CDD BioAssay Express

Expanding the target dimension: How to visualize a lot of SAR

Alex M. Clark

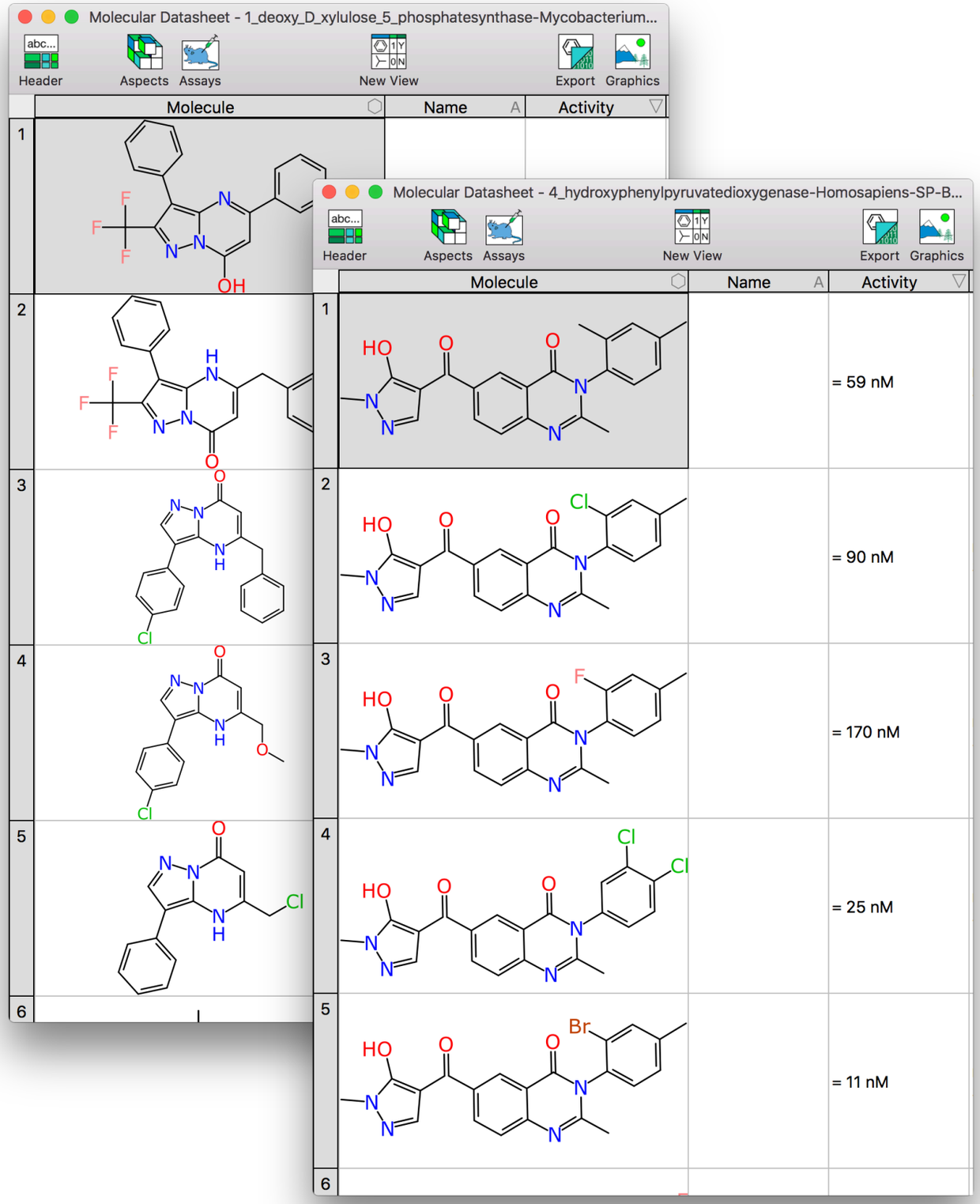
http://**www.bioassayexpress.com**

⚡

COLLABORATIVE DRUG DISCOVERY

Protocols & Datasets

- Structure-activity data usually exists in isolation:
  - one **target**, one **protocol**
  - many **compounds** & **measurements**
- Datasets often related:
  - same target?
  - same protocol?
- *How would we know?*

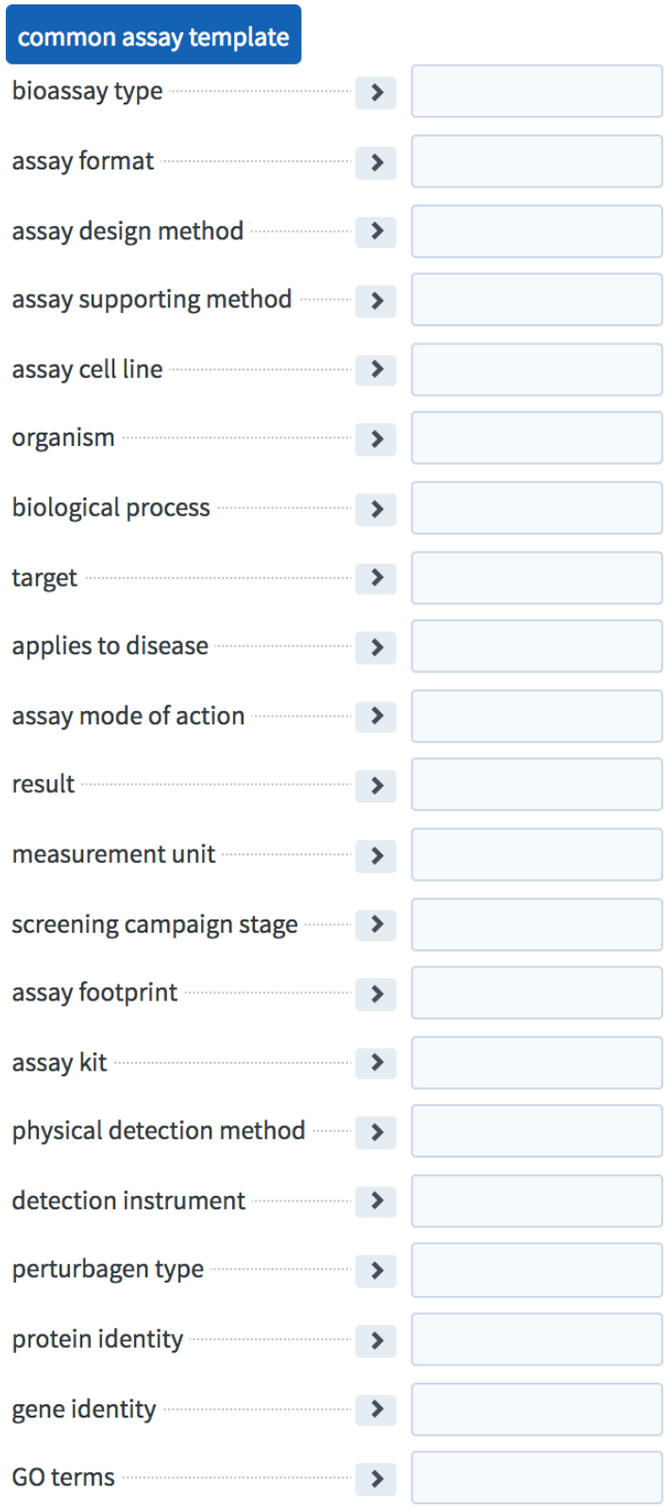


Prerequisite

- Protocols normally described using scientific English
- Alternatively, many ontologies are available:
  - **BAO** (BioAssay Ontology)
  - **DTO** (Drug Target Ontology)
  - **CLO** (Cell Line Ontology)
  - **GO** (Gene Ontology)
  - ... and many others
- Each term has a URI: global meaning, hierarchical

