Big Data Science with the BD2K-LINCS Data Coordination and Integration Center

About this course: The Library of Integrative Network-based Cellular Signatures (LINCS) is an NIH Common Fund program. The idea is to perturb different types of human cells with many different types of perturbations such as: drugs and other small molecules; genetic manipulations such as knockdown or overexpression of single genes; manipulation of the extracellular microenvironment conditions, for example, growing cells on different surfaces, and more. These perturbations are applied to various types of human cells including induced pluripotent stem cells from patients, differentiated into various lineages such as neurons or cardiomyocytes. Then, to better understand the molecular networks that are affected by these perturbations, changes in level of many different variables are measured including: mRNAs, proteins, and metabolites, as well as cellular phenotypic changes such as changes in cell morphology. The BD2K-LINCS Data Coordination and Integration Center (DCIC) is commissioned to organize, analyze, visualize and integrate this data with other publicly available relevant resources. In this course we briefly introduce the DCIC and the various Centers that collect data for LINCS. We then cover metadata and how metadata is linked to ontologies. We then present data processing and normalization methods to clean and harmonize LINCS data. This follow discussions about how data is served as RESTful APIs. Most importantly, the course covers computational methods including: data clustering, gene-set enrichment analysis, interactive data visualization, and supervised learning. Finally, we introduce crowdsourcing/citizen-science projects where students can work together in teams to extract expression signatures from public databases and then query such collections of signatures against LINCS data for predicting small molecules as potential therapeutics.

Who is this class for: Learn various methods of analysis including: unsupervised clustering, gene-set enrichment analysis, interactive data visualization, and supervised machine learning with application to data from the Library of Integrated Network-based Cellular Signature (LINCS) program, and other relevant Big Data from high content molecular omics data and phenotype profiling of mammalian cells.

Created by: Icahn School of Medicine at Mount Sinai

<table>
<thead>
<tr>
<th>Level</th>
<th>Intermediate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Commitment</td>
<td>4-5 hours/week</td>
</tr>
<tr>
<td>Language</td>
<td>English</td>
</tr>
<tr>
<td>How To Pass</td>
<td>Pass all graded assignments to complete the course.</td>
</tr>
<tr>
<td>User Ratings</td>
<td>5.0 stars</td>
</tr>
</tbody>
</table>

Average User Rating 5.0 See what learners said

Syllabus

WEEK 1

The Library of Integrated Network-based Cellular Signatures (LINCS) Program Overview

This module provides an overview of the concept behind the LINCS program; and tutorials on how to get started with using the LINCS L1000 dataset.
8 videos, 2 readings

Metadata and Ontologies

This module includes a broad high level description of the concepts behind metadata and ontologies and how these are applied to LINCS datasets.

2 videos

Serving Data with APIs

In this module we explain the concept of accessing data through an application programming interface (API).

2 videos

WEEK 2

Bioinformatics Pipelines

This module describes the important concept of a Bioinformatics pipeline.

1 video

The Harmonizome

This module describes a project that integrates many resources that contain knowledge about genes and proteins. The project is called the Harmonizome, and it is implemented as a web-server application available at: http://amp.pharm.mssm.edu/Harmonizome/

4 videos

WEEK 3

Data Normalization

This module describes the mathematical concepts behind data normalization.

2 videos

Data Clustering

This module describes the mathematical concepts behind data clustering, or in other words unsupervised learning - the identification of patterns within data without considering the labels associated with the data.

3 videos

Midterm Exam

The Midterm Exam consists of 45 multiple choice questions which covers modules 1-7. Some of the questions may require you to perform some analysis with the methods you learned throughout the course on new datasets.

**Graded:** Midterm Exam

WEEK 4

Enrichment Analysis
This module introduces the important concept of performing gene set enrichment analyses. Enrichment analysis is the process of querying gene sets from genomics and proteomics studies against annotated gene sets collected from prior biological knowledge.

3 videos

Machine Learning

This module describes the mathematical concepts of supervised machine learning, the process of making predictions from examples that associate observations/features/attribute with one or more properties that we wish to learn/predict.

3 videos

WEEK 5

Benchmarking

This module discusses how Bioinformatics pipelines can be compared and evaluated.

2 videos

Interactive Data Visualization

This module provides programming examples on how to get started with creating interactive web-based data visualization elements/figures.

4 videos

WEEK 6

Crowdsourcing Projects

This final module describes opportunities to work on LINCS related projects that go beyond the course.

2 videos, 1 reading

WEEK 7

Final Exam

The Final Exam consists of 60 multiple choice questions which covers all of the modules of the course. Some of the questions may require you to perform some analysis with the methods you learned throughout the course on new datasets.

**Graded:** Final Exam

FAQs

How It Works

Coursework
Each course is like an interactive textbook, featuring pre-recorded videos, quizzes and projects.

Help from Your Peers

Connect with thousands of other learners and debate ideas, discuss course material, and get help mastering concepts.

Certificates

Earn official recognition for your work, and share your success with friends, colleagues, and employers.

Creators

Icahn School of Medicine at Mount Sinai

The Icahn School of Medicine at Mount Sinai, in New York City is a leader in medical and scientific training and education, biomedical research and patient care.

Pricing

<table>
<thead>
<tr>
<th></th>
<th>Audit</th>
<th>Purchase Course</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access to course materials</td>
<td>Available</td>
<td>Available</td>
</tr>
<tr>
<td>Access to graded materials</td>
<td>-</td>
<td>Available</td>
</tr>
<tr>
<td></td>
<td>Not available</td>
<td></td>
</tr>
<tr>
<td>Receive a final grade</td>
<td>-</td>
<td>Available</td>
</tr>
<tr>
<td></td>
<td>Not available</td>
<td></td>
</tr>
<tr>
<td>Earn a shareable Course Certificate</td>
<td>-</td>
<td>Available</td>
</tr>
<tr>
<td></td>
<td>Not available</td>
<td></td>
</tr>
</tbody>
</table>

Ratings and Reviews

Rated 5 out of 5 of 5 ratings

Share

Tweet

Email

You May Also Like