CS 466
Introduction to Bioinformatics

Instructor: Jian Peng
Teaching Assistant: Baqiao Liu & Shayan Tabe Bordbar

Introduction

Instructor:

- Jian Peng
  Office hour: Mon, 3:00pm-4:00pm
  Zoom link: Same as the class link
  Email: ***jianpeng@illinois.edu***
- My own research:
  Computational Biology and Machine Learning

Teaching Assistants:

- Baqiao Liu, PhD student
  Office hour: TBD
  Email: ***baqiaol2@illinois.edu***
- Shayan Tabe Bordbar, PhD student
  Office hour: TBD
  Email: ***tabebor2@illinois.edu***

Prerequisites

- Programming skills (equivalent to CS 225) for doing the mini-project.
- Knowledge of basic probability and statistics for understanding several lectures.
- No biology background is necessary.

Course logistics

- Course website:
  https://courses.engr.illinois.edu/cs466/sp2021/

- Piazza website:
  https://piazza.com/illinois/spring2021/cs466/home

- Lecture slides will be released before each class.
- Participation is encouraged.
- Come to class having read the day's lecture slides and reading assignments, if any.

Course Objectives

Introduction to bioinformatics

- Basic problems in computational biology
- Statistics and machine learning for data analysis
- Algorithms for data processing
- Advanced applications to biology

Assignments

- See the University Policy on Academic Integrity, especially the section on plagiarism.
- Late submission within 3 days (72 hours) is worth 80% credit.
- A student may request an extension of 3 days at most once in the semester.

Grading

- Five problem sets (30%)
- Midterm (30%)
- Final (40%)

Approximate data from a recent offering:

- Enrollment (who completed course): 43
- 27 A grades (2 A+, 23 A, 2 A-)
- 16 B grades (10 B+, 6 B)

This is not a statement about what the distribution this semester will be.

Questions about the course logistics?

Introduce yourself

Bioinformatics

- Is not about one problem (e.g., designing better computer chips, better compilers, better graphics, better networks, better operating systems, etc.)

- Is about a family of very different problems, all related to biology, all related to each other

- How can computers help solve any of this family of problems ?

Bioinformatics and You

- You can learn the tools of bioinformatics
- These tools owe their origin to computer science, information theory, probability theory, statistics, etc.
- You can learn the language of biology, enough to understand what the problems are
- You can apply the tools to these problems and contribute to science

Important Biological Questions?

"Why do humans have so few genes?"

"Can we understand DNA code?"

"Can we understand gene function?"

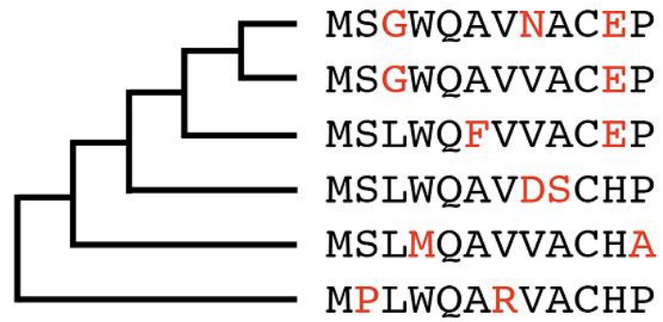
"How did cooperative behavior evolve?"

"Can we cure cancer?"

……

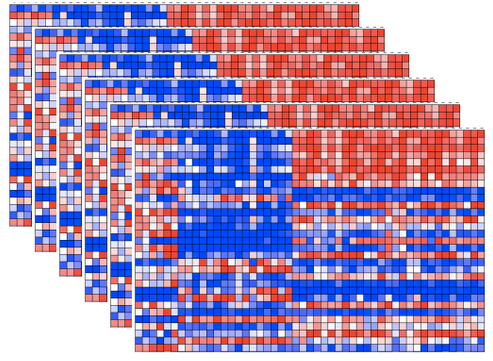
What does biological data look like?

