Introduction to Chem2Bio2RDF

Presented at LODD CALL 05/12/2010

By Xiao Dong

(On Behalf of Chem2Bio2RDF Team)

Team Members

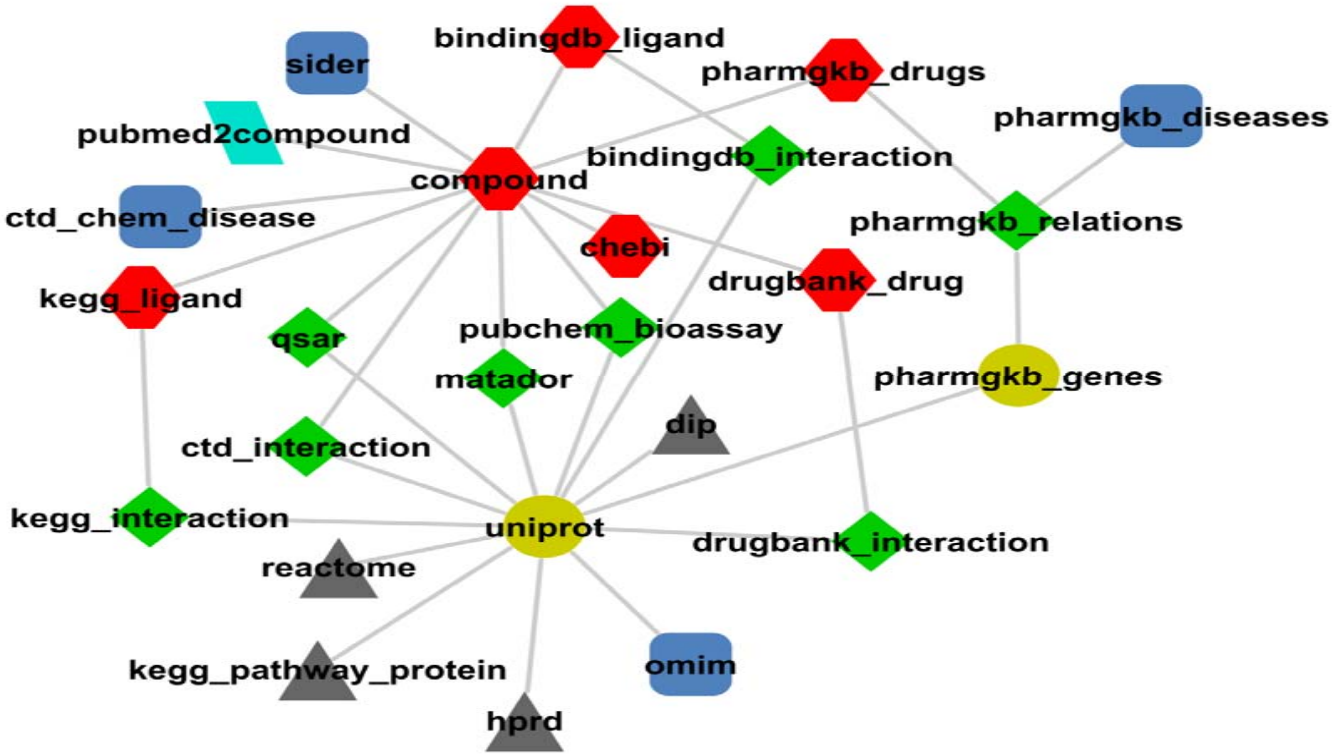
- Advisor:
  - Prof. David Wild (Chemical Informatics)
  - Prof. Ying Ding (Information Science)
- Students:
  - Bin Chen (Chemical Informatics, Novartis Intern 08, Pfizer intern 09,10)
  - Huijun Wang (Chemical Informatics, Pfizer intern 08, 09)
  - Xiao Dong (Chemical Informatics, Eli Lilly intern 08)

