# Open Phenotypic Drug Discovery Resource (OPDDR)
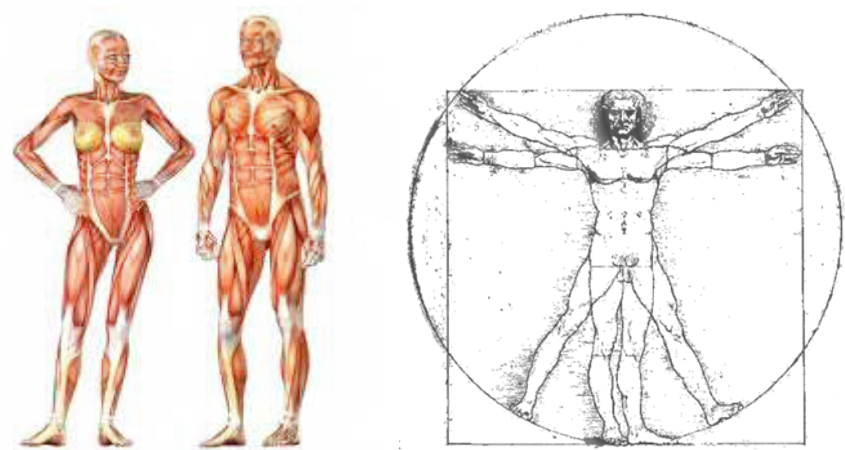
Jeremy J Yang[1,2], Natalie I Franklin[1,3], Rajarshi Guha[4], Ajit Jadhav[4] and David J Wild[1]

[1]Integrative Data Science Lab, Indiana University, Bloomington, Indiana, USA
[2]Translational Informatics Division, School of Medicine, University of New Mexico, USA
[3]Open Innovation Drug Discovery Program, Eli Lilly & Co., Indianapolis, Indiana, USA
[4]NIH National Center for Advanced Translational Science, Rockville, Maryland, USA
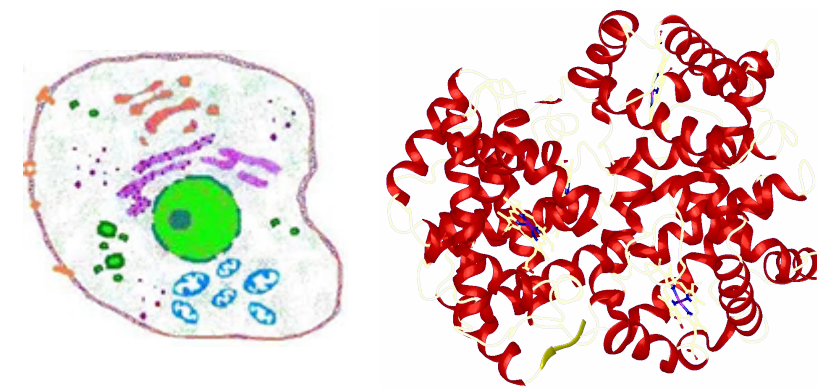
## Collaboration

