Integrative LINCS (iLincs): Connecting diseases, drugs and mechanisms of actions

Marcin Pilarczyk, Mehdi Fazel Najafabadi, Michal Kouril, Naim Mahi, Nicholas Clark, Shana White, Mark Bennett, Wen Niu, John Reichard, Juozas Vasiliauskas, Jarek Meller, Mario Medvedovic
Laboratory for Statistical Genomics and Systems Biology, Division of Biostatistics and Bioinformatics, Department of Environmental Health, University of Cincinnati College of Medicine

iLINCS (Integrative LINCS) is an integrative web platform for analysis of omics data and signatures of cellular perturbations. The portal consists of biologists-friendly user interfaces for finding and analyzing datasets and signatures, backend databases with a large collection of datasets (>3,000), pre-computed signatures (>200,000) and their connections (>2*10^9). The portal integrates R analytical engine via several R tools for web-computing (rserve, opencpu, shiny, rgl) and other public domain web tools and open-source applications (e.g., FTreeView, Enrichr, L1000CDS^2) into a coherent web platform for omics data analysis. Analytical tools are organized into three interconnected analytical workflows. The "Dataset" workflow facilitates comprehensive analysis of primary omics datasets. In a typical use case, the user starts with an omics dataset of interest (e.g., GEO dataset corresponding to a disease of interest), performs differential gene expression analysis to construct the signature of the disease. Performs enrichment, pathway and network analysis of differentially expressed genes. Identifies "connected" drug signatures that can implicate a potential therapeutic agent for the disease. The "Signatures" workflow facilitates "connectivity analysis" with a large collection of pre-computed signatures that include LINCS drug perturbation signatures, ENCODE transcription factor binding signatures and a library of "disease related signatures" extracted from public domain omics datasets. User can either select one or more pre-computed signatures, or upload their own signatures to use in the analysis. One of the use-cases involves uploading a custom disease signature, identifying the connected LINCS chemical perturbagen signatures which can then provide putative agents for treating the disease. The "Genes" workflow starts with a user supplied list of genes which are then used to query and analyze primary data and pre-computed signatures. The portal can be accessed freely and does not require user registration (http://ilincs.org).

- Paste or construct a gene list
- Select a LINCS or non-LINCS dataset
- Query, analyze and visualize gene/ protein expression profiles for selected genes in the selected dataset

- Select a dataset
- Construct a signature
- Functional enrichment and pathway analysis
- Correlation with signature libraries
- Use signature to interrogate a different dataset

- Find or upload signature
- Find concordant signature
- Analyze expression/binding profiles of concordant signatures
- Assess biological underpinning of groups of concordant signatures

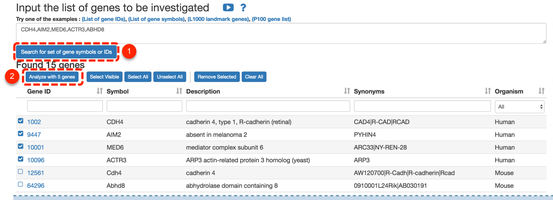
- Select a dataset
- Construct a signature
- Functional enrichment and pathway analysis
- Correlation with signature libraries

GENES

What are my genes doing in LINCS data?

Query, analyze, and visualize gene/protein expression profiles for the selected genes in the selected dataset and signatures.

> Select Genes



Beginning with an list of genes, iLINCS users may explore portals datasets and signatures

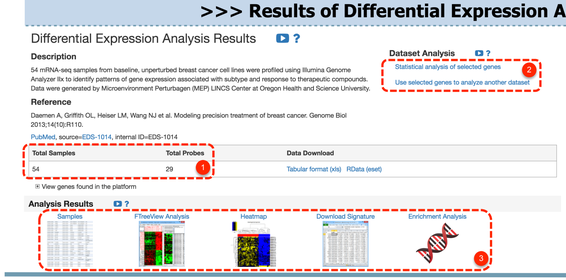
List of genes may be inputted using gene symbols or gene IDs. They can be inputted in separate lines as well as separated with other characters, such as a space, tab, or comma. After submitting a list (1), it can be refined to match ones which are relevant to desired species or other criteria.