Sequence data



- Protein/DNA sequence
- Probabilistic models for sequences
- Dynamic programming

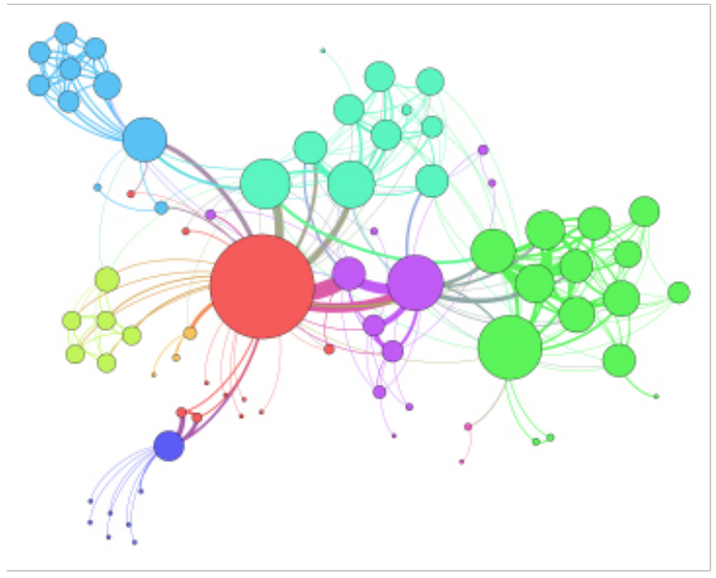
Matrix data



- Gene expression
- Dimensionality reduction and feature selection
- PCA and clustering

Biological Data

Graph data



- Molecular interaction networks
- Graph algorithms

Heterogeneous data

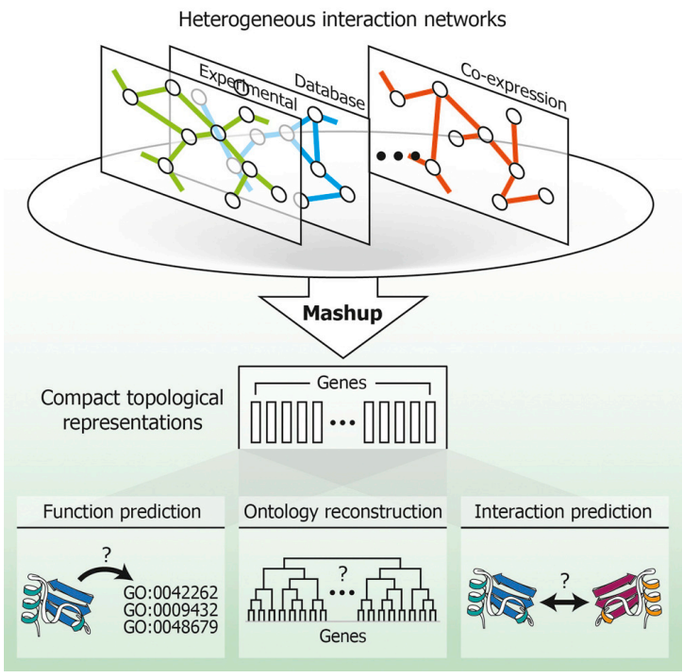
- Dimensionality reduction
- Probabilistic models for data integration
- Network-based data integration

TODO after this class

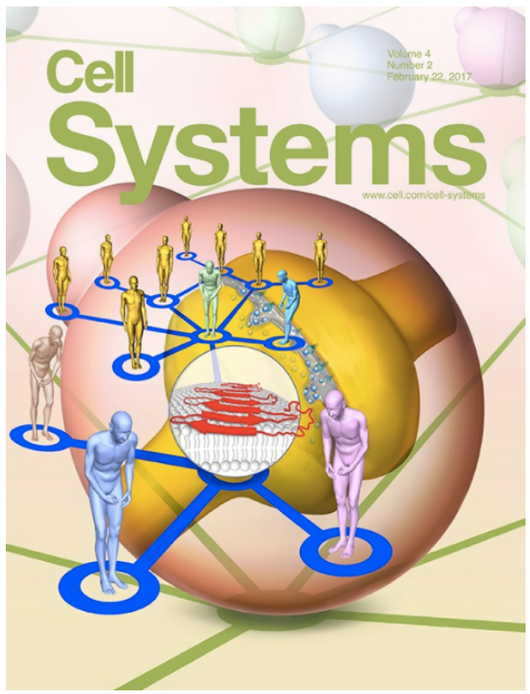
Please read “Molecular Biology for Computer Scientists” by Lawrence Hunter

