Hypothesis Fusion to Improve the Odds of Successful Drug Repurposing

Alexander Tropsha, Charles Schmitt, Eugene Muratov
UNC-Chapel Hill

Weifan Zheng, NCCU

Nabarun Dasgupta, Epidemico
A second concern held by some is that a new class of research person will emerge — people who had nothing to do with the design and execution of the study but use another group’s data for their own ends, possibly stealing from the research productivity planned by the data gatherers, or even use the data to try to disprove what the original investigators had posited. There is concern among some front-line researchers that the system will be taken over by what some researchers have characterized as “research parasites.”
Information resources for bioactive chemicals are abundant and growing.

- PubMed: Over 18 million citations from MEDLINE and other life science journals for biomedical articles back to the 1950s.
- FDA Orange Book of Approved Drug Products.
- Electronic Orange Book.
- FDA approved labels for marketed drugs.
- Compound Assay data for proteins and cytotoxicity.
- DrugBank: The DrugBank database combines detailed drug data (8200+ drug entries) with comprehensive drug target information.
- FDA Safety Information and Adverse Event Reporting Program.
- Human Liver Adverse Effects Database.
- “Potential Safety Issue” data
  - “Drug Interactions” table
  - “Drug Interactions” table
  - FDA approved labels for marketed drugs.
  - FDA Orange Book of Approved Drug Products.
  - Electronic Orange Book.
  - Compound Assay data for proteins and cytotoxicity.
  - DrugBank: The DrugBank database combines detailed drug data (8200+ drug entries) with comprehensive drug target information.
  - FDA Safety Information and Adverse Event Reporting Program.
  - Human Liver Adverse Effects Database.

Modified from a slide provided by Julie Barnes, Biowisdom.
**Data Science and Data Cycle**

- **Data Analysis and Modeling**
  - Structured Data Repository
  - Electronic Databases:
    - Lab collections
    - Literature data
  - Data Analysis
  - Text Mining
  - Processing
  - Literature data

- **Data Reproducibility and Data Curation** are critical, otherwise:
  - BD2K = Bogus
  - Data to Knonsense

- **Unstructured test:**
  - Facebook
  - Twitter
  - Other Social Media

- **Experimental Validation**
- **Experimental Design**
- **Predictive data models & tools**
- **Decision support**
Data set curation workflows: Trust but Verify!

Fourches D. et al.  

Fourches D. et al.  
Disease-gene signatures

Disease related genes or proteins

Text/database mining

Network mining

Functional data

Binding data

QSAR

Predictive models

Database mining

Target related ligands

Disease-Target Association

New hypothesis about connectivity between chemicals and diseases

Structural hypothesis “putative drug candidates”

QSAR modeling and Virtual Screening: Hit identification in external libraries

- CHEMICAL STRUCTURES → CHEMICAL DESCRIPTORS → PREDICTIVE QSAR MODELS → PROPERTY/ACTIVITY

- CHEMICAL DATABASE

- VIRTUAL SCREENING

- HITS (confirmed actives)

- INACTIVES (confirmed inactives)

~10^6 – 10^9 molecules
5-HT$_6$ receptor QSAR models & QSAR-based VS

Dataset

196 cps.

94 Inactives $K_i \geq 10$ µM
102 Actives $K_i < 10$ µM

Model statistics

Virtual screening

300 VS Hits “Actives”
The connectivity map

Step 1: upload signature
Step 2: query the cmap
Step 3: list of correlated compounds

Querying the cmap

Upload signature

Query the cmap

List of compounds

Alzheimer’s disease gene signatures
97 COMMON HITS with S1
106 COMMON HITS with S2

73 COMMON HITS with S1 & S2

Further selection

34 Higher Confidence Hits

- Antipsychotics
- Antidepressants
- Calcium Channel Blockers
- Selective Estrogen Receptor Modulators (SERMs)
Exploring PubMed as one of the largest Chemical Biology Databases: the ChemoText Project

- 2008 Medline baseline: 16,880,015 records
- 6,635,344 records had subject chemicals

- 134,184 distinct Subject Chemicals
- 61,329 distinct Proteins
- 7,761 distinct Drug Effects
- 4,865 distinct Diseases

- 9,088,747 relationships
- 13,157,701 relationships
- 20,466,335 relationships
- 9,360,330 relationships

http://chemotext.mml.unc.edu/
Swanson’s ABC approach to drug discovery via text mining*

ABC Method as applied to discern chemical-target-disease associations (using Chemotext)

![Diagram of ABC Method]

http://chemotext.mml.unc.edu/
Raloxifene identified as a 5-HT$_6$ receptor ligand and potential treatment for the Alzheimer’s disease

- Raloxifene binds to 5-HT$_6$ receptor with a $K_i = 750$ nM.*


---


**Competition binding at 5-HT$_6$ receptors for raloxifene (yellow triangle) and chlorpromazine (square) versus [3H] LSD. Tested by our collaborators at PDSP.**
Aim 1: From Man

- Social Media (e.g., twitter, facebook, patientslike.me, Etc.)
- Data Integration
- Data Curation
- Curated Database of Assertions
- Analysis
- Cancer-Related Assertions

Hypothesis generation

Aim 2: To Molecule

- On-line Databases (e.g., CHEMBL, PDB, PubChem, Etc.)
- Curated Cancer-Related Bioassay Database
- Virtual screening platform
- OSAR
- Primary Hits

Hypothesis confirmation

Aim 3: To Man

- Experimental validation in-vitro and in-vivo
- Drug-Target-Disease Database
- Candidates for Repurposing
- Disease ↔ Effect

Hypothesis enrichment

NIH 1U01CA207160-01. Drug repurposing: From Man to Molecules to Man