Assessing drug-target association using semantic linked data

"Information is cheap. Understanding is expensive" (Karl Fast)

David Wild, Bin Chen and Ying Ding
http://djwild.info

David Wild, Assistant Professor and Director,
Cheminformatics & Chemogenomics Research Group (CCRG)
Indiana University School of Informatics and Computing
djwild@indiana.edu

The question

Can we make accurate (*useful)*
drug-target association predictions using a wide range of public data?

Chemogenomics informed by systems chemical biology

Alternative to multiple QSAR / target based / similarity

The methods

Semantic Technologies

- Heterogeneous networks
- Semantic integration
- Prediction in its infancy

Networks & Graphs

- Homogeneous networks
- Association prediction and data mining

QSAR

- Scope defined by +/- data avalable
- Decent binding predictions in scope

Systems chemical biology + Semantic Web

Reviews • INFORMATICS

Systems chemical biology and the Semantic Web: what they mean for the future of drug discovery research

David J. Wild[1], Ying Ding[2], Amit P. Sheth[3], Lee Harland[4], Eric M. Gifford[5] and Michael S. Lajiness[6]

[1] Indiana University School of Informatics and Computing, Bloomington, IN, USA
[2] Indiana University School of Library and Information Science, Bloomington, IN, USA
[3] Wright State University Department of Computer Science and Engineering, Dayton, OH, USA
[4] ConnectedDiscovery, Deal, Kent, UK
[5] Pfizer Global Research and Development, Groton, CT, USA
[6] Eli Lilly, Indianapolis, IN, USA

Systems chemical biology, the integration of chemistry, biology and computation to generate understanding about the way small molecules affect biological systems as a whole, as well as related fields such as chemogenomics, are central to emerging new paradigms of drug discovery such as drug repurposing and personalized medicine. Recent Semantic Web technologies such as RDF and SPARQL are technical enablers of systems chemical biology, facilitating the deployment of advanced algorithms for searching and mining large integrated datasets. In this paper, we aim to demonstrate how these technologies together can change the way that drug discovery is accomplished.

Introduction

Traditionally, drug discovery paradigms involve identifying a protein target that is implicated in disease processes, and then identifying one or more chemical compounds that can safely

off-target effects of drugs using network methods [2,3]; repurposing of known drugs for new targets [4]; drug–target interaction networks for exploring the kinome [5]; mapping assay networks onto biological networks to relate compounds and

Drug Discovery Today, 2012, in press.

Semantic Technologies: an enabler for integration

- Allows flexible description of **heterogeneous** graphs of data relationships (RDF), following the rules of an ontology (OWL)
- "Semantic linking" refers to annotation of relationships with meaningful, organized labels (like node and edge labels on a graph) using an ontology
- "Semantic Web" received a bad press, as not mature enough to be practically useful until the last 2-3 years. Practicality radically changed by advent of:
  - Fast triple stores – e.g. Virtuoso – the equivalent of database management systems for RDF
  - SPARQL – a query language that allows expression of integrative queries in a simple(ish) fashion
  - Sensible use of OWL ontologies – small to medium scope, aligned with other ontologies through open resources such as NCBO BioPortal and OBO
- Currently main focus is on data integration and description. Lots of current work in this area (OpenPHACTS, CSHALS, SWHCLS, Pistoia Alliance) – topic of recent JChemInf special issue http://www.jcheminf.com/series/acsrdf2010

