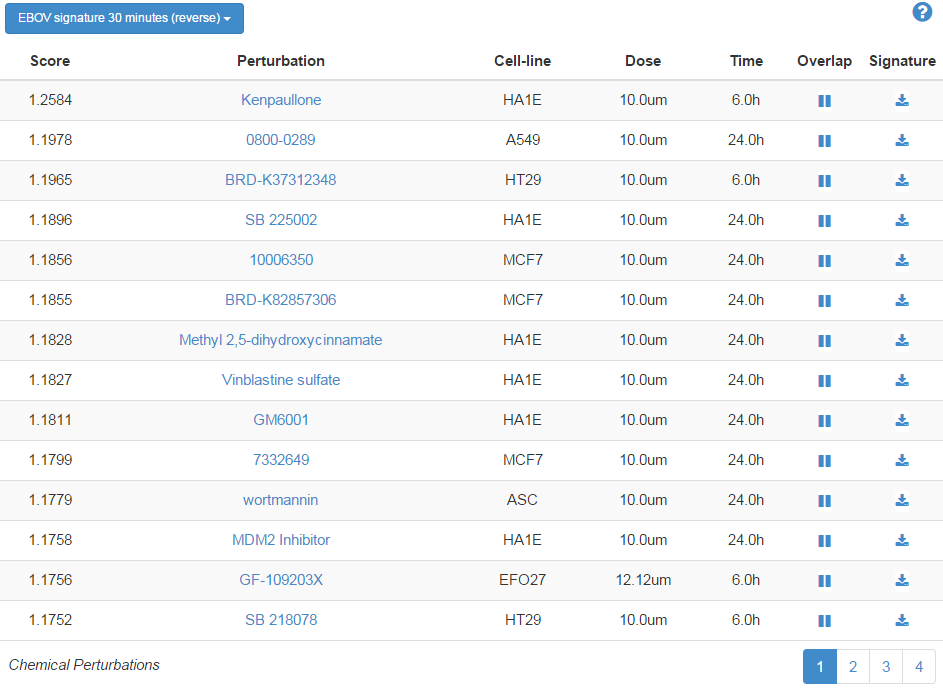
Any metadata associated with the input signature can be entered in the metadata section. By default, the section provides four input fields for metadata: Tag, Cell, Perturbation and Time Point (Fig. 3). Users can add new input fields for additional types of metadata by clicking the plus sign at the bottom or remove one by clicking the minus sign on the right of each row. The minus sign will only appear when the mouse cursor is hovering over a row. The tag field is used to enter few words which are most descriptive of the input signature.

**Recent Searches**

Recent search history will be displayed in this section as links (Fig 6). Clicking a link will show the results for that search. Recent searches are stored in the browser's local storage buffer. Clearing browsing data would result in a loss of these records. A maximum of 50 recent searches are stored for each user’s browser.

*Result*

**Table**

****

*Fig.6 Screenshot of the paginated results table and the metadata header*

The search results are rendered as a paginated table with 14 entries per page (Fig 6). Each entry provides seven pieces of information about the signature: rank, metric, perturbation, cell-line, dose, time point and overlap with the input.

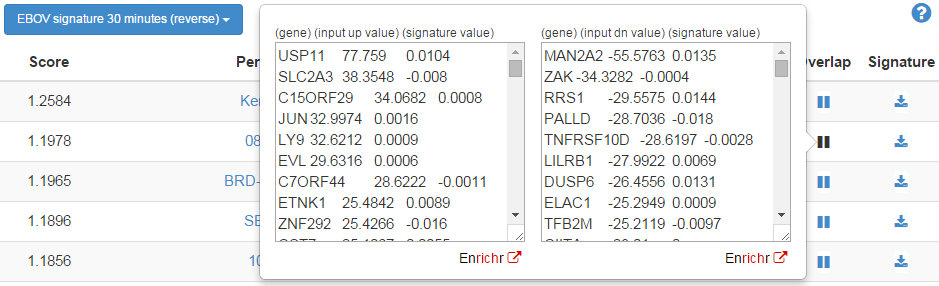
***Rank***: The rank of a signature based on the metric.

***Metric:*** The metric is the overlap between the input DE genes and the signature DE genes divided by the effective input. The effective input is the length of the intersection between the input genes and the L1000 genes since some input lists contain genes that are not present in the L1000 dataset. This includes all ~22,000 L1000 genes, not just the measured ~1000. This metric is visualized as a venn diagram in the venn column. For the cosine distance search, the metric is the cosine distance between the input signature and the L1000 characteristic direction signatures. The consensus signatures are sorted by their metrics in descending order in reverse mode and in ascending order in mimic mode. This metric is visualized as an angle in the α column.

***Perturbation:*** The perturbation column shows the names of the chemical perturbations.  The perturbation column shows the names of the chemical perturbations. User can click on the three **L** **P** **D** icons to look up perturbations in [LIFE](http://life.ccs.miami.edu/life/), [PubChem](https://pubchem.ncbi.nlm.nih.gov/) and [DrugBank](http://www.drugbank.ca/) that catalog detailed information about the chemical compound. Not every perturbation is available in all the three resources. There is also a signature column in the table.

***Cell-line, Dose and Time***: The cell-line, dose and time point used for generating the signature.

***Overlap***: Clicking the overlap button will show the overlapping genes (and their values) in two text boxes. If the user input type is up/down gene lists, the first box will show the overlap genes between the input up genes and the signature up (down) genes and the second will show the overlap between the input down and the signature down (up) in mimic (reverse) mode. If the input is a signature, the first box will show genes with a positive value in input and their values in the signature; the second box will show genes with a negative value in input and their values in the signature. The signature values and input values in both boxes are expected to be mostly in the same sign in mimic mode and in the opposite sign in reverse mode. The Enrichr button under each text box will send the genes to Enrichr for enrichment analysis.



