Clinical Data for Translational Science Research; Bioinformatics and High-Density Data, Session 3

These online video lectures describe the increasingly rich sources of health data, such as electronic health records, available for scientific investigation by clinical and translational researchers, and introduces the methods and issues surrounding their use. The lecturer also reviews bioinformatic tools and information technologies that leverage high-density experimental data, with a focus on analysis of the human genome.

**About this Resource**

Computer databases and information systems have transformed virtually all aspects of health care and clinical research in the U.S. and internationally, impact that will only continue to grow in significance and breadth. Understanding these resources -- what they are, how they are structured, and how to use them -- is critical for clinical/translational researchers who wish to unlock their full potential for study.

In Part 1 of the videos accompanying the short course session, "Clinical Data for Translational Science Research," the ECDE lecturer introduces several of the most important health databases, including electronic health/medical records produced by hospitals and academic health care systems, health care payor billing and administrative systems, tumor and other registries, among others. The lecturer discusses issues involved in accessing, analyzing and aggregating information from these databases for research, and discusses the various ethical concerns and regulations surrounding their use, such as de-identification of data and patient confidentiality.

In Part 2 of the video, "Bioinformatics and High-Density Data," the ECDE lecturer discusses applications of bioinformatics and research involving extremely large quantities of "high-density" data, to address problems in biology, particularly molecular biology. The lecturer describes the rise of bioinformatics, with DNA sequencing in the 1960s, and its sharp growth during recent decades into a multidisciplinary field comprising numerous computational methods and tools that apply computer and information technologies to store and use data.

**Jeanne’s comments below are fine with me.**

**Course Syllabus/Topics**

*Clinical Data for Translational Science Research*

Sources of Clinical Data

Coded Data Vocabularies and Ontologies

LOINC Examples

Data Collection and Chart Abstraction

*Bioinformatics and High-density Data*

What is Bioinformatics?

Overall Strategy in High-Density Data Analysis

Two Step Selection Procedure

Reproducibility of Data Sets

Pearson Correlation Test

Panther Gene List Analysis Tool Suite

**Recommended background**

Please complete the following prior to attending the short course, "Clinical Data for Translational Science Research; Bioinformatics and High-Density Data, Session 3."

Pre-test on Clinical Translational Research

View the two-part video resource

Participant Break-out Session Guide

Read the three articles listed below

**Suggested Readings**

1. [Hude Quan](http://www.ncbi.nlm.nih.gov/pubmed/?term=Quan%20H%5Bauth%5D" \t "_blank), [Bing Li](http://www.ncbi.nlm.nih.gov/pubmed/?term=Li%20B%5Bauth%5D" \t "_blank), [L Duncan Saunders](http://www.ncbi.nlm.nih.gov/pubmed/?term=Duncan%20Saunders%20L%5Bauth%5D" \t "_blank), [Gerry A Parsons](http://www.ncbi.nlm.nih.gov/pubmed/?term=Parsons%20GA%5Bauth%5D" \t "_blank), [Carolyn I Nilsson](http://www.ncbi.nlm.nih.gov/pubmed/?term=Nilsson%20CI%5Bauth%5D" \t "_blank), [Arif Alibhai](http://www.ncbi.nlm.nih.gov/pubmed/?term=Alibhai%20A%5Bauth%5D" \t "_blank), and [William A Ghali](http://www.ncbi.nlm.nih.gov/pubmed/?term=Ghali%20WA%5Bauth%5D" \t "_blank), for the IMECCHI Investigators. Assessing Validity of ICD-9-CM and ICD-10 Administrative Data in Recording Clinical Conditions in a Unique Dually Coded Database. Health Serv Res. Aug 2008; 43(4):1424-1441.
2. Mi H, Muruganujan A, Casagrande JT, Thomas PD. [Large-scale gene function analysis with the PANTHER classification system.](http://www.ncbi.nlm.nih.gov/pubmed/23868073) Nat Protoc. 2013 Aug;8(8):1551-66.

# Hutchins JR. What's that gene (or protein)? Online resources for exploring functions of genes, transcripts, and proteins. Mol Biol Cell. 2014 Apr;25(8):1187-201.

**In-class Exercise**

1. Clinical Data-Leveraging Structured Data Summary
2. Prepare a study question that will require structured data

* Define what data sources you will access to obtain your structured data
* Be specific about the terminologies in question
* What are the limitations on the data?
* What are some areas in which the data will be “non-standardized?”

**FAQ**

**Will I receive a Certificate of Completion after completing this course?** Yes. Participants who complete the course receive a Certificate. Participants must watch the videos and attend the in-person session.

**Do I need to be a K Scholar to take this course?** No. In addition to K Scholars, this course is open to faculty, clinicians, community health workers, fellows, post-docs, TL1 and F trainees, as well as medical, OT/PT, pharmacy and other students who intend to conduct clinical and translational research.