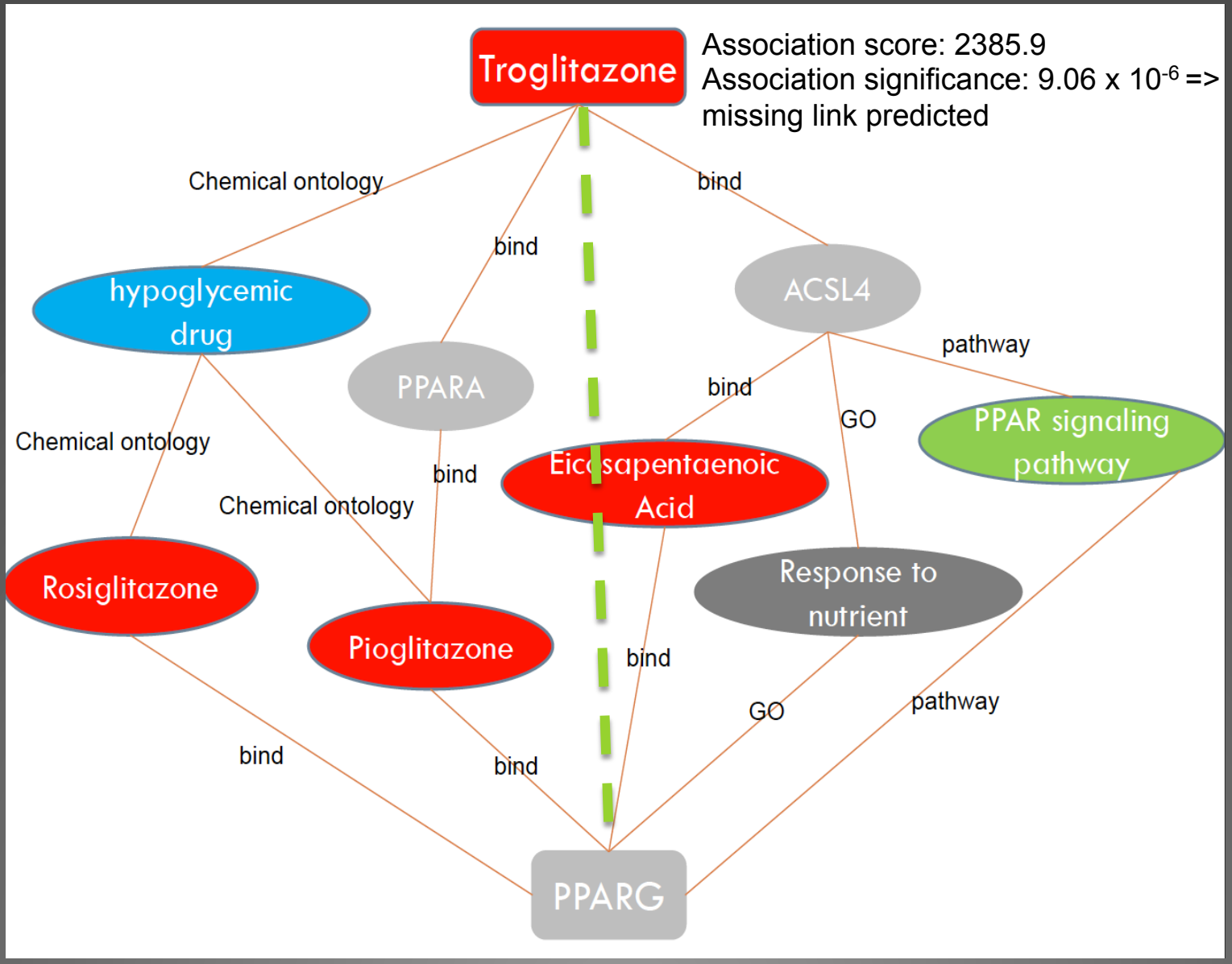
Chem2Bio2RDF – www.chem2bio2rdf.org

- Semantically integrates 42 heterogeneous public datasets related to drug discovery in a fast Virtuoso triple-store with SPARQL endpoint
- Datasets cover chemistry, chemogenomics, biology, systems & pathways, pharmacology, phenotypes, toxicology, glycomics and publications, and biological entities of compounds, drugs, targets, genes, pathways, diseases and side-effects
- Major datasets include PubChem, ChEMBL, DrugBank, PharmGKB, BindingDB, STITCH, CTD, KEGG, SWISSPROT, PDB, SIDER, PubMed. Full set at http://chem2bio2rdf.wikispaces.com/Datasets
- Holds data on ~31m chemical structures, ~5,000 marketed drugs, ~59m bioactivity data points and ~19m publications
- Linked into LOD cloud, and forms part of OpenPHACTS repository
- Data semantically annotated with Chem2Bio2OWL ontology (***Journal of Cheminformatics, 2012, 4:6***)
- For more information, see ***BMC Bioinformatics* 2010, 11, 255.**

Semantic Link Association Prediction (SLAP)

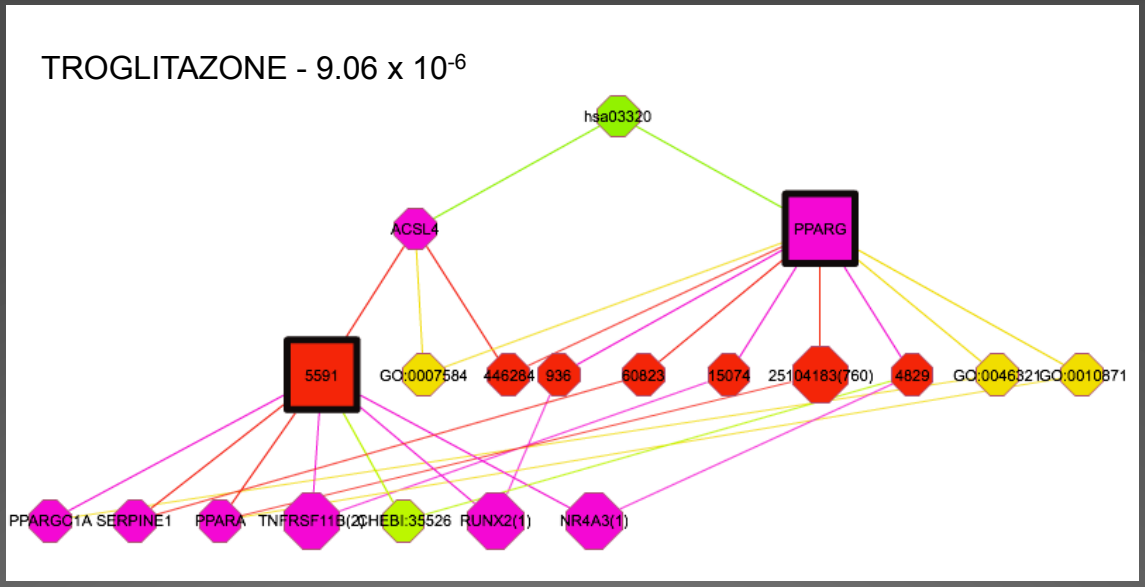
- Predicts a *probability of association* of a compound and a target based on the network paths between them that involve drugs, targets, pathways, diseases, tissues, GO terms, chemical ontologies, substructure and drug side-effects
- It can be primarily considered as a "missing link prediction"
- Data source is a subset of the Chem2Bio2RDF network including 250,000 compounds with known bioactivities and the targets known to associated with these drugs
- *Raw Score* is a measure of the significance of a single path between a compound and target, based on topology and semantics of the path nodes and edges. Raw scores are normally distributed within a *path pattern*
- *Association Score* is a sum of z-scores of raw scores relative to a distribution of random pair scores for different paths and path patterns. Association scores form a normal distribution
- *Association Significance* is a significance p-value of an association score based on the normal distribution of association scores.
- **PLoS Computational Biology – paper in review**