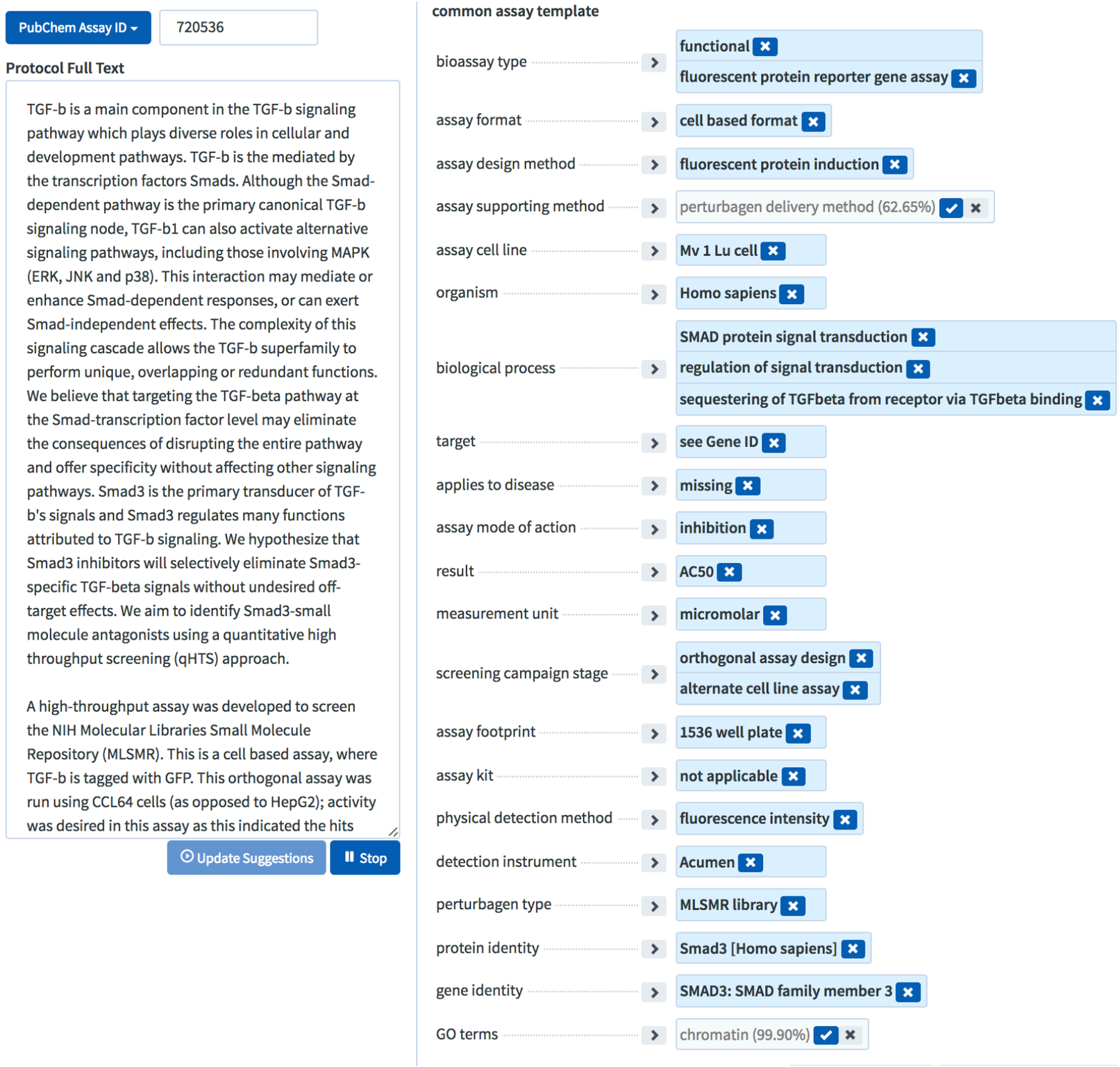
Common Assay Template

- Pulls together ontology categories
- ~22 assignments
- Captures key summary information



Common Assay Template

- Pulls together ontology categories
- ~22 assignments
- Captures key summary information



Protocol Similarity

504854

A Cell Based Secondary Assay To Explore Vero Cell Cytotoxicity of Purified and Synthesized Compounds that Inhibit Mycobacterium Tuberculosis (4)

This functional assay was developed for detection of compounds inhibiting Vero E6 cells viability as a secondary screen to the beta-lactam sensitizing M. tuberculosis bacteriocidal assay.

In this assay, we treated Vero E6 cells with compounds selected as hits in the M. tuberculosis assay for 72 hours over a 10 point 2-fold dilution series, ranging from 0.195 uM to 100 uM. Following 72 hours of treatment, relative viable cell number was determined using Cell Titer Glo from Promega. Each plate contained 64 replicates of vehicle treated cells which served as negative controls.

Outcome: Compounds that showed <70% cell viability for at least one concentration were defined as "Active". If the % viability at all doses was >70%, the compound was defined as "Inactive".

...

69%
similar (?)

449764

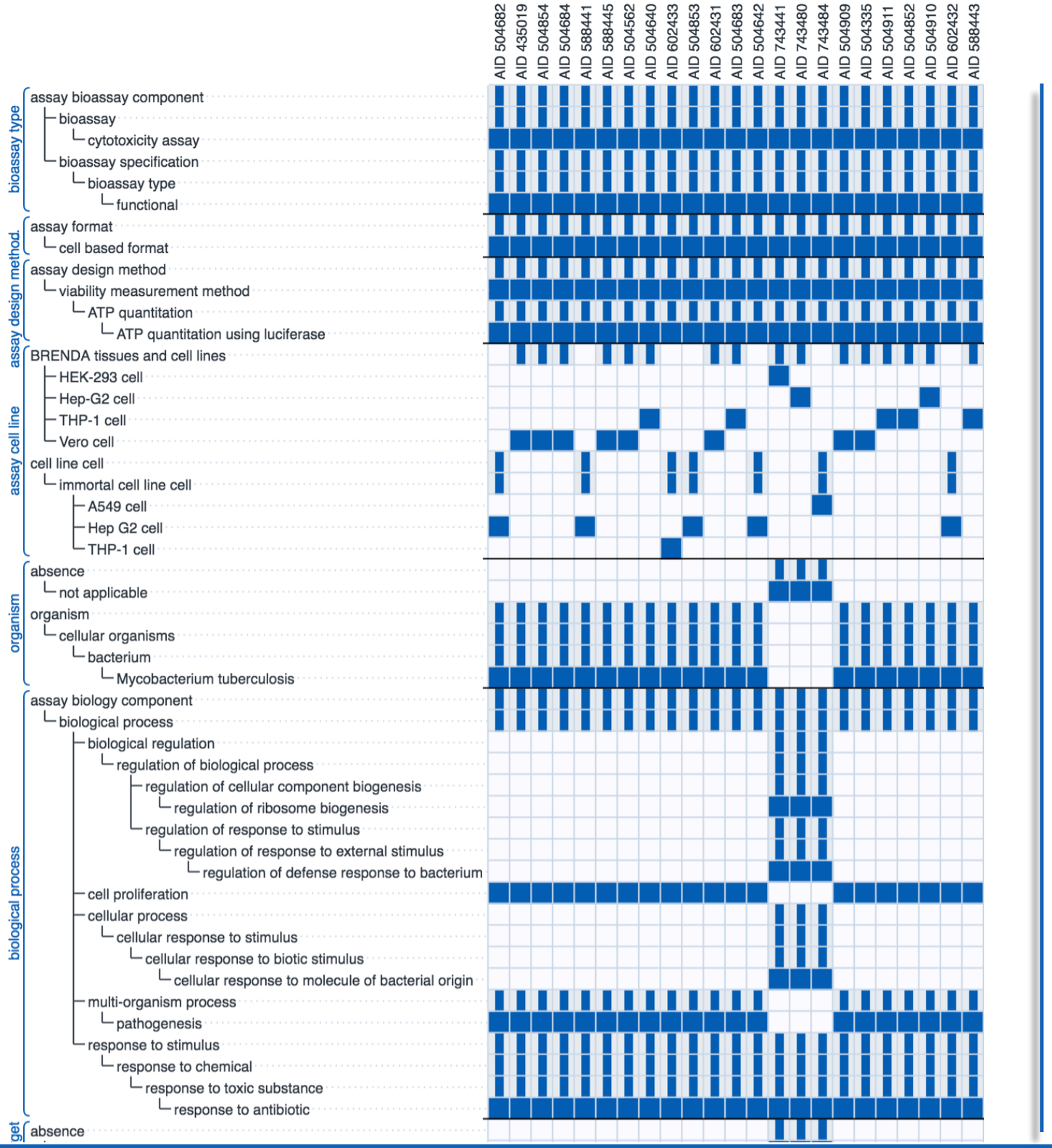
A High Throughput Confirmatory Assay used to Identify Novel Compounds that Inhibit Mycobacterium Tuberculosis in the absence of Glycerol

Outcome: Compounds that showed >30% inhibition for at least one concentration in the Mtb dose response were defined as "Active". If the inhibition at all doses was <30% in the Mtb assay, the compound was defined as "Inactive". In the primary screen a compound was deemed "Inactive" if it had a Percent Inhibition <70.31%. Compounds with a Percent Inhibition >70.31% but were not selected for follow up dose response were labeled "Inconclusive."