- Lilly OIDD - phenotypic assays
- NIH NCATS - Pharmaceutical Collection (NPC) compounds
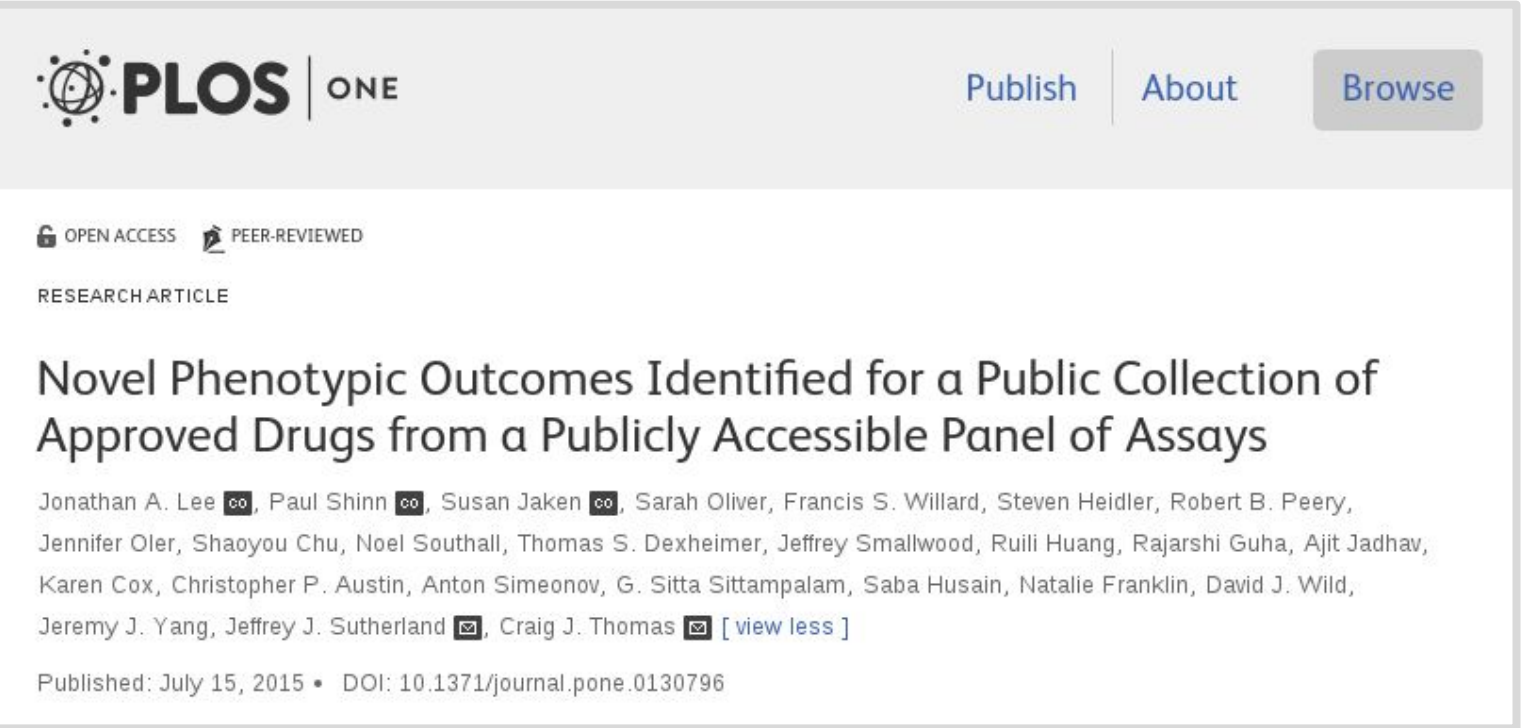- Data2Discovery/IU - informatics, semantics

## Experiments

- NCATS (NPC) compounds (2509)
- OIDD phenotypic assays (35 assays across 5 modules)
- Relevance: cardiovascular, diabetes, cancer, endocrine