Example: Troglitazone and PPARG

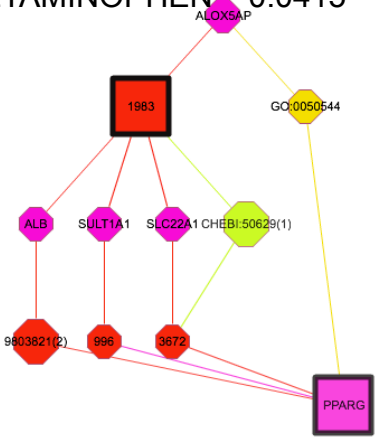


Significance of other drugs vs PPARG

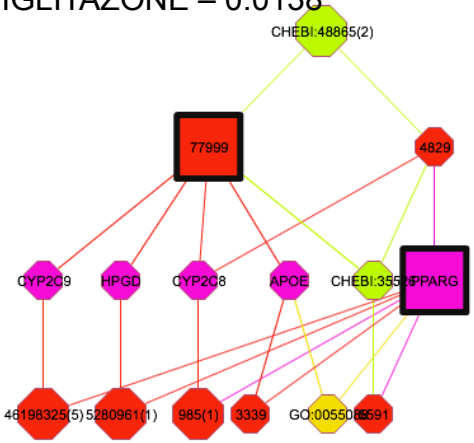
TROGLITAZONE - 9.06 x 10^{-6}



ACETAMINOPHEN – 0.0415



ROSIGLITAZONE – 0.0138



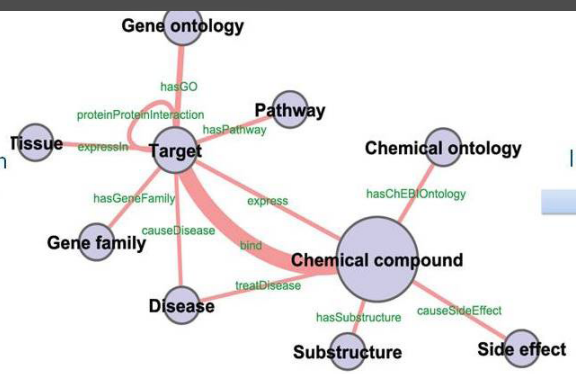
ZALEPLON– 0.4258

SLAP process

PubChem	ChEBI	DrugBank
UniProt	UniProtKB-GOA	HGNC
SIDER	OMIM	KEGG
HPRD	ChEMBL	TTD
BindingDB	CTD	PDSP

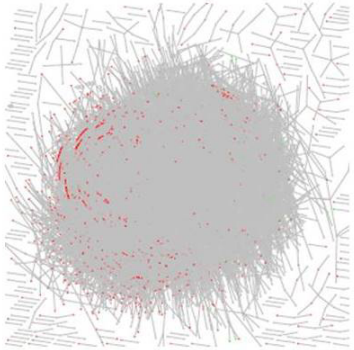
(a) Raw Data Sets

Semantic annotation



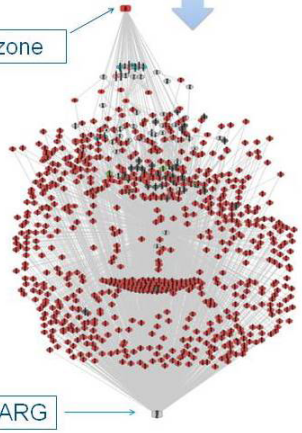
(b) Ontological level schema

Integration



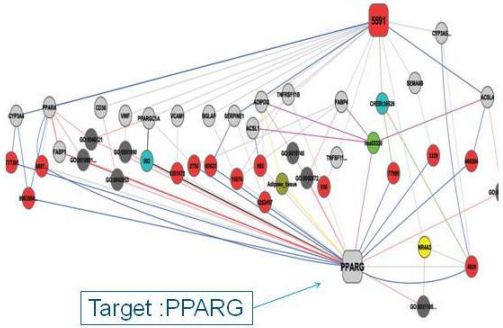
(c) Semantic Linked Data

Path finding



(d) Paths between two nodes

Path filtering



(e) Significant Paths between two nodes

ciation assessing

1. Edge weight:

$$p(e(i \to j)) = \frac{1}{\sum_{k}^{n=1} R_{i,n} == R_{i,j}}$$

2. Path score:

$$p(P_l(t \to s)) = p(P_l(e_{m \to m-1}, \ldots, e_{3 \to 2}, e_{2 \to 1})) = \prod_{i=1}^{m-1} e_{i+1 \to i}$$

$$log(p(P_l(t \to s))) = \sum_{i=1}^{m-1} log(e_{i+1 \to i})$$

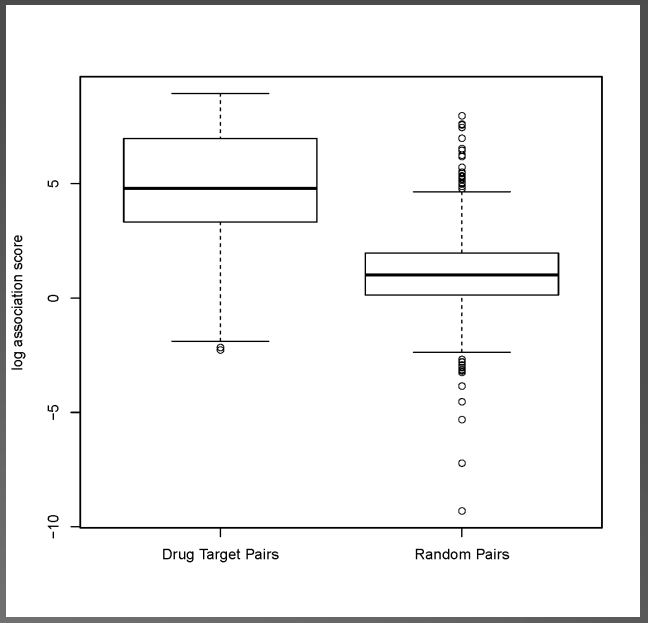
3. Association score

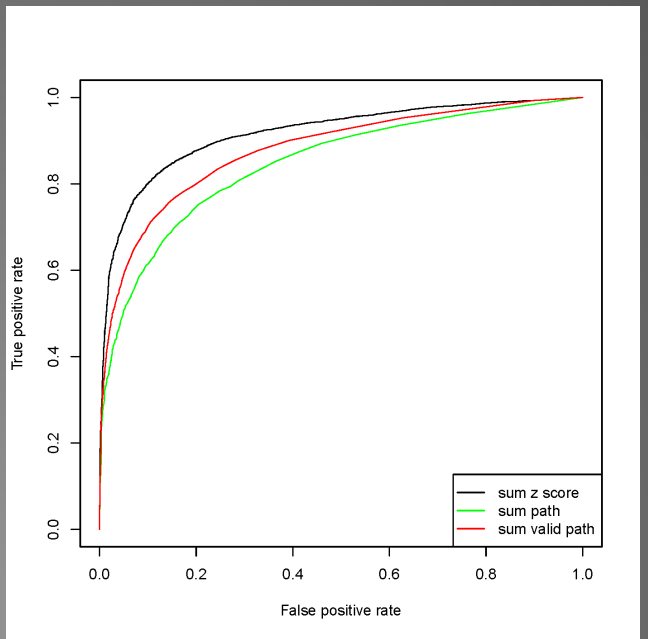
$$raw\ score(s,t) = \sum_{l}^{n} \frac{log(p(P_l)) - \theta(log(P_l))}{\sigma(log(P_l))}$$