The following tiered system has been implemented at Southern Research Institute for use with the PubChem Score. Compounds in the primary screen are scored on a scale of 0-40 based on inhibitory activity where a score of 40 corresponds to 100% inhibition. In the confirmatory dose response screen, active compounds were scored on a scale of 41-80 based on the IC50 result in the Mtb assay while compounds that did not confirm as actives were given the score 0.

...

Protocol Similarity



semantic annotations

≡

fingerprints

Matrix: Assays vs. Compounds

- Curated ~3500 assays from **PubChem**:
  - mostly from *Molecular Libraries* project
  - natural language/machine learning helped
  - carefully analyzed by expert biologists
- Enough data for proof of concept *assay informatics*
- Adding the **assay** dimension to **structure-activity relationship** analysis

Now With Compounds

- Each assay has compounds & data from **PubChem**:

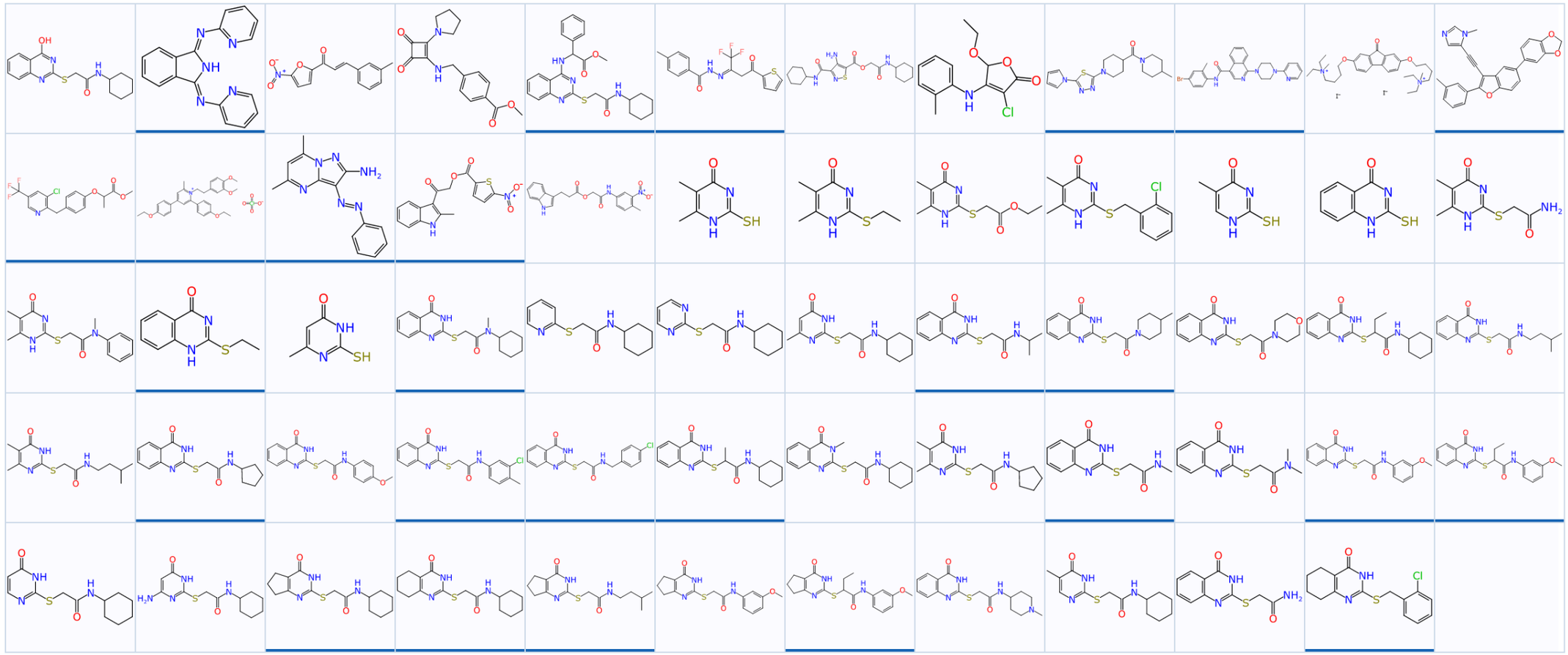
Assay ID: *4011*, PubChem Assay ID: 504854, PubChem Source: Southern Research Specialized Biocontainment Screening Center

PubChem Cross References

Taxonomy	https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=9534
Other	dburl: http://www.southernresearch.org

+ Annotation History

1 of 59



Download as SDfile

Now With Compounds

- Each assay has compounds & data from **PubChem**:

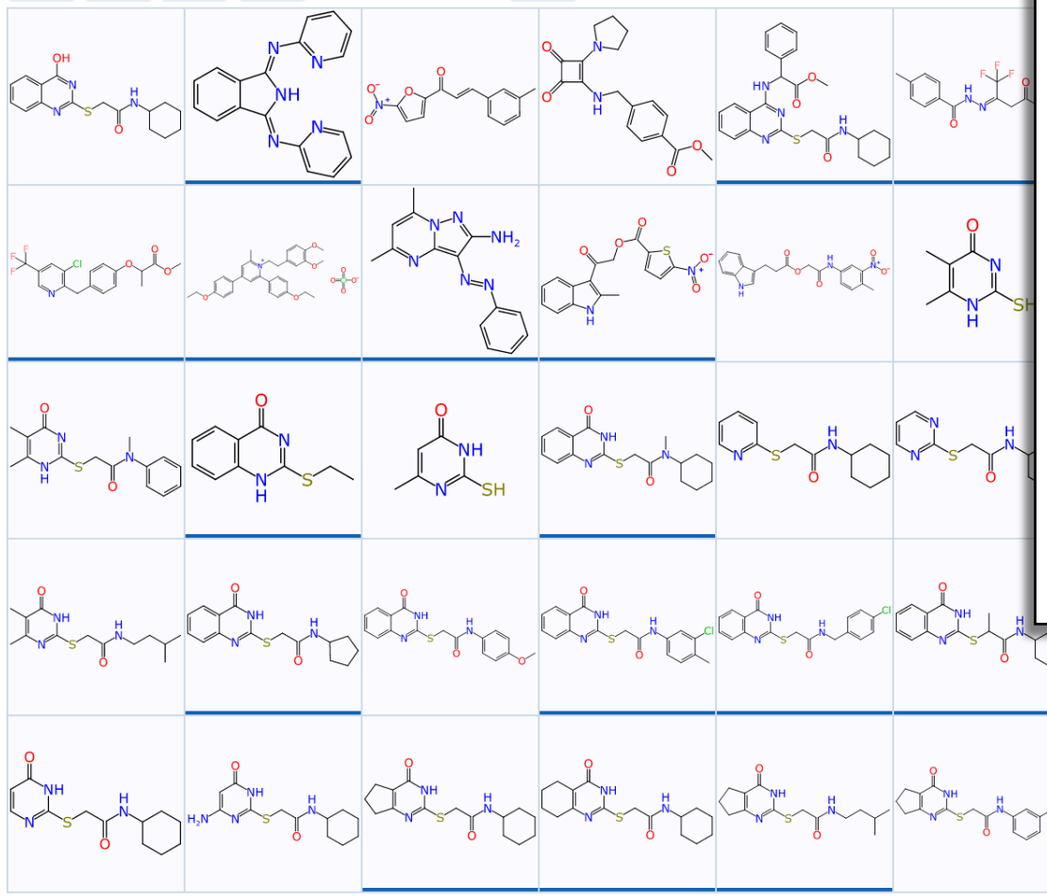
Assay ID: *4011*, PubChem Assay ID: 504854, PubChem Source: Southern Research Speci