## Why phenotypic?

- Phenotypic assays more *biologically* relevant.
- But, require analytics for *molecular* inferences.
- Phenomics reflects systems biology.
- Phenotypic assay phenotypes are rigorously defined, observable biological effects, often well associated with disease states.
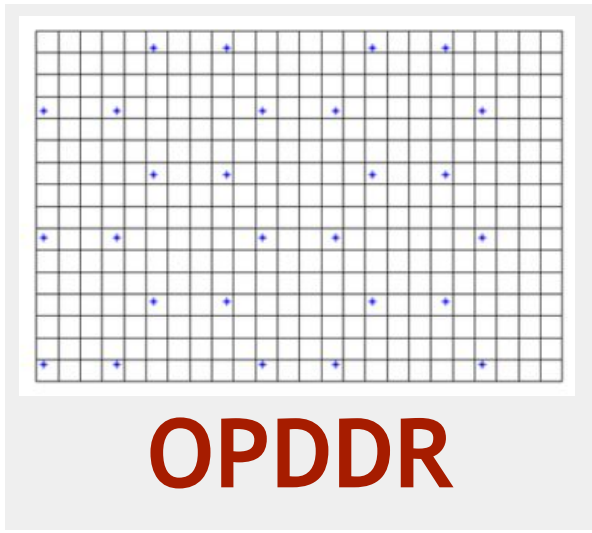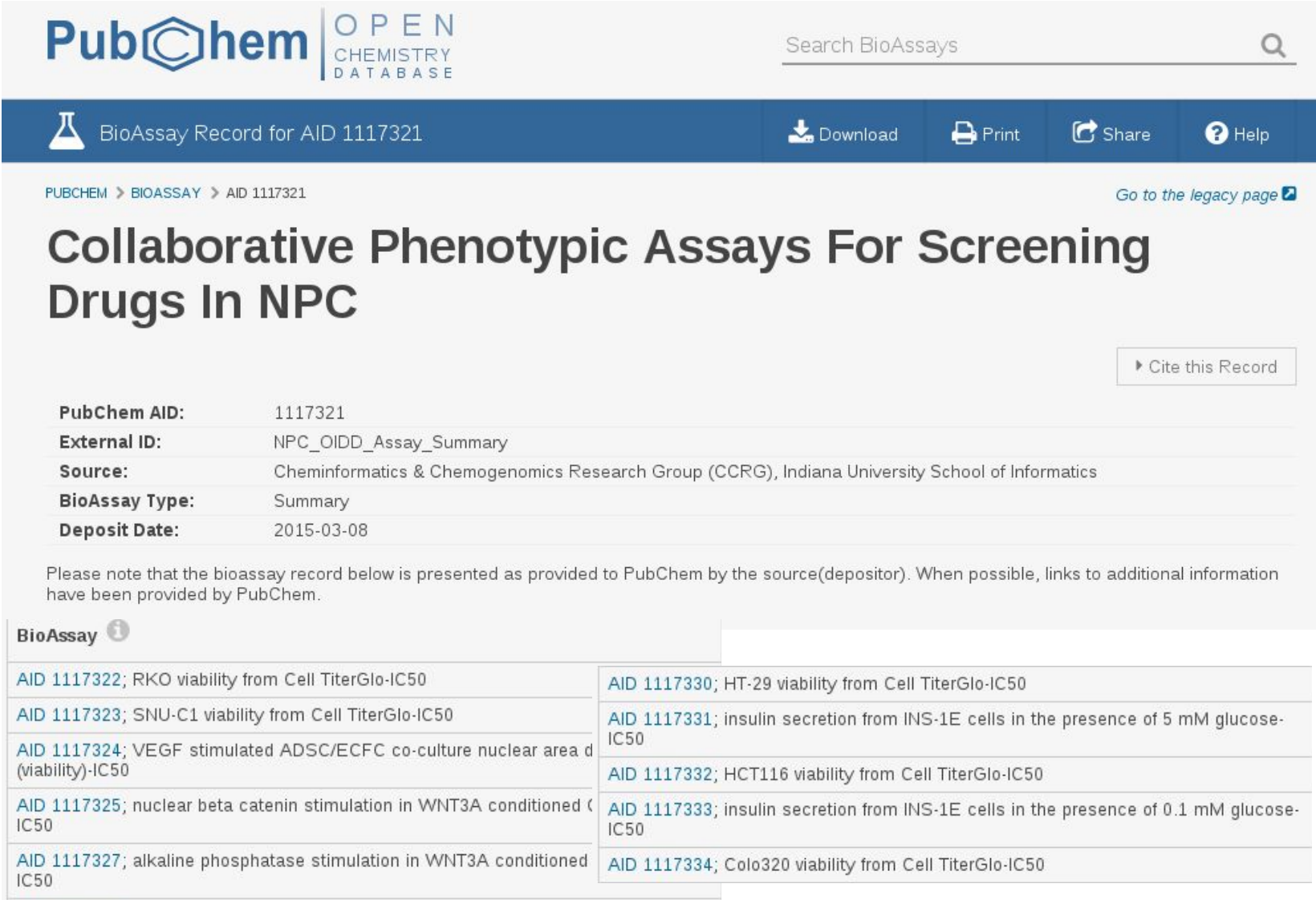
## Semantic engineering