(f) Statistical Models

Validation of SLAP with DrugBank and MATADOR

- Leave-one-out study on 1,000 known drug-target pairs from DrugBank compared with 1,000 randomly selected pairs of drugs and targets shows significantly higher association scores for known pairs ($p<2.2x10^{-16}$)
- ROC curve based on 4,508 pairs in DrugBank SLAP with other missing link prediction methods, shows SLAP outperforming other methods
- Leave out study on 1065 MATADOR pairings (444 not present in our source dataset) identified 560/621 and 170/444 pairs
- SLAP shows no preference for particular target classes





Cross-check with SEA

- SEA analysis (Nature 462, 175-181, 2009) predicts 184 new compound-target pairs, 30 of which were experimentally tested
- 23 of these pairs were experimentally validated (<15uM) including 15 aminergic GPCR targets and 8 which crossed major receptor classification boundaries
- 9 of the aminergic GPCR target pairings were correctly predicted by SLAP ($p<0.05$) – for the other 6 the compounds were not present in our set
- 1 of the 8 cross-boundary pairs was predicted

Table 2 | Prediction and testing of new cross-boundary targets for drugs

Drug	Canonical target	*E*-value	Predicted target	K_i (nM)
Xenazine	VMAT2 (transporter)	1.4×10^{-61}	α_2 adrenergic receptor (GPCR)	α_{2A}, 960; α_{2C}, 1.3×10^3
Rescriptor	HIV-1 reverse transcriptase (enzyme)	1.1×10^{-30}	Histamine H_4 receptor (GPCR)	5.3×10^3
Vadilex	NMDAR (ion channel)	5.1×10^{-13} 2.0×10^{-4}	μ-opioid receptor (GPCR) 5-HTT; serotonin transporter (transporter)	1.4×10^3 77
RO-25-6981	NMDAR (ion channel)	1.5×10^{-8} 1.9×10^{-6} 3.6×10^{-6} 9.1×10^{-5}	5-HTT; serotonin transporter (transporter) Dopamine D_4 receptor (GPCR) NET; noradrenaline transporter (transporter) κ-opioid receptor (GPCR)	1.4×10^3 120 1.3×10^3 3.1×10^3

K_i values are accurate ±20% at two significant figures. The MDDR database did not specify the α_2 adrenergic receptor subtype, requiring a separate assay for each one (α_{2A}, α_{2C}).

Table 1 | Prediction and testing of new aminergic GPCR targets for drugs

Drug	Pharmacological action	*E*-value	Predicted target	K_i (nM)
Sedalande	Neuroleptic	8.3×10^{-136} 1.7×10^{-14}	α_1 adrenergic blocker* 5-HT_{1D} antagonist	α_{1A}, 1.2; α_{1B}, 14; α_{1D}, 7.0 140
Dimetholizine	Antihistamine; antihypertensive	1.6×10^{-129} 2.7×10^{-113} 7.4×10^{-56}	α_1 adrenergic blocker* 5-HT_{1A} antagonist Dopamine D_2 antagonist	α_{1A}, 70; α_{1B}, 240; α_{1D}, 170 110 180
Kalgut	Cardiotonic	3.1×10^{-79}	β_3 adrenergic agonist	2.1×10^3
Fabahistin	Antihistamine	5.7×10^{-57}	5-HT_{5A} antagonist	130
Prantal	Anticholinergic; antispasmodic	5.5×10^{-32}	δ-opioid agonist	1.4×10^4
N,*N*-dimethyltryptamine	Serotonergic hallucinogen	3.1×10^{-21} 1.2×10^{-13} 1.1×10^{-7}	5-HT_{1B} agonist 5-HT_{2A} agonist† 5-HT_{5A} antagonist	130 130 2.1×10^3
Doralese	Adrenergic α_1 blocker; antihypertensive; antimigraine	5.0×10^{-6} 2.8×10^{-27}	5-HT_7 modulator Dopamine D_4 antagonist	210 18
Prozac	5-HT reuptake inhibitor; antidepressant	3.9×10^{-15}	β adrenergic blocker*	β_1, 4.4×10^3
Motilium	Antiemetic; peristaltic stimulant	4.8×10^{-11}	α_1 adrenergic blocker*	α_{1A}, 71; α_{1B}, 530; α_{1D}, 710
Paxil	5-HT reuptake inhibitor; antidepressant	1.3×10^{-7}	β adrenergic blocker*	β_1, 1.0×10^4