>> Select iLINCS Datasets or Signatures to Analyze

Submitted list of genes may be used to analyze and visualize gene/protein expression profiles across datasets. Submitted list could also be used to enrich the signatures or perform **Signature Group Analysis** with the user selected genes (1).

Datasets may be chosen directly from the list or using **iLINCS** search capabilities across **LINCS** and **non-LINCS** datasets (2).

>>> Results of Differential Expression Analysis



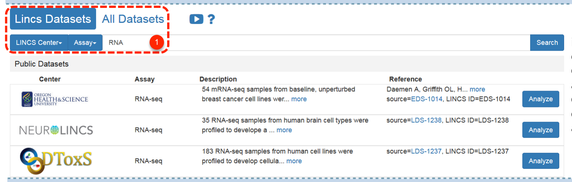
If any of the submitted genes are found in the dataset, preliminary results are shown for all samples (1). To further specify a particular subset of samples in the inquiry, narrow down the list of samples by selecting **Statistical analysis of selected genes** and use inclusion/ exclusion criteria for the sample selection. Genes that are found in both the selected list and dataset may be further used to query other datasets and signatures (2). The results can be visualized and further analyzed by the tools described below (3).

DATASETS

Mining LINCS datasets

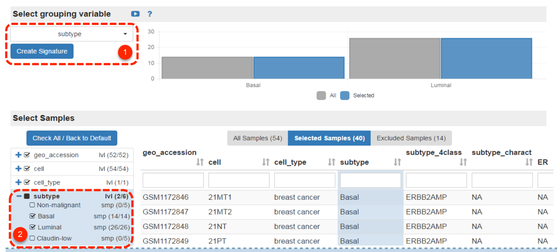
Differential gene/protein analysis, functional enrichment analysis, pathway analysis, correlation with signatures libraries, use newly created signature to interrogate another dataset (LINCS or non-LINCS).

> Select Dataset



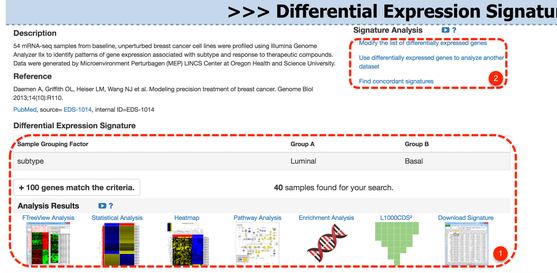
iLINCS allows users to choose among currently released LINCS datasets. Some datasets may be available for certain authorized personnel. Aside from LINCS datasets, there are more than 3000 others (e.g., **TCGA, GEO, CGH**) available to be analyzed (1).

>> Create a Signature



Create a **Differential Expression Signature** using inclusion/exclusion criteria and **sample grouping variable** for the sample selection (1). To create a signature, **only two** groups can be selected. However, the users are able to select additional criteria to focus on particular subsets (2). Currently selected samples counts are displayed on a bar plot and can be explored in a tabular view.

>>> Differential Expression Signature



A signature is generated using the top 100 differentially expressed genes, but may be revised by using **Construct another set...** according to a P-value and a differential expression value cut-off. Results can be vizsalized or analyzed with other **iLINCS** or **DCIC** tools (1)

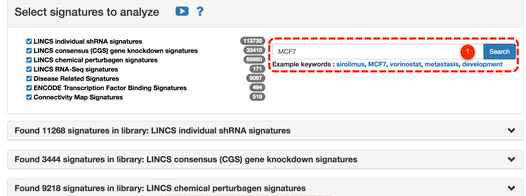
Use differentially expressed genes... allows users to analyze another dataset or signatures with the extracted list of genes, as described in the "Genes" pipeline. **Find concordant signatures** performs similarity tests against selected library of signatures (2).

SIGNATURES

Mining and explaining perturbation signatures

Interrogate signatures similar to the signature of interest, assess biological underpinning of groups of concordant signatures.

