HMS LINCS Database - Tutorial

(updated 2013-03-20)

This document provides a tour through the HMS LINCS Database. It is a "learn by example" guide where we show some of the functionality of the interface.

Overview

The HMS LINCS Database has a web interface that provides information for the experimental reagents (small molecule perturbagens, cells, and proteins) as well as for the experimental datasets of the HMS LINCS project. In addition to allowing the user to query for and view each of these items independently, the interface also displays the connections between these items. The first section below provides a tutorial and examples for usage of the web interface.

The HMS LINCS Database can also be accessed through a **programmatic API**, using the information given in the HMS LINCS Database URL Scheme and Access Guide. The second section below will provide a tutorial with examples for usage of the programmatic interface.

1. Web Interface tutorial methodology:

- For this tutorial, first click on the link on the left side of the screen for each of the numbered items below (Small Molecules, Cells, Proteins, Datasets).
- This will display a table of values for each item, with the search box field on the top.
- After clicking on the proper item, enter the **search term** given, which will display a particular entity or set of entities.
- Clicking on the link for each result will display the detail page for this type of item. The **discussion** will proceed with an explanation of the detail page.

1. Small Molecules
    a. Text based search: 1st: "sorafenib" (HMSL10008); 2nd: "neratinib" (HMSL10018)
    b. Discussion: HMSL10008
        i. The detail information section at the top shows the chemical structure information and provides links to external listings for the compound (PubChem, ChEBI, ChEMBL, LIFE).
        ii. The Nominal Targets section brings in the results of the literature review (dataset HMSL20000), for this compound it indicates that Sorafenib is an inhibitor of B-Raf and KDR.
        iii. The KINOMEscan image is shown next.
        iv. The Batch Information section provides detail for each batch of this compound that is tracked in the HMS LINCS DB.
            1. Clicking on one of the batch information links expands that section to include specific batch information (provider, catalog id...).

v. Orthogonal information about which studies have included this compound can be seen by scrolling to the bottom of the detail page and viewing the "Datasets" table. Sorafenib (HMSL10008) is of interest because it has been annotated in both the KiNativ and the KINOMEscan studies, as well as being further studied in a MGH (CMT) Growth Inhibition Assay.
vi. Neratinib (HMSL10018) is the subject of a KINOMEscan dataset, and also has been studied in a MGH (CMT) Growth Inhibition Assay.

c. Structure search: Searches are performed using the Pubchem power user gateway REST interface to locate CID's for matching compounds. These CIDs are then cross-referenced with the HMS LINCS DB to find matching compounds.
   i. Input can be in the form of a SMILES string or an sdf file.
   ii. Search Modes:
      1. Identity: search for identical structures (same tautomer).
      2. Similarity search for similar structures (2D Tanimoto equation). See Pubchem Identity and Similary Search Help.
      3. Substructure: search for a for particular chemical structure pattern. See Pubchem Substructure Search Help.
   iii. Example: Using SMILES: CC(N1C=NC2=C1N=C(N[C@H](CC)CO)N=C2NCC3=CC=CC=C3)C
      1. Choose "Identity" base searching
      2. results in HMSL10001 (Roscovitine) being displayed

2. Cells
   a. Search term: 1st: "IA-LM" (50017); 2nd: "HEC-1" (50014) (HEC-1)
   b. Discussion: HMSL50017
      i. The top section gives detail information about the cell line: provider and provider catalog identifier, host organism and organ, as well as disease information, genetic information, culture conditions, and literature references for mutations, if available.
      ii. The bottom section lists orthogonal information about which studies use this cell line in their assay (Moerke 2 Color Apoptosis, Tang Proliferation/Mitosis, MGH (CMT) Growth Inhibition Assays).
3. Proteins
   a. Search term: 1st: ABL1 (200002); 2nd: AKT1 (200034, 200482, 200637)
   b. Discussion: search for AKT1 brings up 3 records (KINOMEscan, KiNativ, protein target).
   c. Discussion: search for "ABL1" finds ABL1 and all ABL1 mutants (27 records total, 25 KINOMEscan and 2 KiNativ).
4. Datasets
   a. Search term: "apoptosis"
   b. Discussion: this results in 3 datasets, including 1 Tang datasets: "HMSL20003 - Tang Mitosis/Apoptosis ver.II".
   c. Clicking on 20003 gives the detail view for this dataset. The dataset detail pages consists of the detail view as well as separate tabs for (where they exist) "Small

Molecules studied", "Cells studied", "Proteins studied". Additionally, there will always be a "Results" tab.

i. Details
    1. Dataset lead investigator and lab.
    2. Summary and protocol information.

ii. Small Molecules Studied, Cells Studied, Proteins studied.
    1. Each of these tabs shows a table with the respective entity for this assay only.

iii. Results
    1. Experimental result data for this assay:
        a. Information for each result: the small molecule batch used, the cell line tested, and the protein information, if applicable. Additional information such as the library plate and well can be found in later columns.
        b. Experimental result data:
            i. Microscopy images (via the OMERO server).
                1. Note: additional navigation and features are available in the OMERO interface (e.g. see "split channel" view).
                2. Image downloads: see "5. Microscopy Images" section below.
            ii. Derived information: cell counts, cell phenotype counts.

d. 2nd search term: "sorafenib"

e. Discussion: results in 2 datasets - Sorafenib KINOMEscan (20020) and Sorafenib KiNativ -- dose response experiment (20087)

5. Microscopy images are provided with the dataset views (via the OMERO server). To obtain complete the complete set of images for a dataset, please contact Jay_Copeland @ hms.harvard.edu

2. Programmatic API usage:

- The programmatic interface provides data from the HMS LINCS DB in a format that can easily be parsed by software, or from the command line, using the appropriate tools. Information and examples for this are given in the HMS LINCS Database URL Scheme and Access Guide.
- The programmatic API is used for directly communicating with the University of Miami LIFE system using the SAF data interchange formats.
- For this tutorial, we will access the API using the URL's given in a browser window, and we will specify that the output be displayed in a browser friendly JSON output.

1. Entity listing:
    a. List Cells: http://lincs.hms.harvard.edu/db/api/v1/cell/?format=json&offset=10&limit=30

    b. Note, we have set and “offset” and and “limit” to restrict the size of the returned data.
    c. List Small Molecules: http://lincs.hms.harvard.edu/db/api/v1/smallmolecule/?format=json&offset=10&limit=30
2. Dataset:
    a. Detail: http://lincs.hms.harvard.edu/db/api/v1/dataset/20003/?format=json
    b. Experimental result data (SAF format -- used for data interchange with the University of Miami LIFE system): http://lincs.hms.harvard.edu/db/api/v1/datasetdata/20003/?format=csv
3. Display a single entity by HMSL ID:
    a. Cell: http://lincs.hms.harvard.edu/db/api/v1/cell/50008/?format=json
    b. Small Molecule: http://lincs.hms.harvard.edu/db/api/v1/smallmolecule/10009-101/?format=json