Information-based integrative drug discovery

Jeremy Yang

Associate Instructor
IU School of Informatics and Computing

Instructor: Prof. Joanne Luciano

Based on slides by Prof. David Wild, 2013 edition of course.

DATA SCIENCE FOR DRUG DISCOVERY, HEALTH AND TRANSLATIONAL MEDICINE (DSDHT) INFO I-590

Drug discovery paradigms

- **Empirical**
  - ~200k BCE+
- **Vitamins, vaccines & antibiotics**
  - ~1800+
- **Rational**
  - ~1960+
- **Integrative**
  - ~2010+

New paradigms for disease treatment

- **Complex biological networks**: beyond "lock and key" to understand complex disease states (multi-morbidity), multiple involvement of targets and pathways, and drug treatment including side effects (polypharmacology).
- **Drug repurposing**: finding new applications of safe, well-established drugs (including those available as cheap generics).
- **Personalized/precision medicine**: use of possibly multiple drugs in a targeted fashion that suits individual phenotypes and genotypes and avoids potential rare but serious side-effects

Informatics disciplines related to drug discovery

- **Bioinformatics** – the representation and use of biomolecular information (proteins, genes, biological pathways)
- **Cheminformatics** – the representation and use of chemical information (compounds and drugs and their interaction with proteins)
- **Biomedical informatics** – the representation and use of clinical information
- Computation, data and network driven
- We're now seeing integrative terms like **systems chemical biology** and **translational medicine** that are mapping to informatics disciplines.

Information overload

- There is now an incredibly big problem / rich resource of public information relating compounds, targets, genes, pathways, and diseases. E.g.:
    - 13,664 papers on PPAR-gamma (PubMed)
    - 93M compounds and 1.2M bioassays (PubChem)
    - 2.3M compounds tested in bioassays (PubChem)
    - 140M sequences (UniParc), 550k proteins (UniProt/SwissProt)
    - 125k protein 3D structures (PDB)
    - 198M human nucleotide sequences (GenBank)
    - 26M life science publications, +1M/year (PubMed)
    - Multitude of other sets (drugs, toxicogenomics, chemogenomics, metagenomics …)

Data obtained Dec 2016

Take home message

- Effective drug discovery now requires understanding of the complexeties of disease, and the complex ways drugs interact system-wide in the body.

- To do this requires informatics to be at the core of the drug discovery process – one person or group can't amass all the knowledge needed to make good decisions.

- Computational disciplines such as bioinformatics, cheminformatics, and biomedical informatics have emerged over the last few decades. These provide a toolbox for integrative approaches.