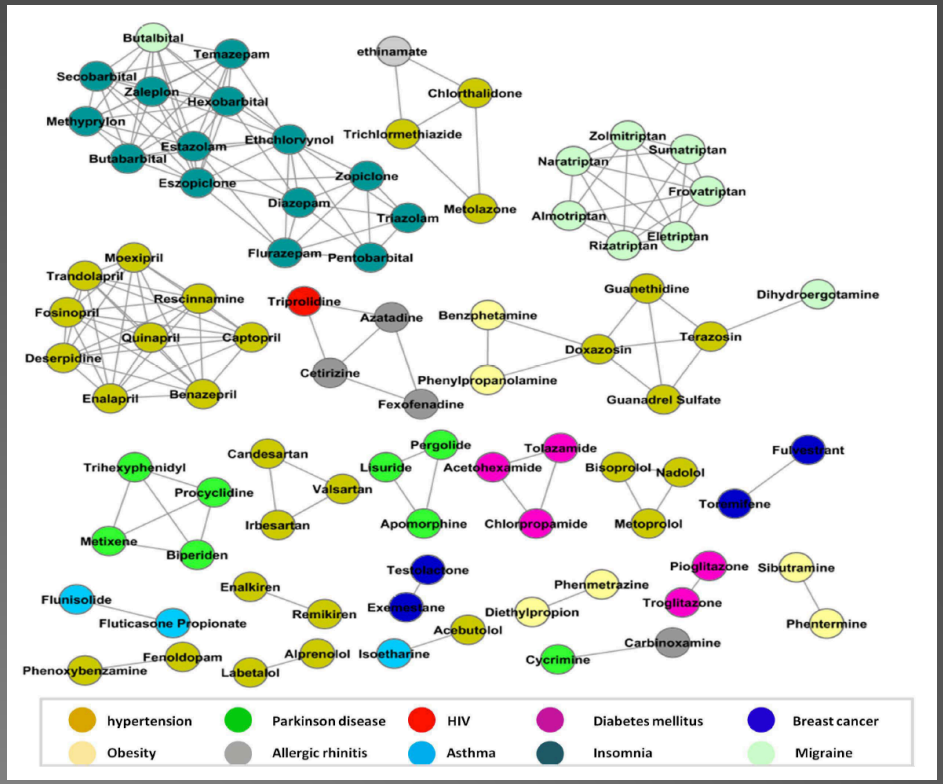
K_i values are accurate ±20% at two significant figures.
* For the targets marked, the reference data set did not specify the receptor subtype, requiring a separate assay for each one. For instance, the MDDR contains an 'α_1 adrenergic blocker' set, for which it was necessary to test the α_{1A}, α_{1B} and α_{1D} subtypes.
† 5-HT_{2A} is a known target of DMT, but is shown here with its retrospective SEA *E*-value for comparison purposes.

Comparison with Connectivity Map

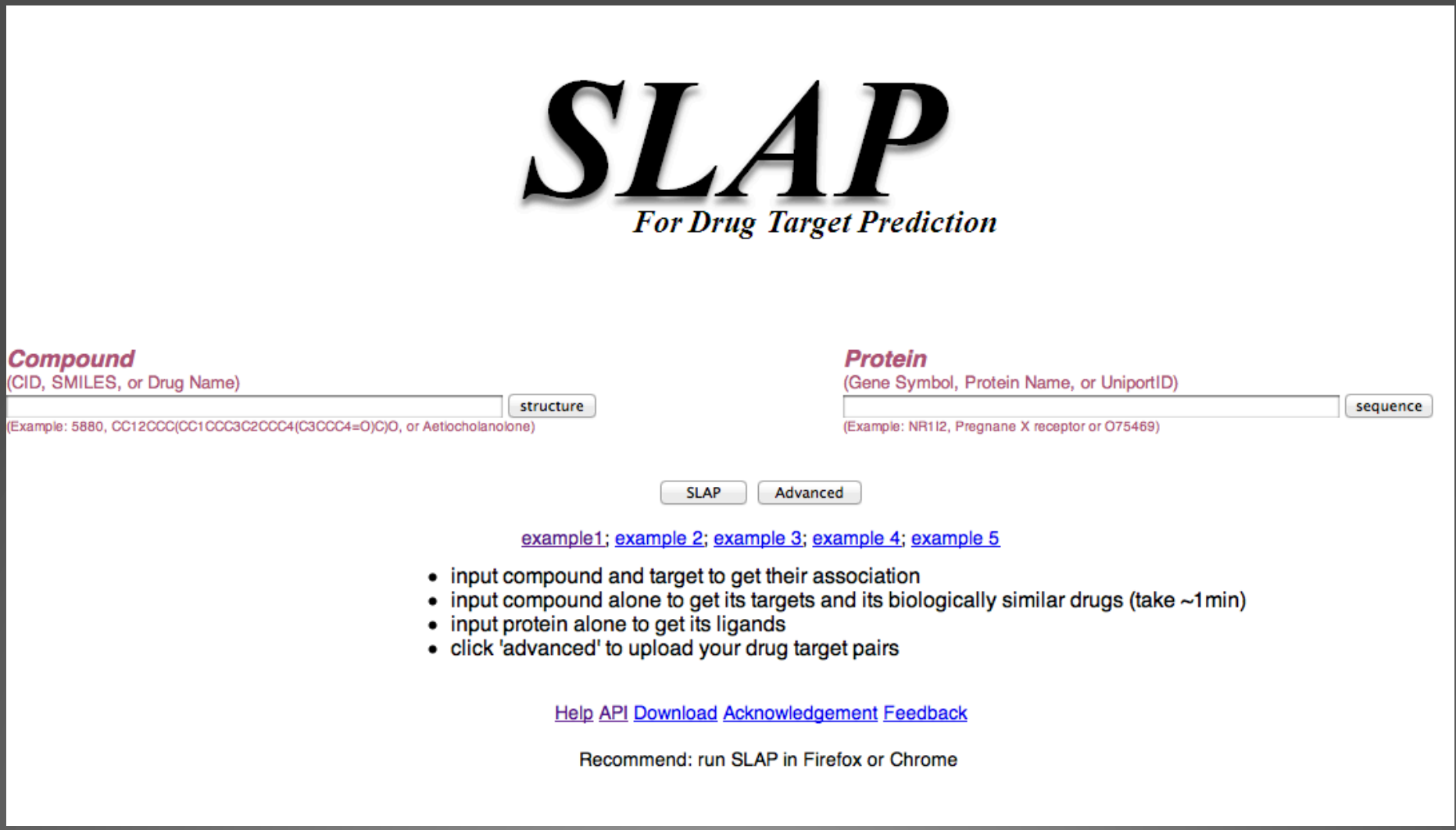
- SLAP can be used to build a polypharmacology profile of a compound across targets without a full experimental data matrix
- Connectivity Map (CMap) maps experimental drug-gene-disease links through expression changes in samples treated with drugs
- Association scores were calculated for 113 drugs from the CMap library also present in our dataset against 1,683 targets in our set; 679 targets were associated with at least one drug
- Both predictive SLAP and experimental CMAP profiles were used to calculate the most similar drugs for 8 query drugs (similarity > 0.75)
- Of 21 pairs identified by SLAP, 19 were verified using CMAP experimental profiles
- Indicates SLAP profiles closely match expression profiles: wider study needed

Assessing drug similarity from biological function

- Took 157 drugs with 10 known therapeutic indications, and created SLAP profils against 1,683 human targets
- Pearson correlation between profiles > 0.9 was used to create associations between drugs
- Drugs with the same therapeutic indication unsurprisingly cluster together – also subcluster by MOA
- Some drugs with similar profile have different indications – potential for use in drug repurposing?