Examples of my research projects

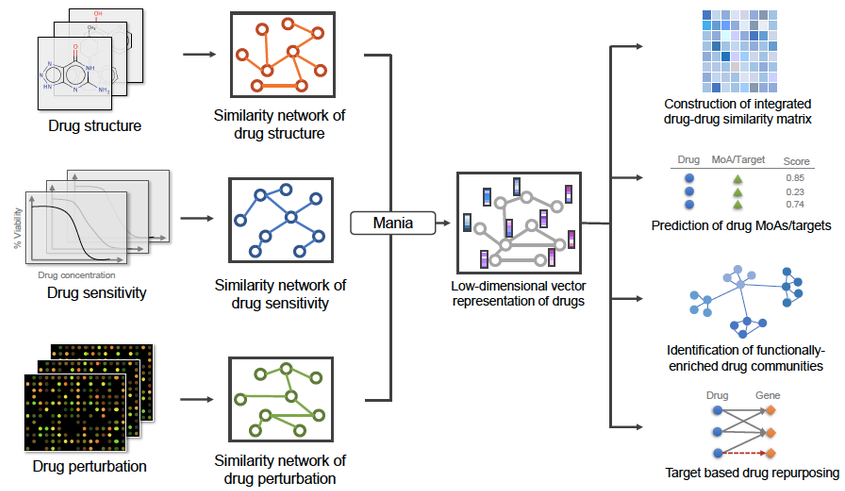
Recent research



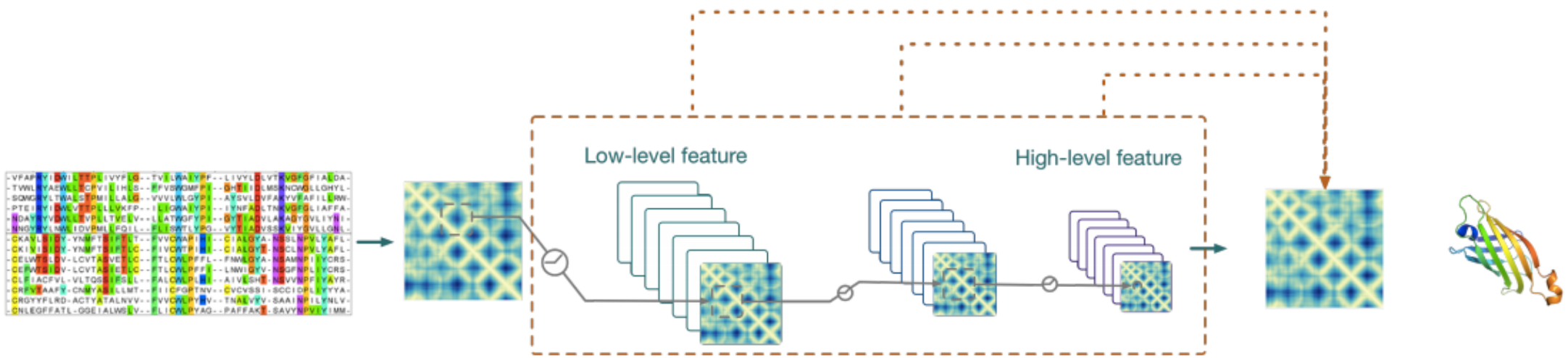
Cell Systems, 2016



Cell Systems, 2017



Nature Communications, 2017