Impact

- Association ranking methodology paper presented at the FWCS2010 two weeks ago.
- One BMC Bioinformatics Paper (forthcoming)
- Winner of 2010 CINF Scientific Excellence Poster Competition @ ACS Spring Meeting

  (connects to ACS outreach effort)

Chem2Bio2RDF



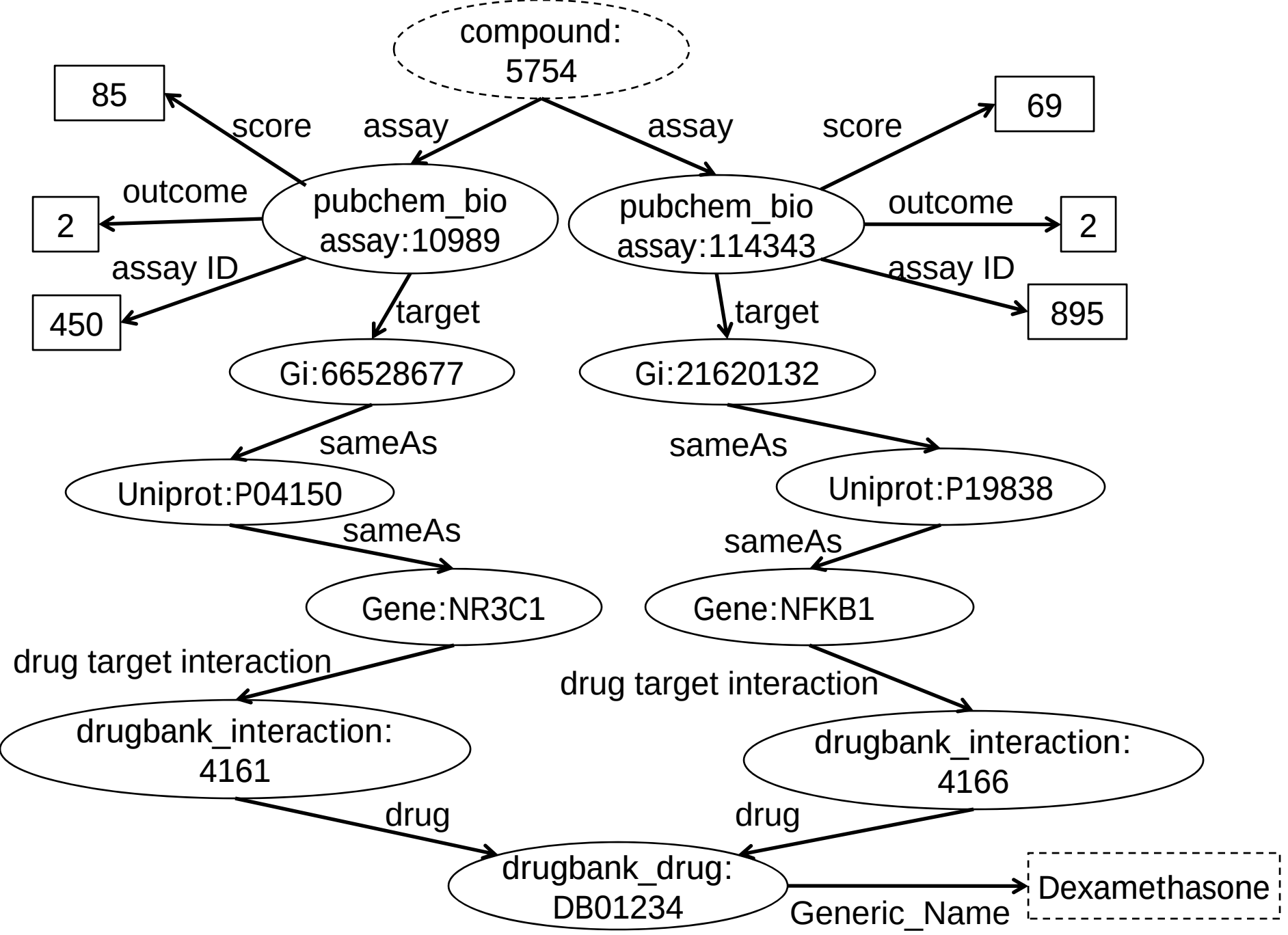
- www.chem2bio2rdf.org
- Focus on Chemical Systems Biology
- 24 databases & 82,795,324 triples
- Overlapping portions extended to LODD and Bio2RDF by dereferenceable URL's