SLAP web tool http://chem2bio2rdf.org/slap



SLAP potential uses

- Polypharmacology profiling
  - Predicts association for a compound across targets (currently 1,683 human targets from Drugbank)
  - For a compound, identifies most biologically similar drugs to a compound via target association profiles
- Building predictive drug and target networks
  - We predicted the association scores for 174 drugs from 10 disease areas against 1683 human targets, and measured their *biological* similarity using a 174*1683 score matrix.
  - The similarity network indicates that drugs from the same disease area tend to cluster together in ways which are not captured by structural similarity, with several potential new drug pairings being identified (see next slide)
- Mechanism of action hypothesis generation and testing
  - Given a compound and a gene, will provide numeric prediction of likelihood of a relationship between them, and will allow exploration of the paths between them
- Virtual Screening
  - Given a target/gene, will rank all PubChem "biologically active" compounds by biological association
  - Given a compound, will rank all targets/genes in a set by strength of association to the compound
  - Uses chemical structure or BLAST similarity for compounds or targets not in the dataset

SLAP – target profile for Ibuprofen

COX2 – main target

Regulate neurotransmitter release

COX1

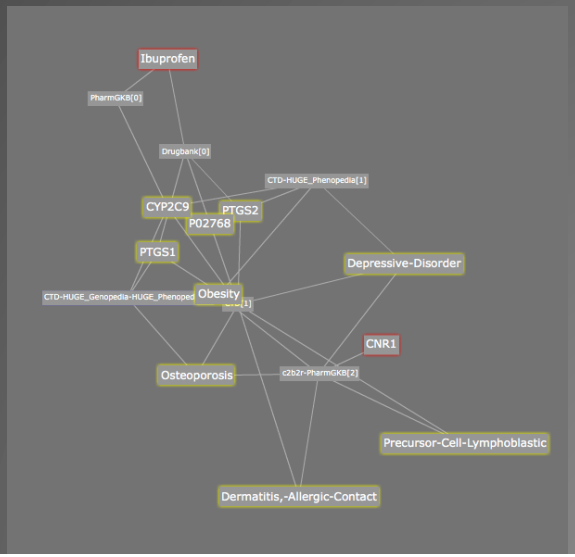
Dopamine receptor

Seratonin receptors

Cannaboid receptors

Muscarinic receptor (motor control)

target	p value	score	type	chemohub
PTGS2	3.26050638978e-05	1509.09	approved interaction	see paths
ADRA2A	3.5824898966e-05	1457.32	approved interaction	see paths
ADRA2C	4.21593691741e-05	1371.39	approved interaction	see paths
ADRA2B	4.28844317198e-05	1362.64	approved interaction	see paths
PTGS1	4.54838408268e-05	1332.83	approved interaction	see paths
DRD2	5.2290304563e-05	1264.4	approved interaction	see paths
HTR1B	5.51387174742e-05	1239.16	approved interaction	see paths
HTR1D	5.65393044882e-05	1227.38	approved interaction	see paths
PPARA	6.38373424718e-05	1171.69	predicted	see paths
CNR1	6.62383224405e-05	1155.19	approved interaction	see paths
HTR2C	6.88598161789e-05	1138.06	approved interaction	see paths
HTR2A	7.0120066649e-05	1130.12	approved interaction	see paths
CHRM4	9.60843321586e-05	999.63	predicted	see paths
DRD3	9.89314009341e-05	988.21	approved interaction	see paths
CHRM3	1e-04	943.88	predicted	see paths
CNR2	1e-04	872.05	approved interaction	see paths
DRD1	1e-04	978.37	approved interaction	see paths
DRD4	1e-04	860.45	approved interaction	see paths
DRD5	1e-04	879.33	approved interaction	see paths
HTR1F	1e-04	957.54	predicted	see paths
CHRM1	2e-04	812.45	predicted	see paths
HRH1	2e-04	711.14	approved interaction	see paths
HTR1A	2e-04	780.84	approved interaction	see paths
HTR2B	2e-04	760.87	approved interaction	see paths
CYP3A4	4e-04	536.09	predicted	see paths
HRH2	4e-04	535.76	approved interaction	see paths
HTR7	4e-04	576.89	approved interaction	see paths
CHRM2	5e-04	503.53	approved interaction	see paths
PPARD	5e-04	525.77	predicted	see paths
CTSD	6e-04	465.05	predicted	see paths
ADRA1A	7e-04	444.41	approved interaction	see paths
ALOX5	7e-04	426.67	approved interaction	see paths
SLC6A4	9e-04	391.26	approved interaction	see paths
HTR1E	0.001	374.64	approved interaction	see paths
HTR6	0.001	365.97	approved interaction	see paths
KCNH2	0.001	375.03	predicted	see paths
HRH4	0.0011	358.5	predicted	see paths
HTR4	0.0011	360.28	predicted	see paths
NISCH	0.0012	345.55	predicted	see paths
ADRA1B	0.0013	324.26	approved interaction	see paths
ADRA1D	0.0013	329.71	predicted	see paths