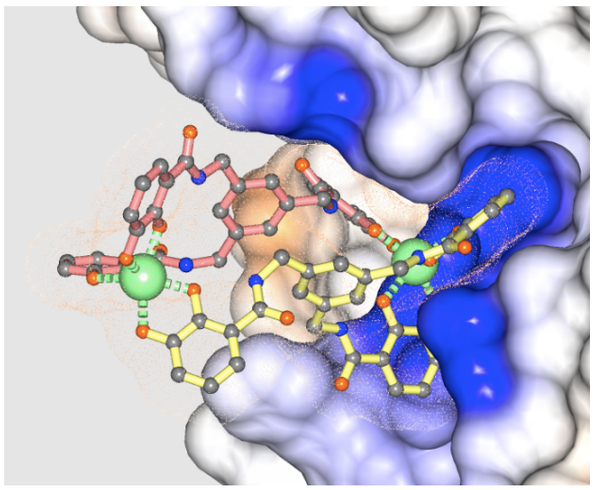
Cell Systems, 2018

Protein sequence, structure and function

sequence

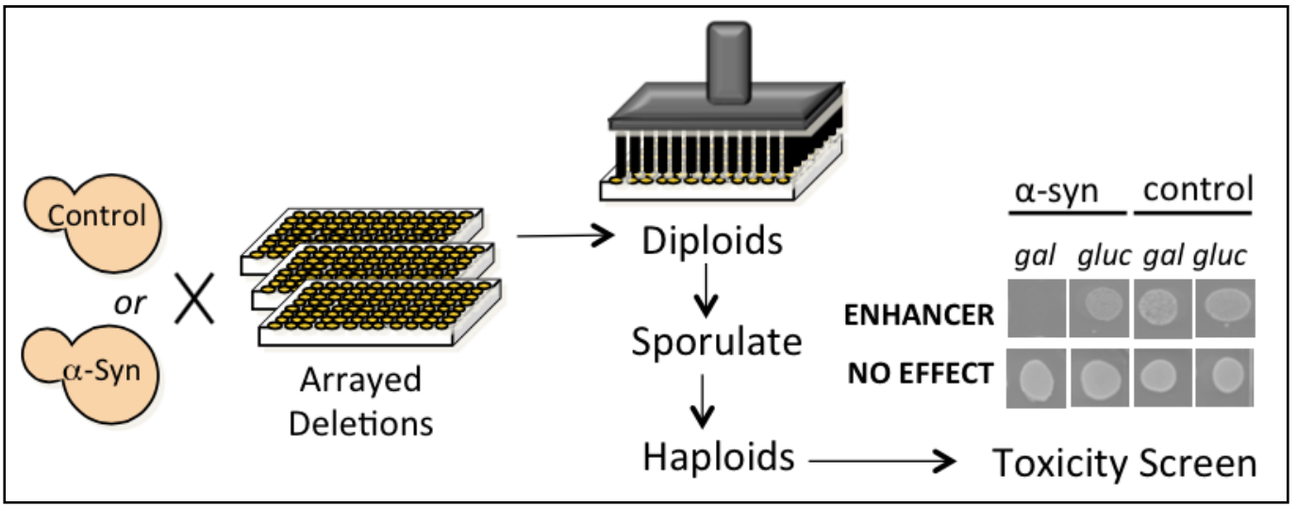
```
ACDEEEFGHIKL----MPQRSTVWY
ACDE--FGHIKLRMQP----STVWY
```