- OPDDR RDB to RDF transformation
- Manual annotation via BAO
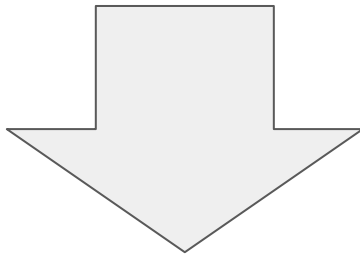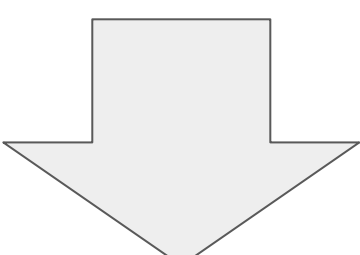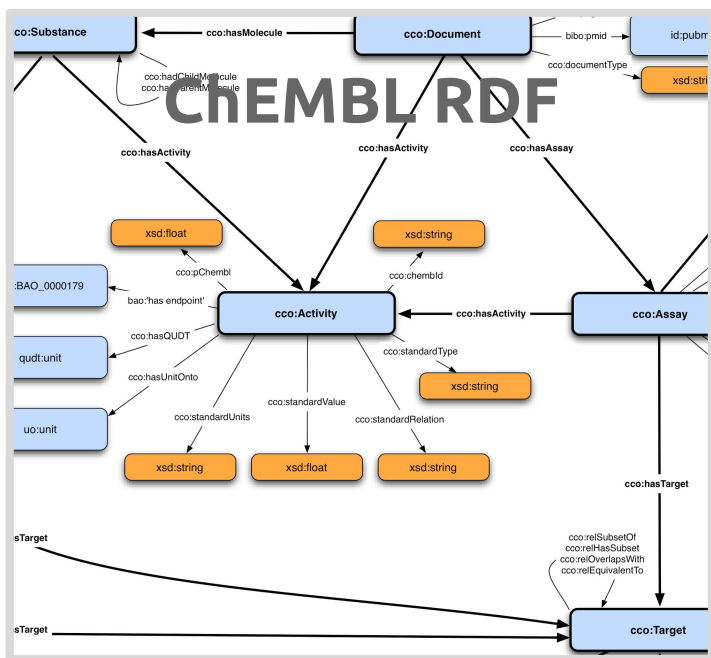- Integration: PubChem, Chembl, Open PHACTS





Novel Phenotypic Outcomes Identified for a Public Collection of Approved Drugs from a Publicly Accessible Panel of Assays

Jonathan A. Lee, Paul Shinn, Susan Jaken, Sarah Oliver, Francis S. Willard, Steven Heidler, Robert B. Peery, Jennifer Oler, Shaoyou Chu, Noel Southall, Thomas S. Dexheimer, Jeffrey Smallwood, Ruili Huang, Rajarshi Guha, Ajit Jadhav, Karen Cox, Christopher P. Austin, Anton Simeonov, G. Sitta Sittampalam, Saba Husain, Natalie Franklin, David J. Wild, Jeremy J. Yang, Jeffrey J. Sutherland, Craig J. Thomas [ view less ]

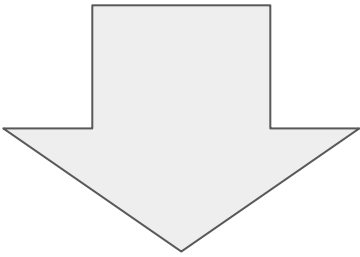Published: July 15, 2015 • DOI: 10.1371/journal.pone.0130796

## Publication

- PubChem Bioassay (March 2015)
- PLOS One (July 2015)
- NCATS site: https://ncats.nih.gov/expertise/preclinical/pd2