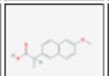
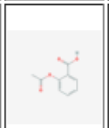
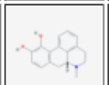
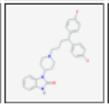
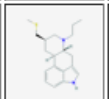
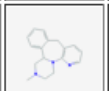
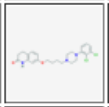
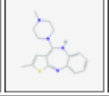
BBC NEWS | Health | Brain '**cannabis**' **Parkinson's** hope
news.bbc.co.uk/2/hi/health/6338173.stm +1
Feb 8, 2007 – Boosting the brain's natural **cannabis**-like chemicals could help treat **Parkinson's** disease, a study suggests.

target	p value	score	type	chemohub
CA2	1.65306498257e-06	4210.28	approved interaction	see paths
CA1	1.74645404993e-06	4136.42	approved interaction	see paths
CA9	3.31332517922e-06	3355.42	predicted	see paths
CA4	4.01729181176e-06	3147.13	approved interaction	see paths
CA12	1.40813617761e-05	2044.74	approved interaction	see paths
CA5A	1.70469337978e-05	1910.31	predicted	see paths
CA5B	1.84232712477e-05	1857.9	predicted	see paths
CA14	2.12933809418e-05	1763.48	approved interaction	see paths
CA13	3.01933662642e-05	1552.5	predicted	see paths
PTGS2	3.08519257795e-05	1540.2	approved interaction	see paths
CA7	3.72309234737e-05	1436.6	approved interaction	see paths
PTGS1	4.59732695355e-05	1327.47	approved interaction	see paths
MAOA	8.66262518232e-05	1041.07	approved interaction	see paths
MAOB	9.68885985199e-05	996.36	approved interaction	see paths
CA3	1e-04	905.88	approved interaction	see paths
CYP2C9	2e-04	693.39	predicted	see paths
ALOX5	3e-04	625.81	predicted	see paths
CYP1A2	5e-04	499.14	predicted	see paths
CYP3A4	5e-04	505.25	approved interaction	see paths
CA11	7e-04	428.26	predicted	see paths
CA8	7e-04	429.39	predicted	see paths
CA10	9e-04	390.41	predicted	see paths
FAAH	0.0016	294.36	approved interaction	see paths
ABCC1	0.0021	256.18	predicted	see paths

target	p value	score	type	chemohub
ADRA2A	2.2583754428e-05	1726.31	approved interaction	see paths
DRD2	2.44050813141e-05	1678.33	approved interaction	see paths
ADRA2C	2.68583734929e-05	1620.66	approved interaction	see paths
ADRA2B	2.74390344807e-05	1608.01	approved interaction	see paths
ITGB3	3.12162596043e-05	1533.55	approved interaction	see paths
HTR2A	3.23337738672e-05	1513.76	approved interaction	see paths
HTR2C	3.39930096688e-05	1485.98	approved interaction	see paths
PTGS2	3.89837859761e-05	1412.16	approved interaction	see paths
HTR1D	4.26993941995e-05	1364.86	approved interaction	see paths
HTR1B	4.52979812635e-05	1334.89	approved interaction	see paths
ITGAV	4.82133278797e-05	1303.86	predicted	see paths
DRD3	5.10143416836e-05	1276.3	approved interaction	see paths
CNR1	5.26649608524e-05	1260.97	approved interaction	see paths
CHRM4	5.29594702321e-05	1258.31	predicted	see paths
DRD1	5.4408421828e-05	1245.47	approved interaction	see paths
CHRM3	5.54364524297e-05	1236.63	predicted	see paths
ITGA2B	5.76549687449e-05	1218.27	approved interaction	see paths
PTGS1	6.12612623494e-05	1190.34	approved interaction	see paths
DRD4	7.59017015987e-05	1096.02	approved interaction	see paths
DRD5	7.96090789503e-05	1075.91	approved interaction	see paths

vs acetaminophen and aspirin

SLAP - Biologically Similar Drugs

PubChem CID	structure	Drug Name	Similarity	Related Diseases	ATC
5090		Rofecoxib	0.949	Osteoarthritis	M01AH02
1302		Naproxen	0.947	Osteoarthritis Rheumatoid arthritis	G02CC02 M01AE02 M02AA12
2244		Aspirin	0.891	Lupus erythematosus Myocardial infarction Osteoarthritis Rheumatoid arthritis Systemic lupus erythematosus	A01AD05 B01AC06 N02BA01
6005		Apomorphine	0.785	Parkinson disease	G04BE07 N04BC07
16362		Pimozide	0.784	Rett syndrome	N05AG02
4168		Metoclopramide	0.778		A03FA01
47811		Pergolide	0.775	Parkinson disease	N04BC02
4205		Mirtazapine	0.769	Major depressive disorder	N06AX11
60795		Aripiprazole	0.757	Schizophrenia	N05AX12
4585		Olanzapine	0.752	Bipolar disorder Schizophrenia	N05AH03

Taking Ibuprophen/**Naproxen** regularly reduces **parkinsons** by 45%...
www.prohealth.com/me-cfs/blog/boardDetail.cfm?id=261855 +1
Aug 20, 2003 – The risk of developing the disease was 45 percent lower in people who used drugs such as ibuprofen and **naproxen** at least twice weekly than ...

Dopamine agonist, used in Parkinson's

J Neurol Neurosurg Psychiatry. 1975 Apr;38(4):331-5.

Metoclopramide and pimozide in Parkinson's disease and levodopa-induced dyskinesias.

Tarsy D, Parkes JD, Marsden CD.

Abstract

Metoclopramide is an antiemetic drug which occasionally produced acute dystonic reactions. Although known to interfere with central dopamine mechanisms, it is frequently used in Parkinson's disease to prevent levodopa-induced nausea and vomiting. In this study metoclopramide did not increase Parkinsonism or reduce levodopa-induced involuntary movements in patients with Parkinson's disease. Pimozide, by contrast, increased Parkinsonism and reduced involuntary movements. The capacity of metoclopramide to produce acute dyskinesias while being apparently free of Parkinsonism effects is pharmacologically unique and differentiates this drug from the phenothiazines and butyrophenones.

Dopamine agonist, used in Parkinson's

Improvement of Movement Disorders with **Mirtazapine**: A Preliminary ...
www.medicosecuador.com/revecuatneurol/vol12.../improventa.htm
by F Alarcón - Related articles
Mirtazapine is an anti-depressant with a mechanism that is different from .... In: Jankovic J, Tolosa E (eds) **Parkinson's** Disease and Movement Disorders.