*Fig.7 Screenshot of the overlap between the input signature and a small molecule signature*

***Download***: Clicking the download button in that column will download all the information about a signature as a JavaScript Object Notation (JSON) file.

**Header**

On top of the table is a header bar that provides various functions (Fig. 6):

***Reanalyze:*** Clicking this button redirects user back to the input page with input loaded in the input textboxes. User can then reanalyze the input using different configurations or modify the associated metadata. This function also has a bearing on sharing results with others since other people can get the input and reanalyze it with different settings.

***Tag:***

Clicking the header will show the metadata of the input.

**Combination Table**

A header that displays the tag and search mode is shown on the upper left corner of the result page (Fig. 6). Clicking the header will show the metadata of the input.

**API**

**[POST]** [**http://amp.pharm.mssm.edu/L1000CDS2/query**](http://amp.pharm.mssm.edu/L1000CDS2/query)

**Gene-set Search**

**PAYLOAD (content-type: application/json)**

|  |  |  |
| --- | --- | --- |
| data | Object | An object that saves the input up/down gene lists. |
| data.upGenes | [String] | An array of up-regulated genes. |
| input.data.dnGenes | [String] | An array of down-regulated genes. |
| config | Object | An object that saves the search configuration. |
| config.aggravate | Boolean | True to perform the search in aggravate mode and False in reverse mode. |
| config.seachMethod | String | “geneSet” |
| config.share | Boolean | True to agree to share input data and metadata. |
| metadata | [Object] | An array of objects that saves the metadata of the input. |
| metadata[ ].key | String | The field name of a piece of metadata. |
| metadata[ ].value | String | The value of a piece of metadata. |

Response:

|  |  |  |
| --- | --- | --- |
| shareId | String | Unique ID for sharing the search results |
| topMeta | [Object] | Descriptive information of the top 50 consensus signatures |
| topMeta[ ].score | Number | Score |
| topMeta[ ].cell\_id | String | Cell-line |
| topMeta[ ].pert\_desc | String | Perturbation name |
| topMeta[ ].pert\_id | String | Unique identifier for a perturbation |
| topMeta[ ].pert\_time | String | Time point |
| topMeta[ ].pert\_time\_unit | String | Time point unit |
| topMeta[ ].pert\_dose | String | Dose |
| topMeta[ ].pert\_dose\_unit | String | Dose unit |
| topMeta[ ].sig\_id | String | Unique identifier for a signature |
| topMeta[ ].overlap | Object | The overlapping genes between input DE genes and signature DE genes. |
| topMeta[ ].overlap.up/dn (reverse mode only) | [String] | The overlap between input up genes and signature down genes. |
| topMeta[ ].overlap.dn/up (reverse mode only) | [String] | The overlap between input down genes and signature up genes. |
| topMeta[ ].overlap.up/up (mimic mode only) | [String] | The overlap between input up genes and signature up genes. |
| topMeta[ ].overlap.dn/dn (mimic mode only) | [String] | The overlap between input down genes and signature down genes. |

**Cosine Distance Search**

**PAYLOAD (content-type: application/json)**

|  |  |  |
| --- | --- | --- |
| data | Object | An object that saves the input signature. |
| data.upGenes | [String] | An array of input genes. |
| input.data.dnGenes | [Number] | An array of input values that match to each input gene. |
| config | Object | An object that saves the search configuration |
| config.aggravate | Boolean | True to perform the search in aggravate mode and False in reverse mode. |
| config.seachMethod | String | “CD” |
| config.share | Boolean | True to agree to share input data and metadata. |
| metadata | [Object] | An array of objects that saves the metadata of the input |
| metadata[ ].key | String | The field name of a piece of metadata. |
| metadata[ ].value | String | The value of a piece of metadata. |

Response:

|  |  |  |
| --- | --- | --- |
| shareId | String | Unique ID for sharing the search results |
| uniqInput | Object | Unique input genes overlapped with L1000 genome and their input values. |
| uniqInput.up | Object | Unique input up genes and their input values. |
| uniqInput.up.genes | [String] | Unique input up genes. |
| uniqInput.up.vals | [Number] | Values of unique input up genes. |
| uniqInput.dn | Object | Unique input down genes and their input values. |
| uniqInput.dn.genes | [String] | Unique input down genes. |
| uniqInput.dn.vals | [Number] | Values of unique input down genes. |
| topMeta | [Object] | Descriptive information of the top 50 consensus signatures |
| topMeta[ ].score | Number | Score |
| topMeta[ ].cell\_id | String | Cell-line |
| topMeta[ ].pert\_desc | String | Perturbation name |
| topMeta[ ].pert\_id | String | Unique identifier for a perturbation |
| topMeta[ ].pert\_time | String | Time point |
| topMeta[ ].pert\_time\_unit | String | Time point unit |
| topMeta[ ].pert\_dose | String | Dose |
| topMeta[ ].pert\_dose\_unit | String | Dose unit |
| topMeta[ ].sig\_id | String | Unique identifier for a signature |
| topMeta[ ].overlap | Object | The signature values of unique input genes overlapped with L1000 genome. |
| topMeta[ ].overlap.up | Object | The signature values of unique input up genes. |
| topMeta[ ].overlap.dn | Object | The signature values of unique input down genes. |

Python Example:

….

Contact

Please contact Avi Ma’ayan and Qiaonan Duan for comments, suggestions, and support:[avi.maayan@mssm.edu](mailto:avi.maayan@mssm.edu), [qiaonan.duan@mssm.edu](mailto:qiaonan.duan@mssm.edu).