## OPDDR

## RDB → RDF



### PubChem RDF

### ChEMBL RDF
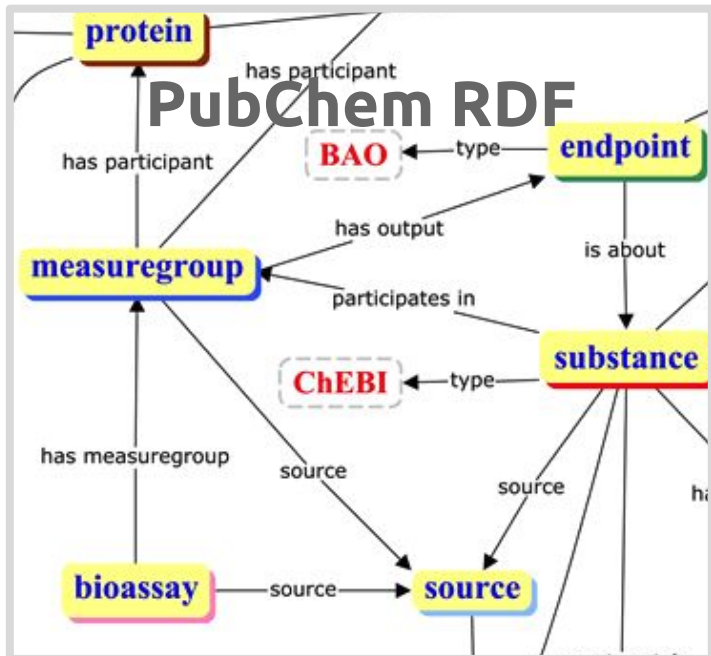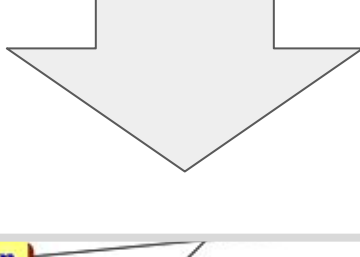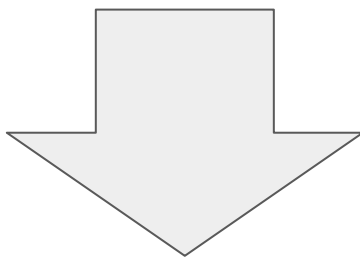
## Open PHACTS

## Community semantics

- Cooperation with PubChem, ChEMBL, Open PHACTS, BAO
- Shared goal: biomedical knowledge discovery ecosystem
- Phenotypic knowledge management as new opportunity

## RDF community progress

- PubChem RDF major revision June 2015, REST API
- ChEMBL RDF, ChEMBL Core Ontology, Sparql enpoint
- Aligned efforts leading to greater results.

## Related projects

- BioAssay Ontology (BAO)
- BioAssay Research Database (BARD)
- Illuminating the Druggable Genome (IDG)
  - Heterogeneous knowledge integration
- D2D: NSF SBIR Predictive Phenotypic Profiler

## Open PHACTS integration

- Open PHACTS v2.0 includes OPDDR beta version
- OPDDR revision plan:
  - ChEMBL RDF schema for tighter API integration
  - Phenotypic curation, e.g. cell line associations
  - Additional BAO annotations