PubChem Cross References

Taxonomy	https://www.ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?id=9534
Other	dburl: http://www.southernresearch.org

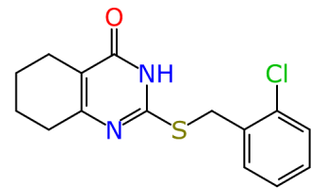
+ Annotation History

1 of 59



Download as SDfile

Compound Measurements — Done



PubChem origins: compound CID 1480154 substance SID 121270649

Name	Type	Relation	Value	Units
Active/Inactive	activity	=	1	binary
Average CC50	measurement	=	100	uM
CC50 Hill Slope Rep 1	measurement	=	-1.05	
CC50 Hill Slope Rep 2	measurement	=	-0.87	
CC50 NormChi2 Rep 1	measurement	=	5.72	
CC50 NormChi2 Rep 2	measurement	=	6.04	
CC50 Rep 1	measurement	=	100	uM
CC50 Rep 2	measurement	=	100	uM
CC50 Std Dev Rep 1	measurement	=	107.34	
CC50 Std Dev Rep 2	measurement	=	307.18	
Cell Viability @ 0.195 uM Rep 1	measurement	=	101.34	%
Cell Viability @ 0.195 uM Rep 2	measurement	=	106.34	%
Cell Viability @ 0.195 uM Rep 3	measurement	=	99.65	%
Cell Viability @ 0.391 uM Rep 1	measurement	=	113.19	%

Now With Compounds

- Each assay has compounds & data from **PubChem**:

Measurement Tagging

Done

Available Measurements

Assay #1

Active/Inactive	binary	activity	IsActive
Average CC50	uM	measurement	CC50_uM
CC50 Hill Slope Rep 1		measurement	
CC50 Hill Slope Rep 2		measurement	
CC50 Hill Slope Rep 3		measurement	
CC50 NormChi2 Rep 1		measurement	
CC50 NormChi2 Rep 2		measurement	
CC50 NormChi2 Rep 3		measurement	
CC50 Rep 1	uM	measurement	
CC50 Rep 2	uM	measurement	
CC50 Rep 3	uM	measurement	
CC50 Std Dev Rep 1		measurement	
CC50 Std Dev Rep 2		measurement	
CC50 Std Dev Rep 3		measurement	
Cell Viability @ 0.195 uM Rep 1	%	measurement	
Cell Viability @ 0.195 uM Rep 2	%	measurement	
Cell Viability @ 0.195 uM Rep 3	%	measurement	
Cell Viability @ 0.391 uM Rep 1	%	measurement	

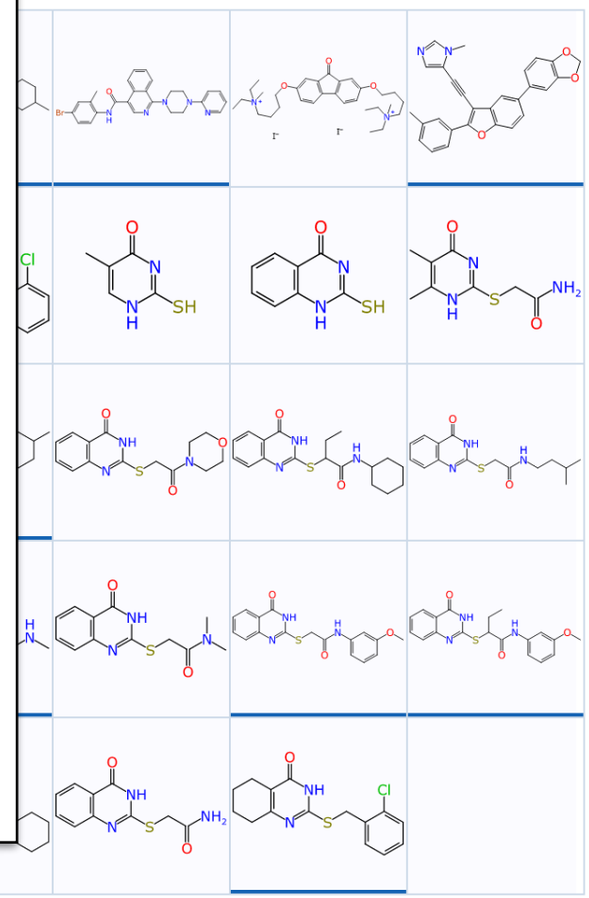
Output Columns

CC50_uM

Average CC50 (measurement, uM)

IsActive

Active/Inactive (activity, binary)



Download as SDfile

Looking at Primary Screens

- *Browse* the curated content from **PubChem**:

Looking at Primary Screens

- *Browse* the curated content from **PubChem**:

PRIMARY **screening campaign stage**

absence
- → **not applicable** (8)

assay screening campaign stage
- → **confirmatory assay** (1106)
- → **lead optimization assay** (269)
- → **preclinical development stage** (5)
- → **primary assay** (349)
- → **secondary assay** (459)
  - → **alternate confirmatory assay** (134)
    - → **alternate assay conditions** (81)
    - → **alternate assay format** (55)
    - → **alternate assay type** (109)
    - → **orthogonal assay design** (107)
    - → **orthogonal assay technology** (33)
  - → **counter screening assay** (693)
    - → **alternate target assay** (250)
    - → **compound toxicity assay** (345)
    - → **parental cell line assay** (20)
    - → **physicochemical profiling assay** (44)
      - → **compound aggregation assay** (3)
      - → **compound fluorescence assay** (27)
      - → **compound redox activity assay** (1)
    - → **variant construct assay** (68)
  - → **selectivity assay** (229)
    - → **alternate cell line assay** (44)
    - → **alternate organism assay** (16)
- → **summary assay** (1)

Looking at Primary Screens

- *Browse* the curated content from **PubChem**:

PRIMARY **screening campaign stage**

- absence
  - → **not applicable** (8)
- assay screening campaign stage
  - → **confirmatory assay** (1106)
  - → **lead optimization assay** (269)
  - → **preclinical development stage** (5)
  - → **primary assay** (349)
  - → **secondary assay** (459)
    - → **alternate confirmatory assay** (134)
      - → **alternate assay conditions** (81)
      - → **alternate assay format** (55)
      - → **alternate assay type** (109)
      - → **orthogonal assay design** (107)
      - → **orthogonal assay technology** (33)
    - → **counter screening assay** (693)
      - → **alternate target assay** (250)
      - → **compound toxicity assay** (345)
      - → **parental cell line assay** (20)
      - → **physicochemical profiling assay** (44)
        - → **compound aggregation assay** (3)
        - → **compound fluorescence assay** (27)
        - → **compound redox activity assay** (1)
      - → **variant construct assay** (68)
    - → **selectivity assay** (229)
      - → **alternate cell line assay** (44)
      - → **alternate organism assay** (16)
  - → **summary assay** (1)