Fact Sheet

- Virtuoso endpoint:
  http://cheminfov.informatics.indiana.edu:8890/sparql
- Faceted Data Source Browser (Powered by SIMILE):
  http://chem2bio2rdf.org/datasets.html
- some sample sparqls:
  http://chem2bio2rdf.wikispaces.com/multiple+sources
- Three cases in BMC paper:
  http://chem2bio2rdf.wikispaces.com/Three+cases+in+the+BMC

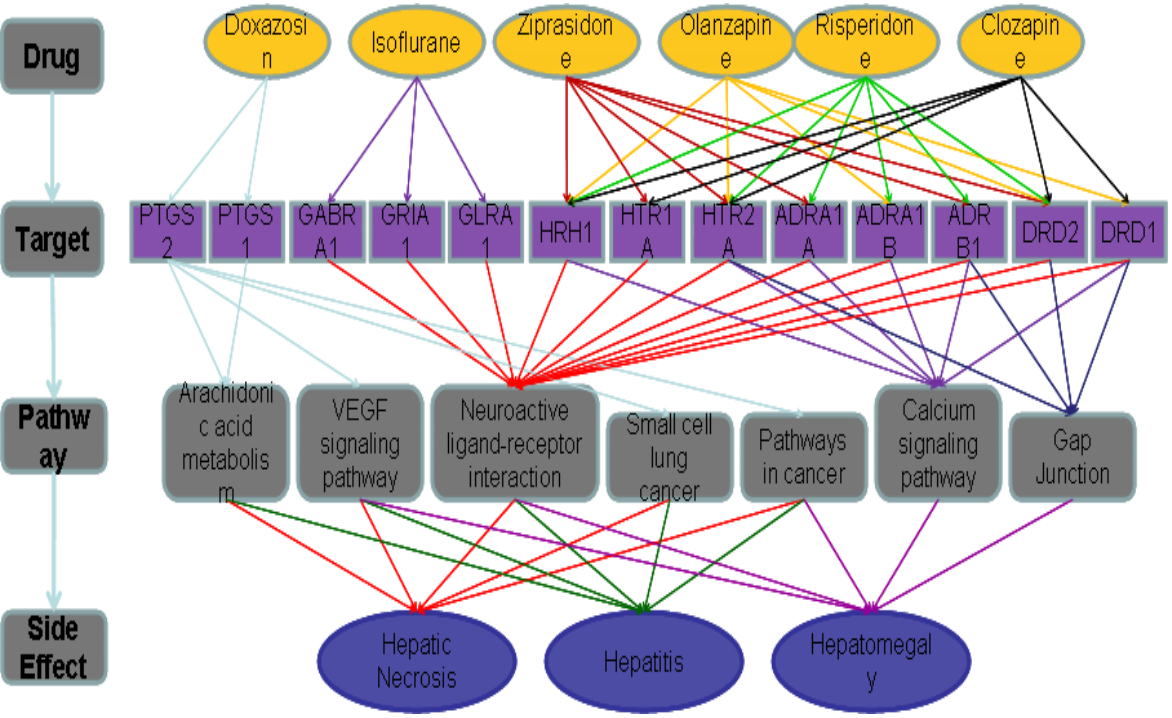
Case I: Polypharmacology

- Conventional drug embraces the dogma “one gene, one drug, one disease”
- Polypharmacology focuses on multi-target drugs
  - Identify target leads to unwanted side effect
  - Enhance therapeutic potency
- Identify compounds sharing targets with drugs of known polypharmacology