Troglitazone vs Rosiglitazone

Predicted Targets

target	p value	score	type	chemohub
PPARA	8.64215768781e-06	2425.72	approved interaction	see paths
PPARG	9.06524962196e-06	2385.9	approved interaction	see paths
PPARD	9.94000104049e-06	2310.75	predicted	see paths
CYP2C9	2e-04	680.81	predicted	see paths
CCND1	0.0014	316.04	predicted	see paths
ITGAL	0.0015	302.82	predicted	see paths
SELE	0.0018	280.63	predicted	see paths
CDK4	0.0019	271.81	predicted	see paths
CYP1A2	0.0029	220.56	predicted	see paths
CYP2C8	0.0029	219.62	predicted	see paths
ITGA4	0.0029	220.07	predicted	see paths
CCNA2	0.0031	212.3	approved expression	see paths
CDK2	0.0032	207.61	approved expression	see paths
RXRA	0.0037	193.91	predicted	see paths
CYP2B6	0.0043	178.6	approved expression	see paths
CYP3A4	0.0045	175.46	approved interaction	see paths
CYP2A6	0.0053	160.41	predicted	see paths
ABCB1	0.0061	148.41	predicted	see paths
ABCC2	0.0072	135.42	predicted	see paths

PubChem CID	structure	Drug Name	Similarity	Related Diseases	ATC
5591		Troglitazone	0.999999999844236	Diabetes mellitus	A10BG01
4829		Pioglitazone	0.975129714002987	Diabetes mellitus	A10BG03
60823		Atorvastatin	0.964568661853574	Hypercholesterolemia	C10AA05
3339		Fenofibrate	0.9587359473525	Hypercholesterolemia	C10AB05
54786		Treprostinil	0.776290578790852	Hypertension Pulmonary hypertension, familial primary	B01AC21
2796		Clofibrate	0.776056284952304	Hyperlipidemia Protein S deficiency	C10AB01 C10AB03

target	p value	score	type	chemohub
PPARA	1.14895307912e-05	2196.68	predicted	see paths
PPARD	1e-04	960.54	predicted	see paths
CYP3A4	2e-04	793.65	predicted	see paths
CYP1A2	0.0015	305.31	predicted	see paths
CYP2C9	0.0035	201.18	approved interaction	see paths
CYP2B6	0.006	149.19	predicted	see paths
CYP2C8	0.0076	131.21	approved interaction	see paths
CYP2C18	0.008	127.03	predicted	see paths
RXRA	0.0093	116.63	predicted	see paths
PPARG	0.0138	92.64	approved interaction	see paths
APOE	0.0686	30.23	approved interaction	see paths
HPGD	0.1079	20.51	approved interaction	see paths

PubChem CID	structure	Drug Name	Similarity	Related Diseases	ATC
77999		Rosiglitazone	1	Diabetes mellitus	A10BG02
4829		Pioglitazone	0.768	Diabetes mellitus	A10BG03

Ongoing work

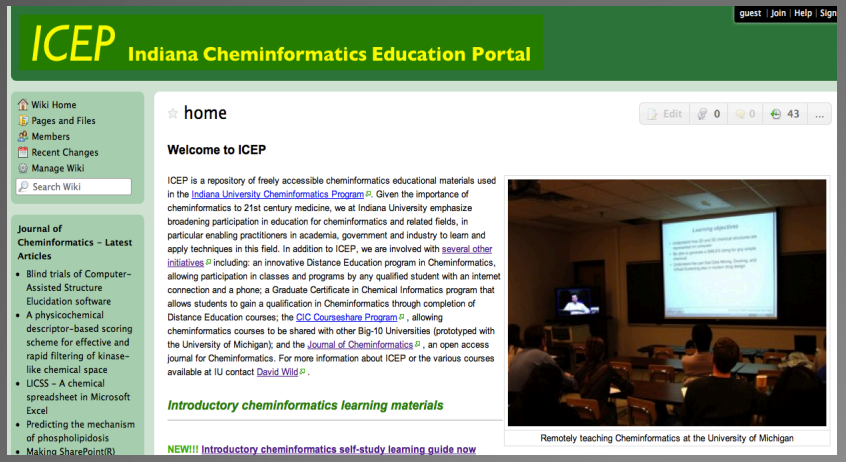
- Proper comparison with other drug-target prediction methods
  - Using experimental matrix (Chen, B., et al., *Bioinformatics*, 2011, 27(21), 3044-3049
- Wider scale drug-target profile similarity network using all of DrugBank
- Experimental validation for virtual screening in PXR and TB projects
  - Including data fusion with predictive models and other LBVS methods
- Many opportunities for improvement
  - Properly include quantitative activity data
    - IC50 etc. subclasses of binding, with weighting by value
  - Qualitative weighting of path patterns

Cheminformatics Education at Indiana University

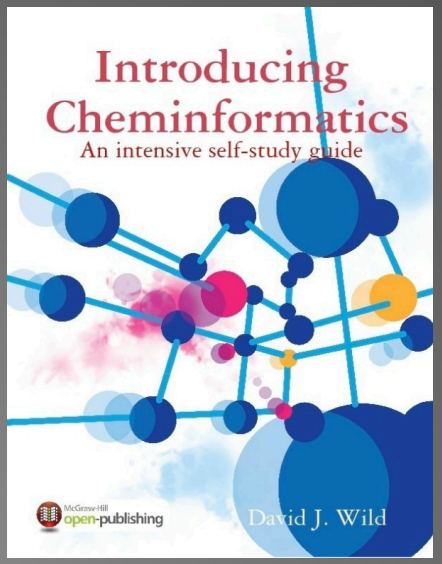
Residential Ph.D. program in Informatics with a Cheminformatics specialty

Distance Graduate Certificate program in Chemical Informatics

http://djwild.info/ed



Free cheminformatics learning resources
http://icep.wikispaces.com



LuLu eBook - $29
http://slg.djwild.info