> Signature Search



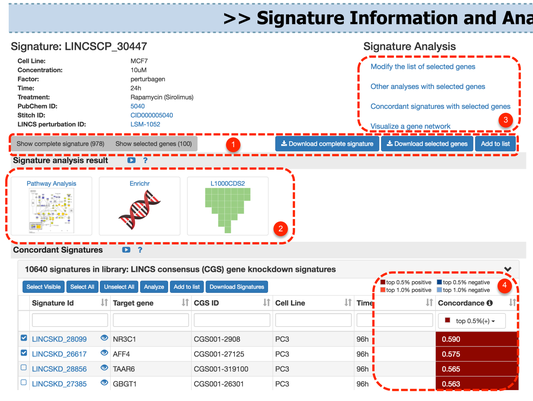
iLINCS provides access to over 200,000 pre-computed signatures organized into 7 libraries based on their origin. Libraries may be searched with by drug, cell name or other desired keywords (1).

Besides **Search Term**, iLINCS provides a vast list of **Pharmacological Actions** to choose from as well as an option to **upload a signature file** that can be compared to signature libraries.

Users may **download signatures of interest** from the results table and can also perform **Group Analysis** of signatures (2).

Information for signatures are provided upon selection (3).

>> Signature Information and Analysis



The **Signature Information and Analysis** page provides information concerning the selected signature. Signature data may be explored and downloaded by the user (1).

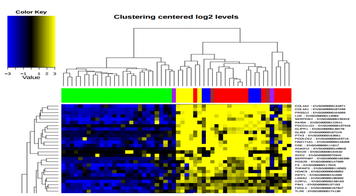
By default one hundred genes with the lowest p-values are selected for the signature. These can analyzed with a variety of **iLINCS** and **DCIC** tools (2, 3). The selection of genes may be revised based on the volcano plot (3).

Every signature is displayed with the list of "concordant" or "connected" signatures calculated based on either Pearson's correlation or TREG Z-score (4).

VISUALIZING ANALYSIS RESULTS

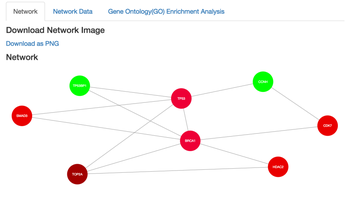
Heatmap

Heatmaps illustrate differential expression values between certain genes and samples, which are clustered according to chosen data properties.



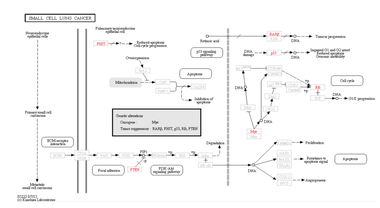
SigNetA

Signature Network Analysis (SigNetA) allows users to illustrate and explore gene subnetworks that were determined based on enrichment analysis.



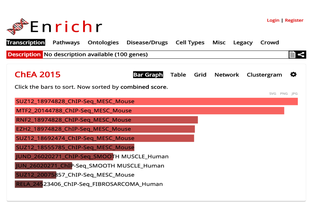
Signaling Pathway Impact Analysis

SPIA uses the information from a list of differentially expressed genes and their log-fold changes in order to identify biological pathways most relevant to the condition under the study.



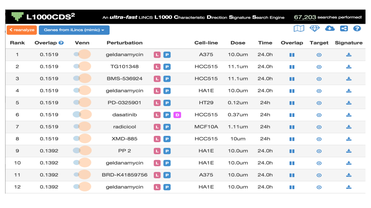
Enrichr

Enrichr is a web-based enrichment analysis tool providing various types of visualization summaries of collective functions of gene lists.



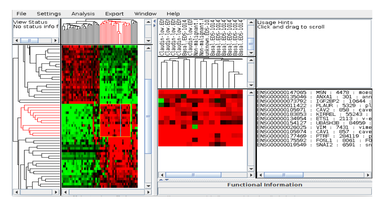
L1000CDS²

L1000CDS² is a LINCS L1000 characteristic direction signature search engine. It enables users to find consensus L1000 small molecule signatures that match user input signatures.



Functional TreeView

FTreeView allows users to interactively view cluster analysis results of microarray data. An extracted gene list may be exported to a file or passed to NCBI, Enrichr or DAVID.



NIH LINCS PROGRAM
DATA COORDINATION AND INTEGRATION CENTER FOR LINCS-BD2K
This work has been funded by the LINCS-BD2K grant U54HL127624.
The NIH Common Fund
DATA COORDINATION AND INTEGRATION CENTER