SECONDARY **organism**

- absence
  - → **unknown** (1 of 7)
- organism
  - → cellular organisms
    - → bacterium
      - → **Bacillus anthracis** (1 of 4)
      - → **Clostridium botulinum** (2 of 21)
      - → **Escherichia coli** (3 of 69)
      - → **Helicobacter pylori** (1 of 10)
      - → **Leuconostoc mesenteroides** (1)
      - → **Mycobacterium tuberculosis** (4 of 54)
      - → **Pseudomonas aeruginosa PAO1** (3 of 58)
      - → **Salmonella enterica subsp. enterica serovar Typhimurium** (1 of 15)
      - → **Vibrio harveyi** (1 of 25)
    - → eukaryote
      - → fungi
        - → **Cryptococcus neoformans** (1 of 17)
        - → multicellular fungi
          - → **Candida albicans** (1 of 50)
        - → unicellular fungi
          - → **Saccharomyces cerevisiae** (18 of 92)
      - → metazoa
        - → invertebrate
          - → **Caenorhabditis elegans** (2 of 25)
          - → **Fasciola hepatica** (1 of 3)
          - → **Haemonchus contortus** (1 of 3)
          - → **Photinus pyralis** (1 of 17)
        - → vertebrate
          - → mammalian
            - → **Bos taurus** (3 of 7)
            - → **Canis lupus familiaris** (1 of 11)
            - → **Homo sapiens** (255 of 2295)
            - → **Mus musculus** (18 of 285)
            - → **Rattus norvegicus** (7 of 89)
          - → non-mammalian
            - → **Danio rerio** (1 of 5)

Matching Assays

Matching assays: 255

All | None | Show Compounds | Assay Grid | Copy Results | Query

Compounds	Result	PubChem
16000	Primary Cell Based High Throughput Screening Assay for Agonists of GALR2	View AID833
150839	qHTS Assay for Inhibitors of HPGD (15-Hydroxyprostaglandin Dehydrogenase)	View AID894
211511	qHTS Assay for Enhancers of SMN2 Splice Variant Expression	View AID1458
194920	Multiplexed high-throughput screen for small molecule regulators of Bcl-2 family protein interactions, specifically Bim-Bcl-XL.	View AID1007
194920	Multiplexed high-throughput screen for small molecule regulators of Bcl-2 family protein interactions, specifically Bim-Bcl-W.	View AID952
194920	Multiplexed high-throughput screen for small molecule regulators of Bcl-2 family protein interactions, specifically Bim-Bcl-B.	View AID951
64925	Primary biochemical high-throughput screening assay for inhibitors of Matrix Metalloproteinase 13 (MMP13) activity	View AID570
302517	MLPCN Platelet Activation -Dense Granule Release	View AID1663
1352	qHTS Assay for Inhibitors of CDC-like Kinase 4 (ADP-Glo Assay)	View AID1771
1352	qHTS Assay for Inhibitors of CDC-like Kinase 4 (Kinase-Glo Assay)	View AID1770
362388	uHTS identification of CXCR6 Inhibitors in a B-arrestin luminescence assay	View AID602244
8590	Inhibitors of USP1/UAF1: Pilot qHTS	View AID504865
345034	A quantitative high throughput screen for small molecules that induce DNA re-replication in MCF 10a normal breast cells.	View AID624296
344074	A quantitative high throughput screen for small molecules that induce DNA re-replication in SW480 colon adenocarcinoma cells.	View AID624297
9984	VCAM-1 Imaging Assay in Pooled HUVECs: Inhibition of TNFa induced VCAM-1 cell surface expression	View AID456
9984	VCAM-1 Imaging Assay in Pooled HUVECs: Augmentation of TNFa induced VCAM-1 cell surface expression.	View AID457
303344	MLPCN Ras selective lethality-BJeLR viability	View AID1554
104742	Inhibitors of Cav3 T-type Calcium Channels: Primary Screen	View AID449739
20540	HTS colorimetric detection of phosphate released in TNAP reaction	View AID614
20540	HTS colorimetric detection of p-nitrophenol released in TNAP reaction	View AID615
64394	HTS discovery of chemical inhibitors of anti-apoptotic protein Bfl-1	View AID432
282587	qHTS for Inhibitors of the Interaction of Thyroid Hormone Receptor and Steroid Receptor Coregulator 2	View AID1469
96416	High Throughput Fluorescence Polarization Screen for Bcl-B Phenotype Converters	View AID748
12344	TNFalpha Induced E-Selectin Expression - Primary screen	View AID487
218117	Primary cell-based high-throughput screening assay to identify agonists of the transient receptor potential channel ML3 (TRPML3)	View AID1448
221370	qHTS Assay for Antagonists of the Neuropeptide S Receptor: cAMP Signal Transduction	View AID1461
16000	Luminescence-based primary cell-based high throughput screening assay to identify inhibitors of NADPH oxidase 1 (Nox1): Maybridge Library	View AID1792
315101	Fluorescence-based primary biochemical high throughput screening assay to identify inhibitors of Protein Phosphatase 5 (PP5).	View AID1987
291075	uHTS HTRF assay for identification of inhibitors of SUMOylation	View AID2006
291075	uHTS fluorescence polarization assay for the identification of translation initiation inhibitors (eIF4H)	View AID2012
315100	Fluorescence polarization-based primary biochemical high throughput screening assay to identify inhibitors of myeloid cell leukemia sequence 1 (MCL1) interactions with BIM BH3 peptide	View AID2057