compound:
5754
85
score
assay
assay
score
69
outcome
2
pubchem_bio
assay:10989
pubchem_bio
assay:114343
outcome
2
assay ID
450
target
target
assay ID
895
Gi:66528677
Gi:21620132
sameAs
sameAs
Uniprot:P04150
Uniprot:P19838
sameAs
sameAs
Gene:NR3C1
Gene:NFKB1
drug target interaction
drug target interaction
drugbank_interaction:
4161
drugbank_interaction:
4166
drug
drug
drugbank_drug:
DB01234
Generic_Name
Dexamethasone

Case II:
Adverse Drug Reaction Study

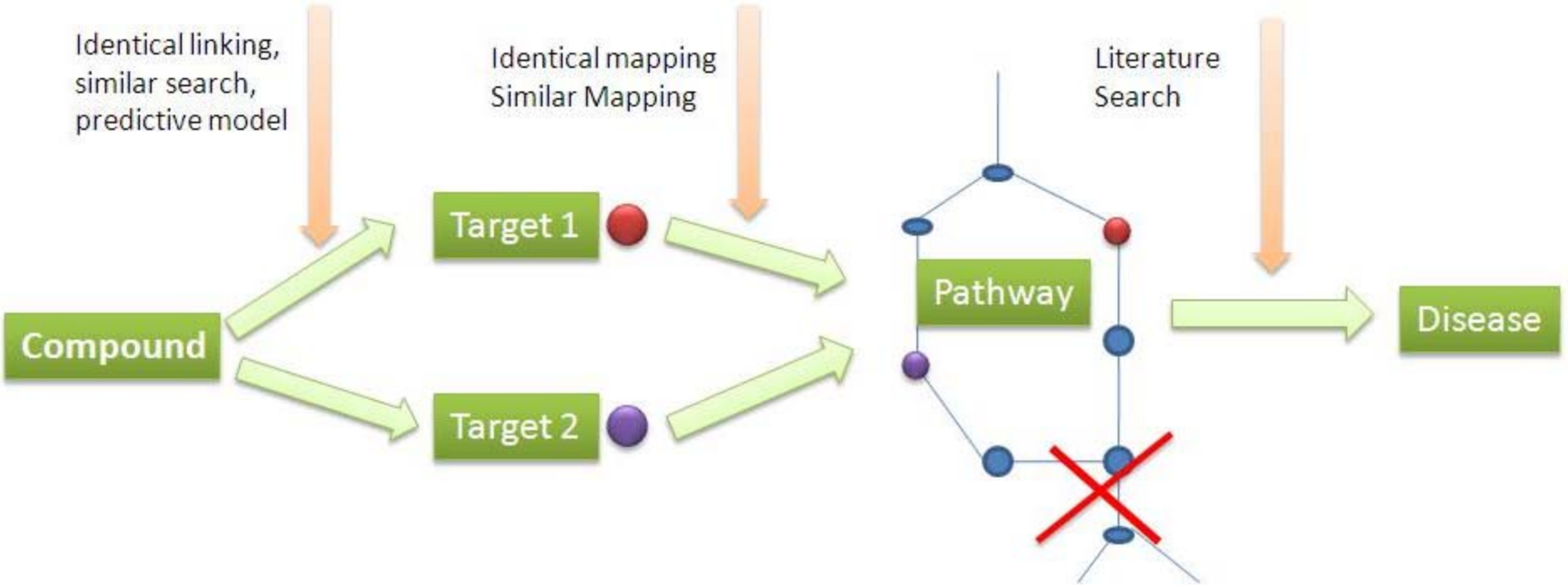
Question: *find the top 5 pathways in the KEGG pathway dataset that contain at least two efficient targets that have drugs that are associated with hepatotoxicity.*