### OPDDR RDF Model
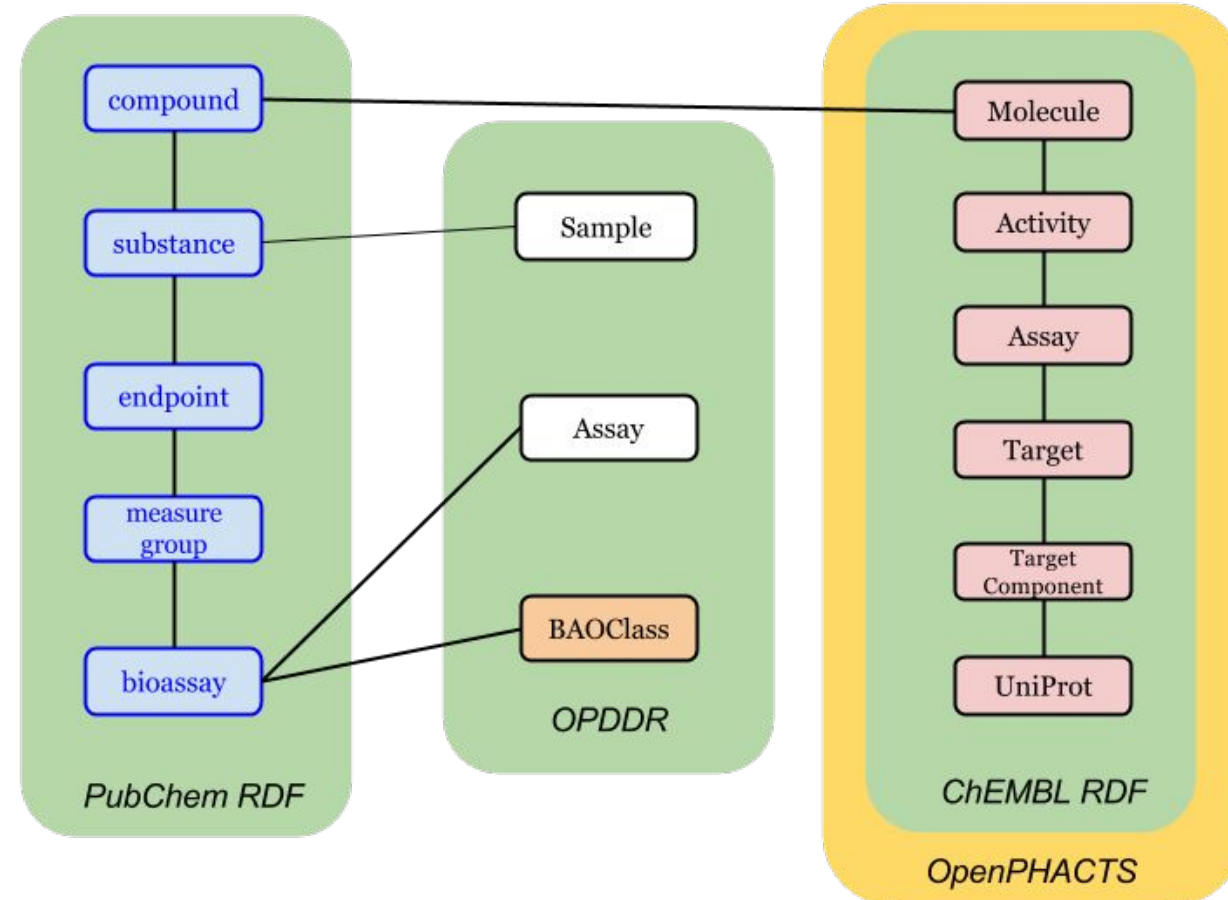


## Applications and use cases

- Semantic assay analytics; finding related data.
- Noise reduction; more data usually allows better sampling.
- Target, MOA deconvolution; interpreting phenotypes.
- Disease relevant lead discovery; diseases as phenotypes.

## Cross-domain query

Find ChEMBL protein kinase targets associated with OIDD Hela cell phenotypic assays via shared active compounds.

```
SELECT DISTINCT
  ?assay ?assayname ?target ?targetname
WHERE
{
  ?substance obo:BFO_0000056 ?measureg .
  ?assay bao:BAO_0000209 ?measureg .
  ?measureg obo:OBI_0000299 ?endpoint .
  ?endpoint obo:IAO_0000136 ?substance .

  ?substance skos:exactMatch ?mol .
  FILTER(REGEX(?assayname, "Hela Cell","i")) .
  ?assay dcterms:title ?assayname .
  ?mol cco:hasActivity ?activity .
  ?chembl_assay cco
  ?target cco:hasAs
```

```
------------------------------------------------------
|  oidd_assay        |  oidd_assayname
======================================================
|  bioassay:AID1117347  |  "Hela CellCycMod PI Cell Number"
|  bioassay:AID1117347  |  "Hela CellCycMod PI Cell Number"
|  bioassay:AID1117347  |  "Hela CellCycMod PI Cell Number"
```

```
------------------------------------------------------
|  target                              |  targetname
======================================================
|  chembl_target:CHEMBL1075034  |  "Thymidine kinase"
|  chembl_target:CHEMBL1075062  |  "Thymidine kinase"
|  chembl_target:CHEMBL1075104  |  "Leucine-rich repeat serine/threonine-protein kin
|  chembl_target:CHEMBL1075115  |  "Dual specificity tyrosine-phosphorylation-regula
|  chembl_target:CHEMBL1075133  |  "G protein-coupled receptor 7"
|  chembl_target:CHEMBL1075155  |  "Serine/threonine-protein kinase 38"
|  chembl_target:CHEMBL1075167  |  "Homeodomain-interacting protein kinase 4"
|  chembl_target:CHEMBL1075189  |  "Pyruvate kinase isozymes M1/M2"
```