Matching Assays

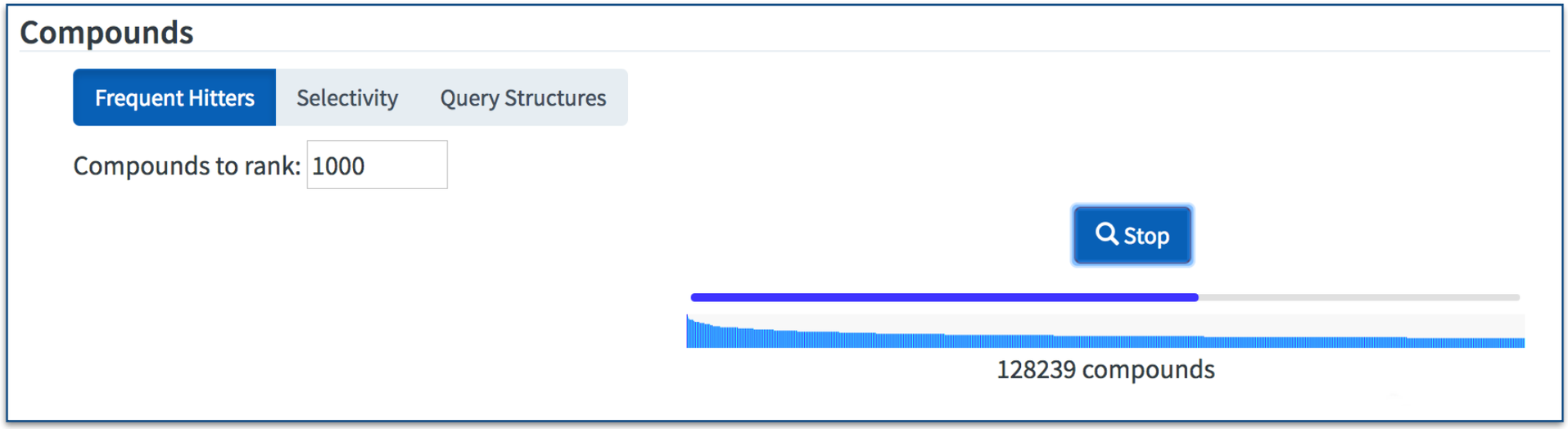
Matching assays: 255

All | None | Show Compound | Assay Grid | Copy Results | Query

Compounds	Result	PubChem
16000	Primary Cell Based High Throughput Screening Assay for Agonists of GALR2	View AID833
150839	qHTS Assay for Inhibitors of HPGD (15-Hydroxyprostaglandin Dehydrogenase)	View AID894
211511	qHTS Assay for Enhancers of SMN2 Splice Variant Expression	View AID1458
194920	Multiplexed high-throughput screen for small molecule regulators of Bcl-2 family protein interactions, specifically Bim-Bcl-XL.	View AID1007
194920	Multiplexed high-throughput screen for small molecule regulators of Bcl-2 family protein interactions, specifically Bim-Bcl-W.	View AID952
194920	Multiplexed high-throughput screen for small molecule regulators of Bcl-2 family protein interactions, specifically Bim-Bcl-B.	View AID951
64925	Primary biochemical high-throughput screening assay for inhibitors of Matrix Metalloproteinase 13 (MMP13) activity	View AID570
302517	MLPCN Platelet Activation -Dense Granule Release	View AID1663
1352	qHTS Assay for Inhibitors of CDC-like Kinase 4 (ADP-Glo Assay)	View AID1771
1352	qHTS Assay for Inhibitors of CDC-like Kinase 4 (Kinase-Glo Assay)	View AID1770
362388	uHTS identification of CXCR6 Inhibitors in a B-arrestin luminescence assay	View AID602244
8590	Inhibitors of USP1/UAF1: Pilot qHTS	View AID504865
345034	A quantitative high throughput screen for small molecules that induce DNA re-replication in MCF 10a normal breast cells.	View AID624296
344074	A quantitative high throughput screen for small molecules that induce DNA re-replication in SW480 colon adenocarcinoma cells.	View AID624297
9984	VCAM-1 Imaging Assay in Pooled HUVECs: Inhibition of TNFa induced VCAM-1 cell surface expression	View AID456
9984	VCAM-1 Imaging Assay in Pooled HUVECs: Augmentation of TNFa induced VCAM-1 cell surface expression.	View AID457
303344	MLPCN Ras selective lethality-BJeLR viability	View AID1554
104742	Inhibitors of Cav3 T-type Calcium Channels: Primary Screen	View AID449739
20540	HTS colorimetric detection of phosphate released in TNAP reaction	View AID614
20540	HTS colorimetric detection of p-nitrophenol released in TNAP reaction	View AID615
64394	HTS discovery of chemical inhibitors of anti-apoptotic protein Bfl-1	View AID432
282587	qHTS for Inhibitors of the Interaction of Thyroid Hormone Receptor and Steroid Receptor Coregulator 2	View AID1469
96416	High Throughput Fluorescence Polarization Screen for Bcl-B Phenotype Converters	View AID748
12344	TNFalpha Induced E-Selectin Expression - Primary screen	View AID487
218117	Primary cell-based high-throughput screening assay to identify agonists of the transient receptor potential channel ML3 (TRPML3)	View AID1448
221370	qHTS Assay for Antagonists of the Neuropeptide S Receptor: cAMP Signal Transduction	View AID1461
16000	Luminescence-based primary cell-based high throughput screening assay to identify inhibitors of NADPH oxidase 1 (Nox1): Maybridge Library	View AID1792
315101	Fluorescence-based primary biochemical high throughput screening assay to identify inhibitors of Protein Phosphatase 5 (PP5).	View AID1987
291075	uHTS HTRF assay for identification of inhibitors of SUMOylation	View AID2006
291075	uHTS fluorescence polarization assay for the identification of translation initiation inhibitors (eIF4H)	View AID2012
315100	Fluorescence polarization-based primary biochemical high throughput screening assay to identify inhibitors of myeloid cell leukemia sequence 1 (MCL1) interactions with BIM BH3 peptide	View AID2057

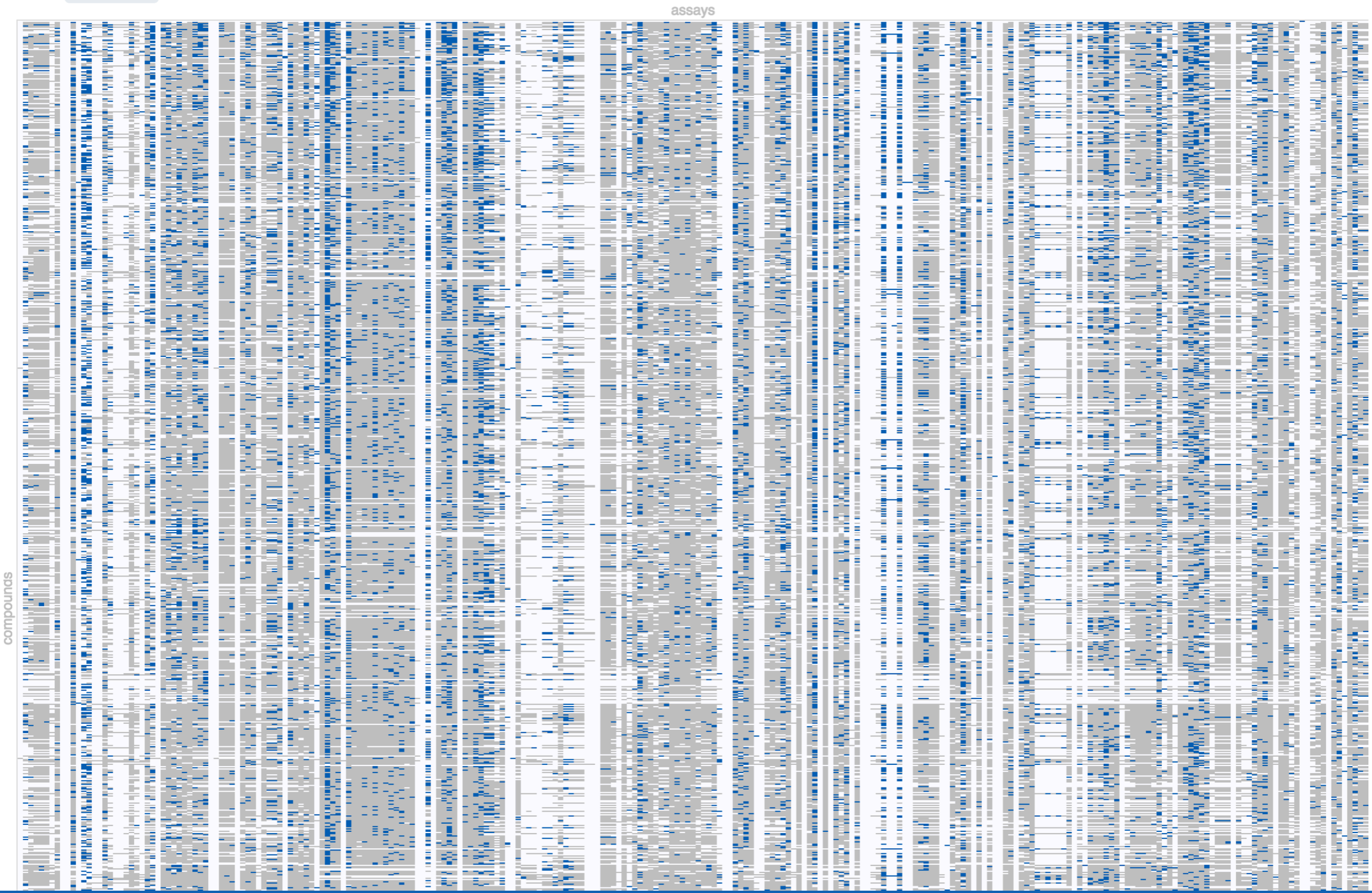
Frequent Hitters

- Rank by **score** = # hits vs. selected assays



Assay Grid (1°)

Assays: Group By

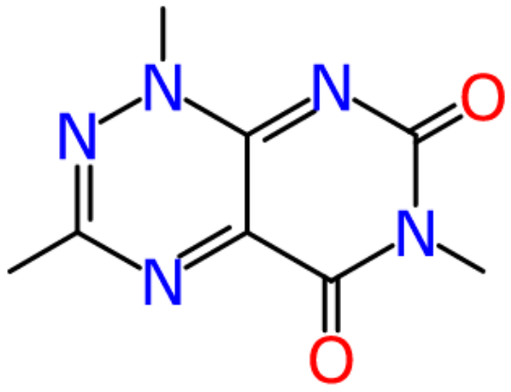


Assay Grid (1°)