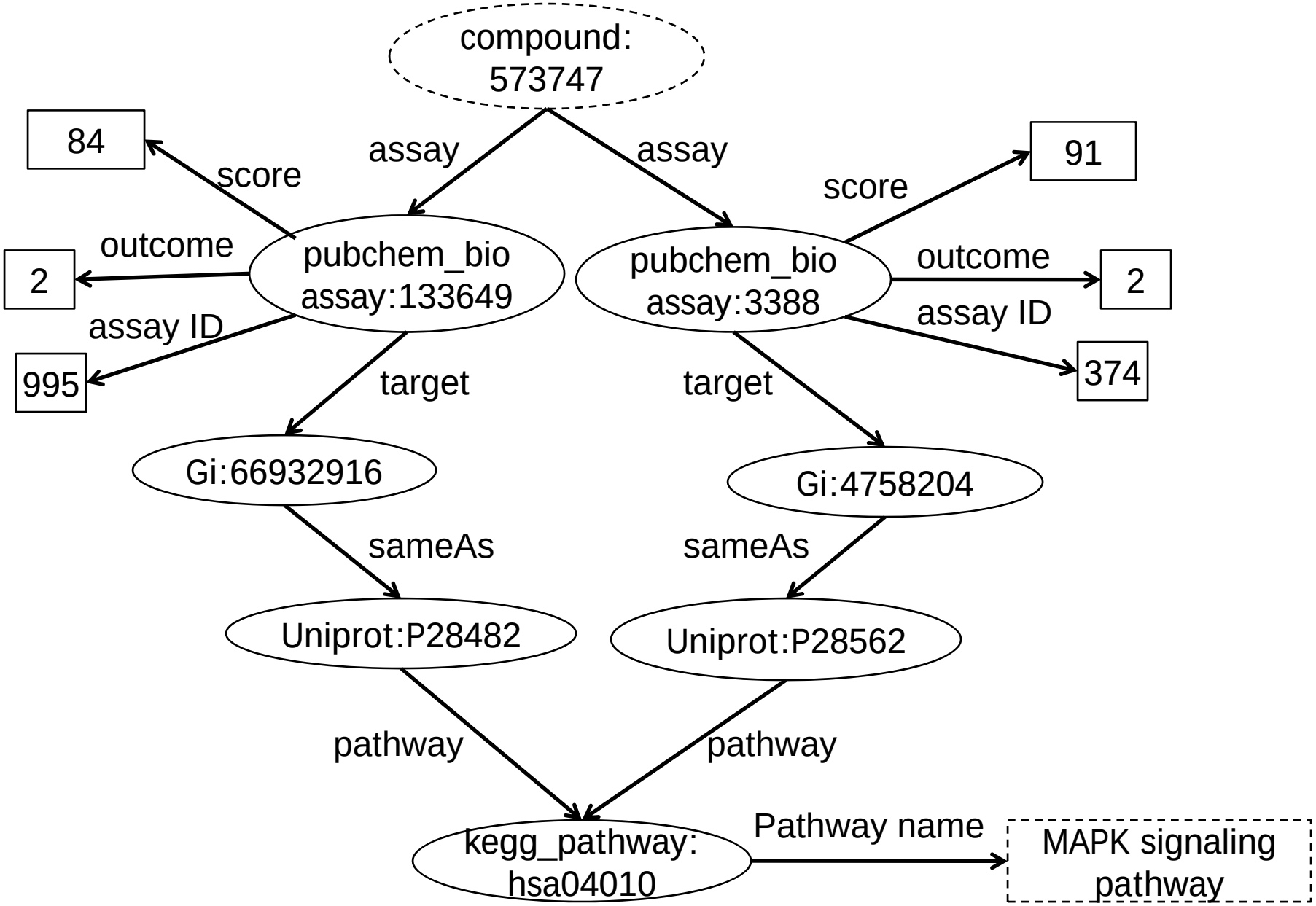


SPARQL:

```
SELECT ?pathway_id (count(?pathway_id) as ?count) WHERE {
 ?sider2compound  sider:side_effect  ?side_effect  .   FILTER
regex(?side_effect,"hepatomegaly","i") .
 ?sider2compound sider:cid ?compound .

 ?drug drugbank_drug:CID ?compound .
 ?drug2target drugbank_interaction:DBID ?drug .
 ?drug2target drugbank_interaction:SwissProt_ID ?uniprot .

 ?kegg_pathway kegg_pathway_protein:Uniprot ?uniprot .
 ?kegg_pathway kegg_pathway_protein:PathwayID ?pathway_id .

} GROUP BY ?pathway_id ORDER BY ?count
```

Case III:
Multiple Pathway Inhibitor