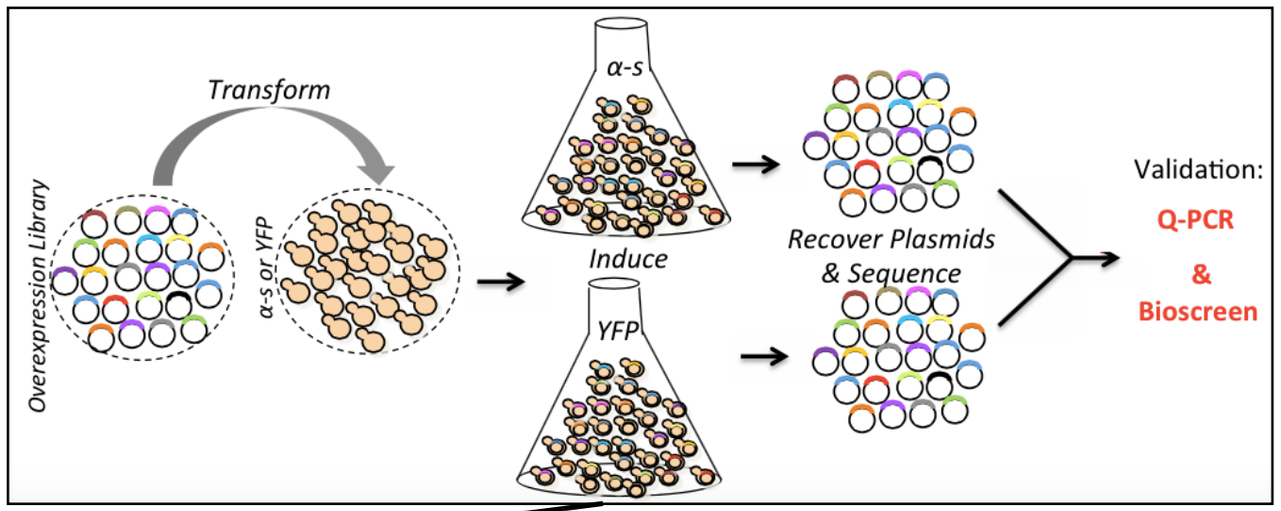
structure



function

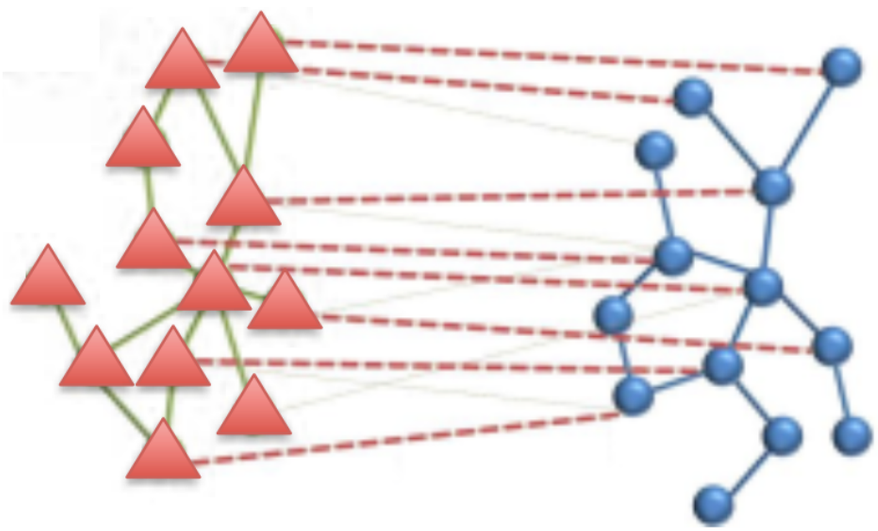
Network analysis for disease modeling





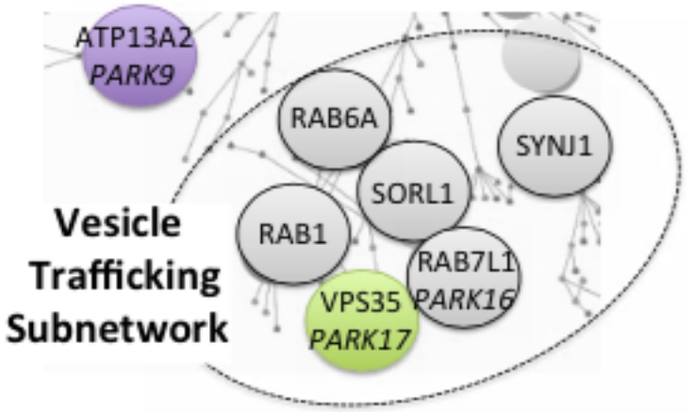
Yeast Screen Hits

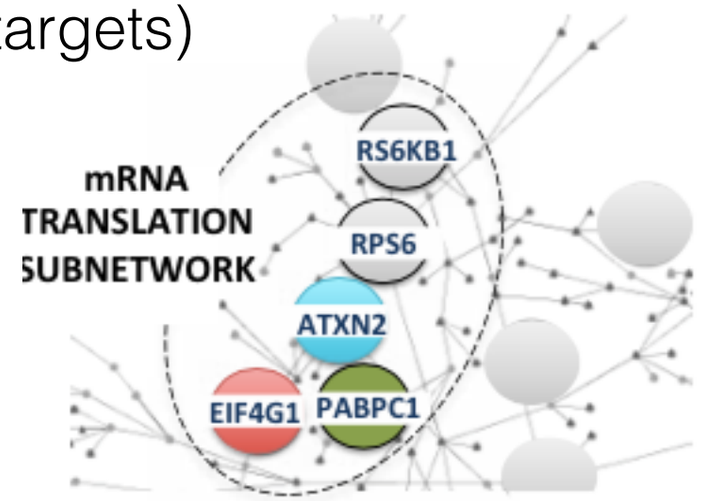
network analysis