compound:
573747
84
score
assay
assay
91
score
outcome
2
pubchem_bio
assay:133649
pubchem_bio
assay:3388
outcome
2
assay ID
995
assay ID
374
target
target
Gi:66932916
Gi:4758204
sameAs
sameAs
Uniprot:P28482
Uniprot:P28562
pathway
pathway
kegg_pathway:
hsa04010
Pathway name
MAPK signaling
pathway

Possibility Within LODD

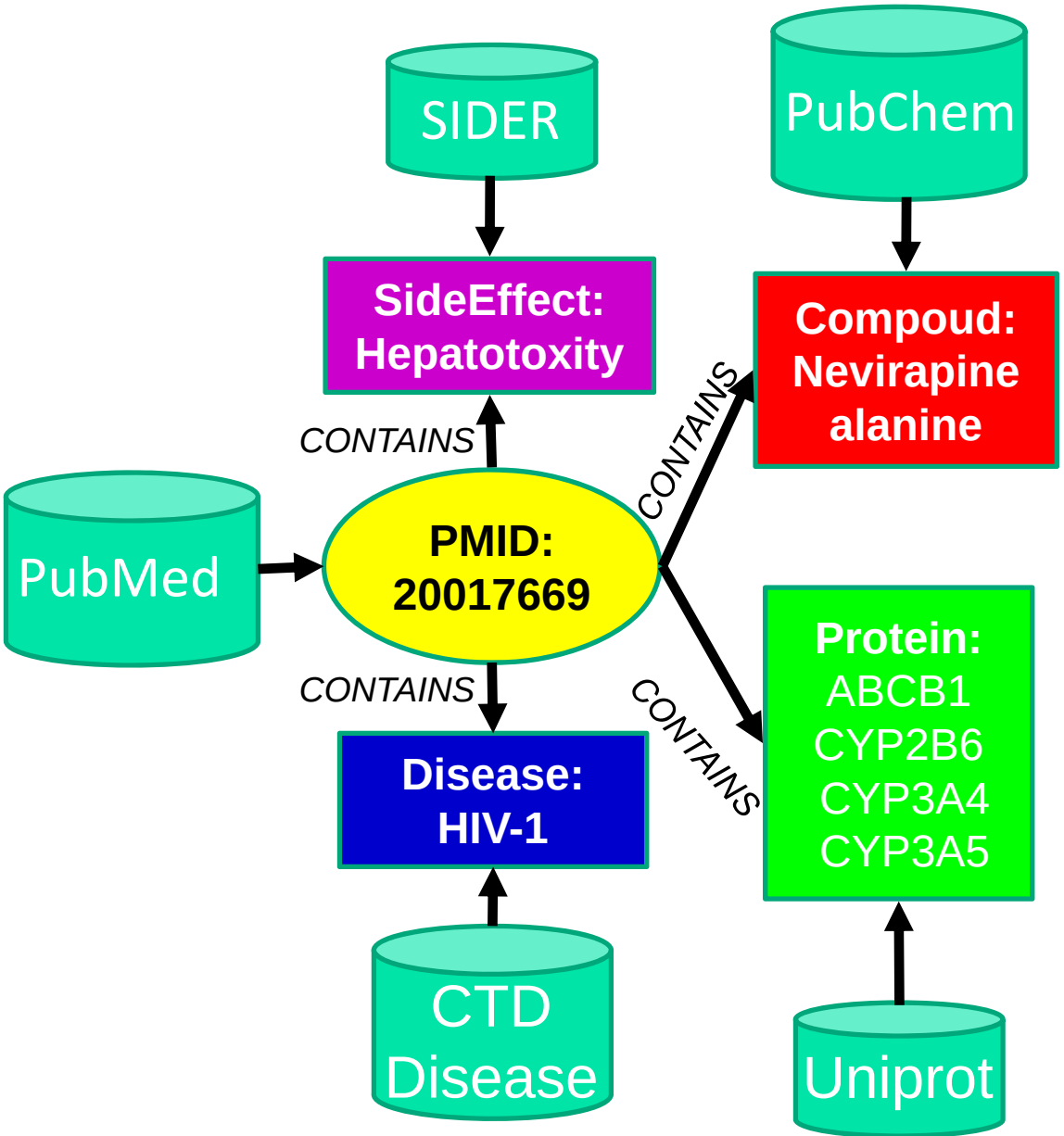
- Comprehensive chemogenomics coverage by Chem2Bio2RDF
- Use case development
  - With emphasis on chemogenomics and system chemical biology
  - Compound, aggregated use case development
- Domain entity extraction from literature

Chem/Bio Entity Extraction from PubMed

PMID: 20017669

Title: Nevirapine-induced hepatotoxicity and pharmacogenetics: a retrospective study in a population from Mozambique.

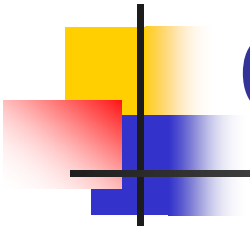
Abstract: Aims: Nevirapine is widely used to treat HIV-1 infection to prevent mother-to-child transmission; unfortunately adverse drug reactions have been reported. Our aim was to identify genes/variants involved in nevirapine-induced hepatotoxicity. MATERIALS & METHODS: Patients from Mozambique, 78 with nevirapine-induced hepatotoxicity and 78 without adverse events, were genotyped for ABCB1, CYP2B6, CYP3A4 and CYP3A5 gene variants. We conducted a case-control association study and a genotype/phenotype correlation analysis. RESULTS: The ABCB1 c.3435C>T SNP was associated with hepatotoxicity ($p = 0.038$), with the variant T allele showing a protective effect (odds ratio: 0.42). Moreover, four SNPs in the CYP2B6 and CYP3A5 genes resulted significantly correlated with transaminase values. In particular, for the CYP2B6 c.983T>C SNP the difference in the alanine aminotransferase mean values were highly significant between TT and TC genotypes ($p < 0.001$). CONCLUSION: Our preliminary results confirm the contribution of the ABCB1 c.3435C>T SNP in nevirapine-induced hepatotoxicity risk and, at the same time, suggest the necessity for further studies.



Biomedical Entity Extraction from Medline

- 18,502,916PubMed/Medline literature records from 1865-2009
  - 56,383 compounds and 11,775,891 literature-compound pairs
  - 2,820 drugs and 5,624,529 literature-drug pairs
  - 13,022 human genes and 5,252,844 literature-gene pairs
  - 3,848 diseases and 12,612,636 literature-disease pairs
  - 1363 side effects and 10,489,676 literature-side effect pairs
  - 180 human pathways and 916,754 literature-pathway pairs
- Structure/biology activities for 56,911,891 compounds and 4764 drugs
- Systems chemical biology data
- 70,134protein-proteininteraction
  - 1,992,907 gene-compound
  - 297422 disease-compound
  - 1,872,260 disease-gene
  - 103,830 pathway-gene
  - 1,527,229 pathway-compound Literature

Courtesy of Huijun Wang



Q&A