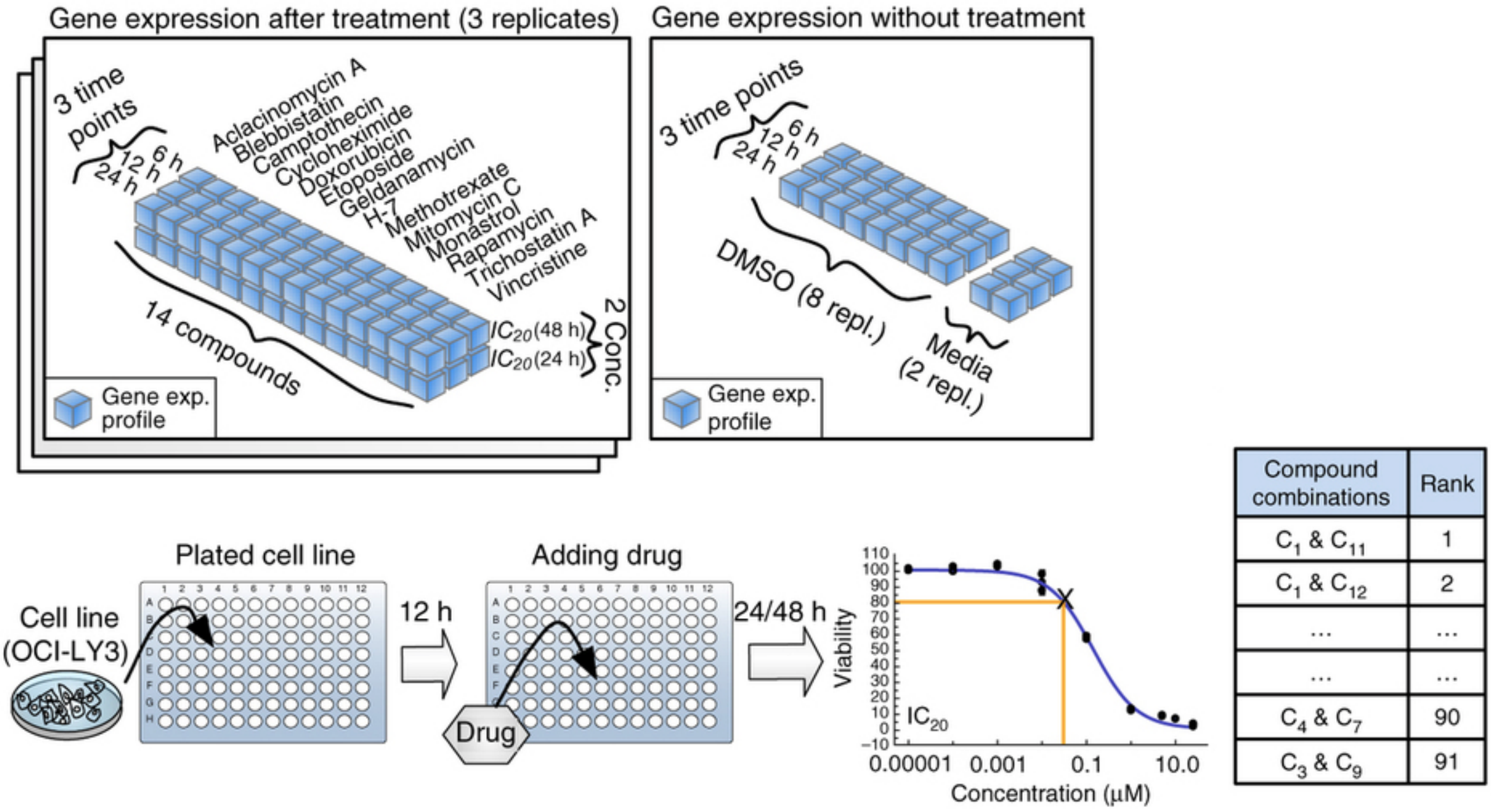
human disease network

new disease biology
(potential drug targets)





Pharmacogenomics and cancer genomics



Compound combinations	Rank
C_1 & C_{11}	1
C_1 & C_{12}	2
…	…
…	…
C_4 & C_7	90
C_3 & C_9	91

Figure from the DREAM challenge website