Assays: Group By

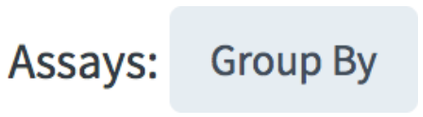
assays

compounds

Assay	MITF Measured in Cell-Based System Using Plate Reader -... View ↘ AID488899
Active	 PubChem compound CID 460747 PubChem substance SID 460747 Active vs. PubChem Assay ID 488899

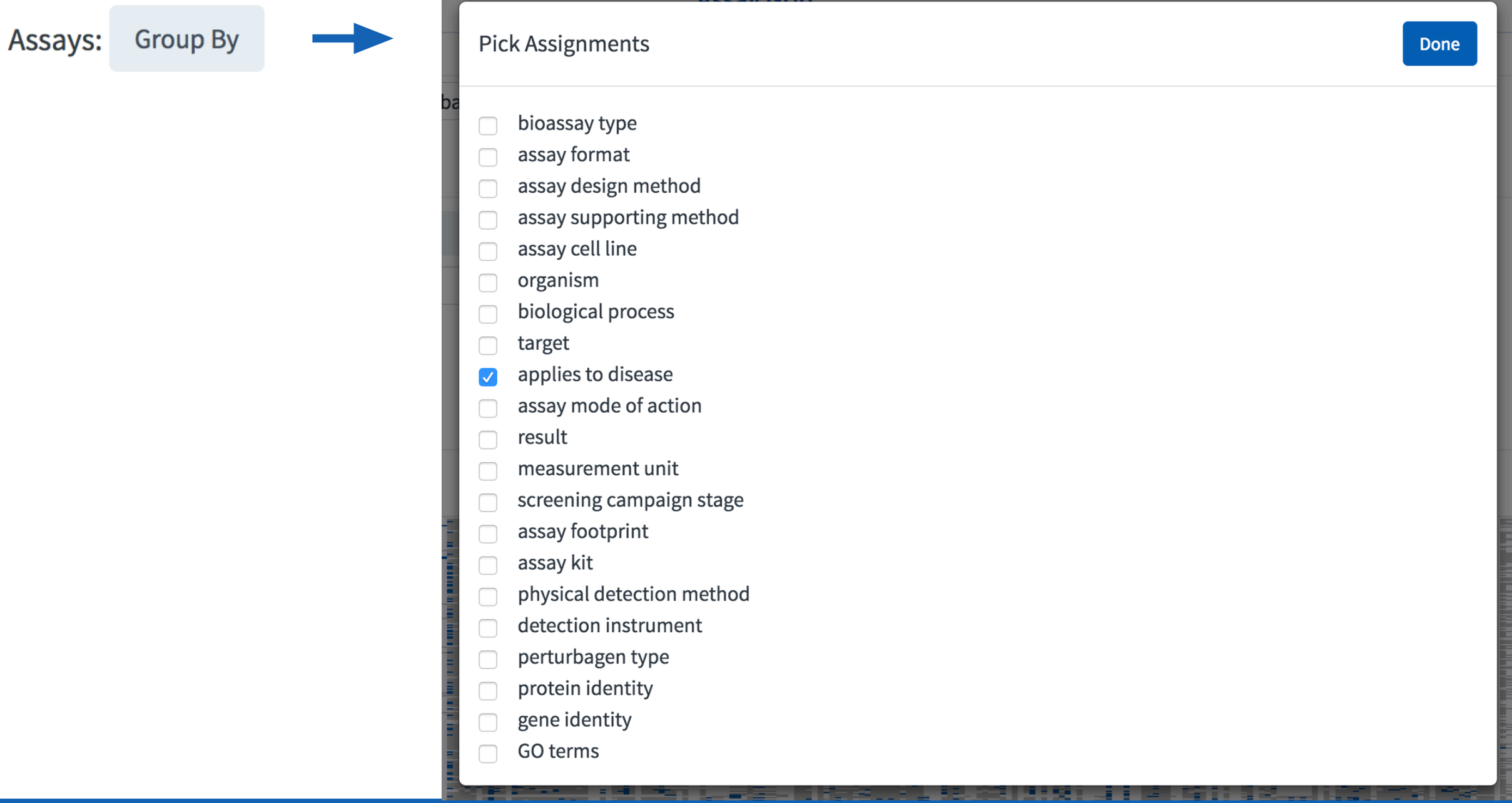
Assay Clustering

- Leverage hierarchical structure of ontology terms:

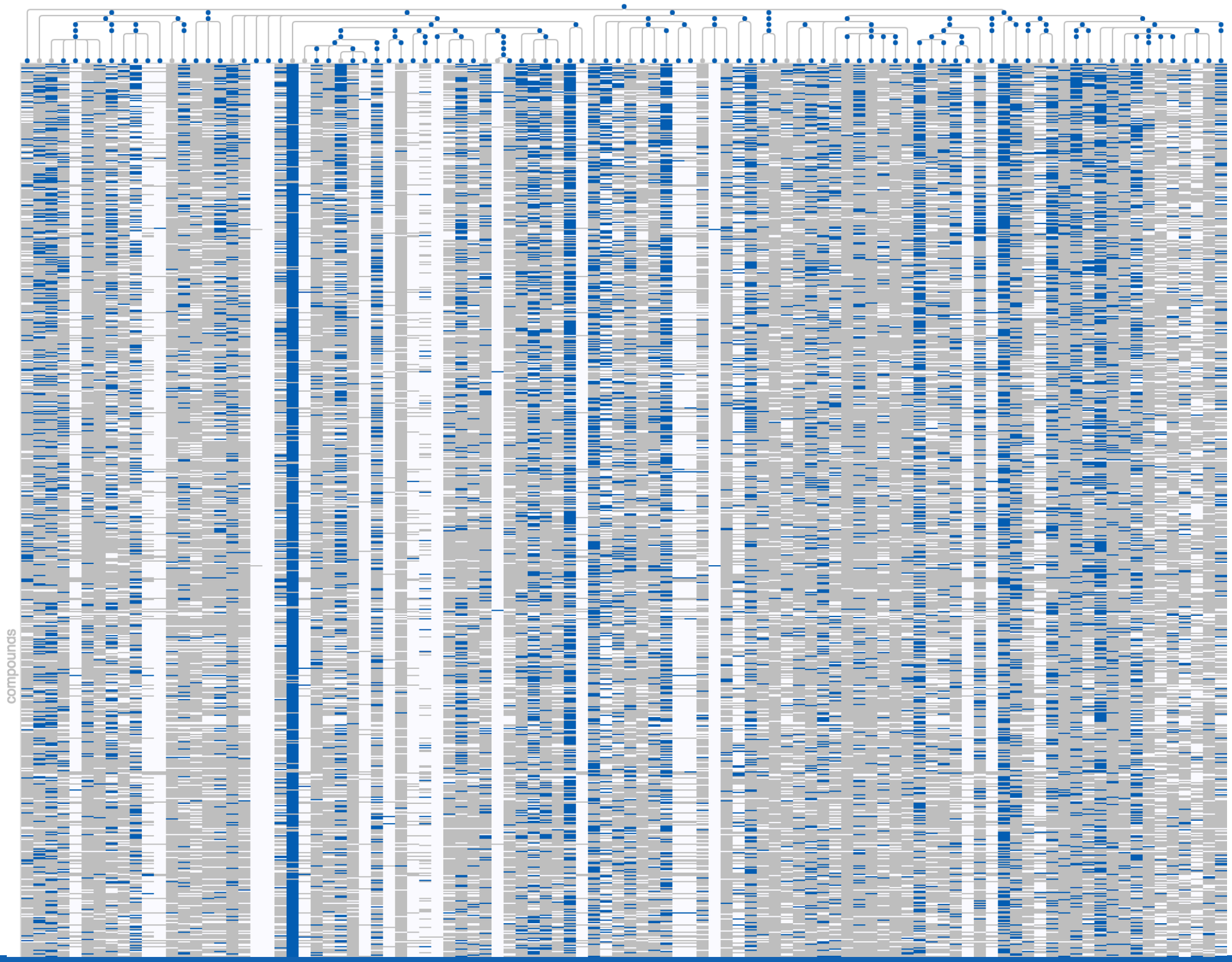


Assay Clustering

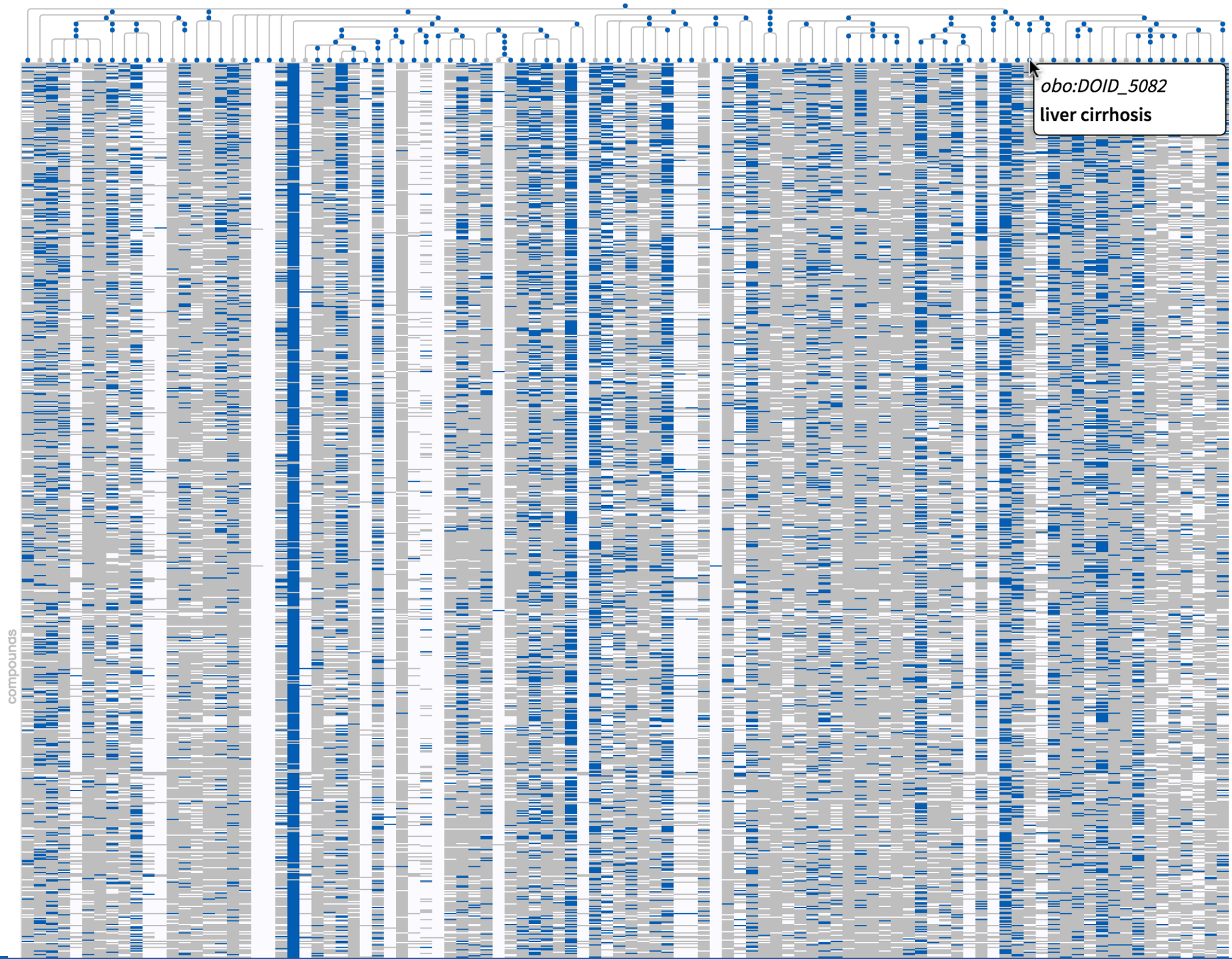
- Leverage hierarchical structure of ontology terms:



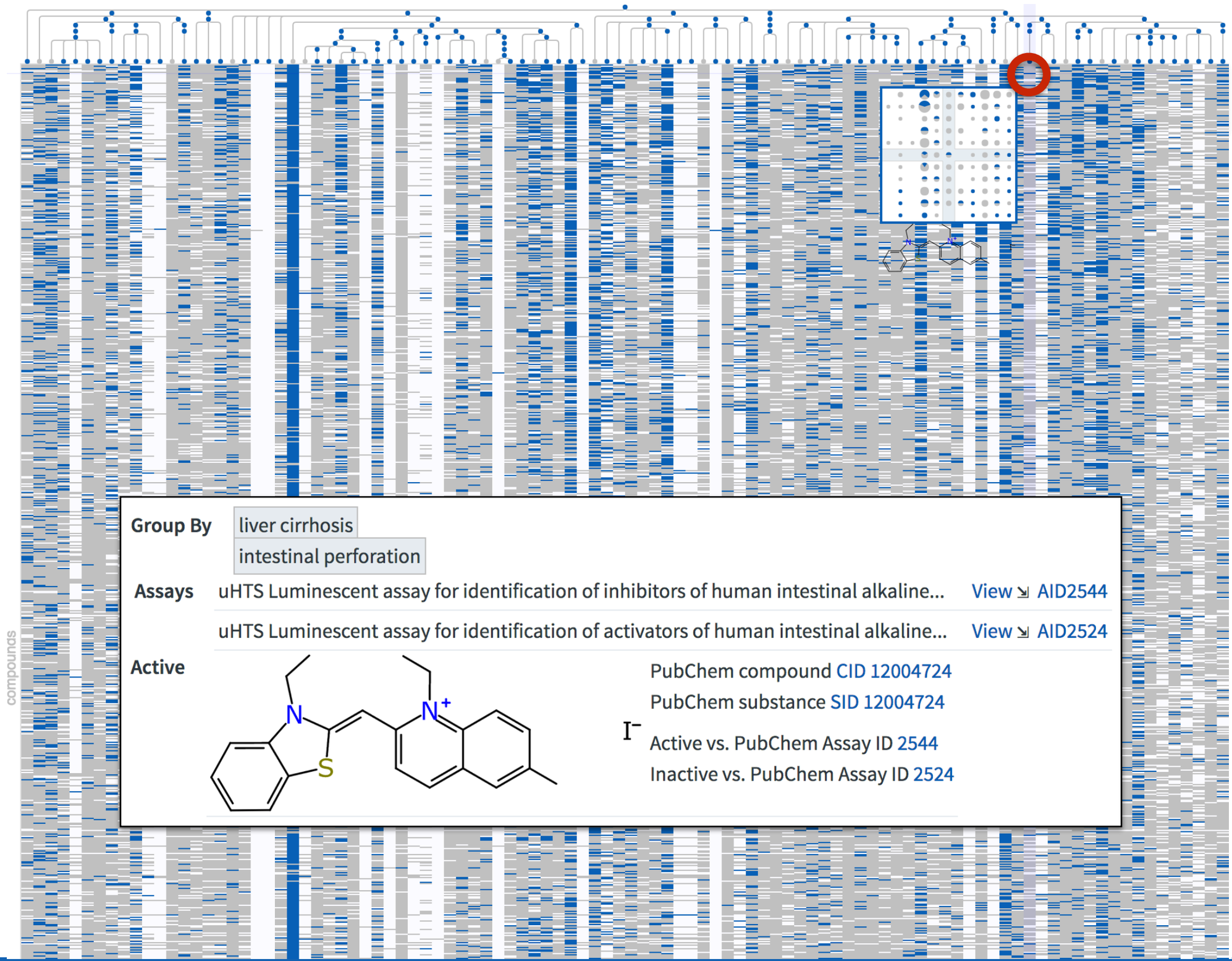
Group by Disease



Group by Disease



Group by Disease



Selectivity

- Using secondary assays
- Ranking score =

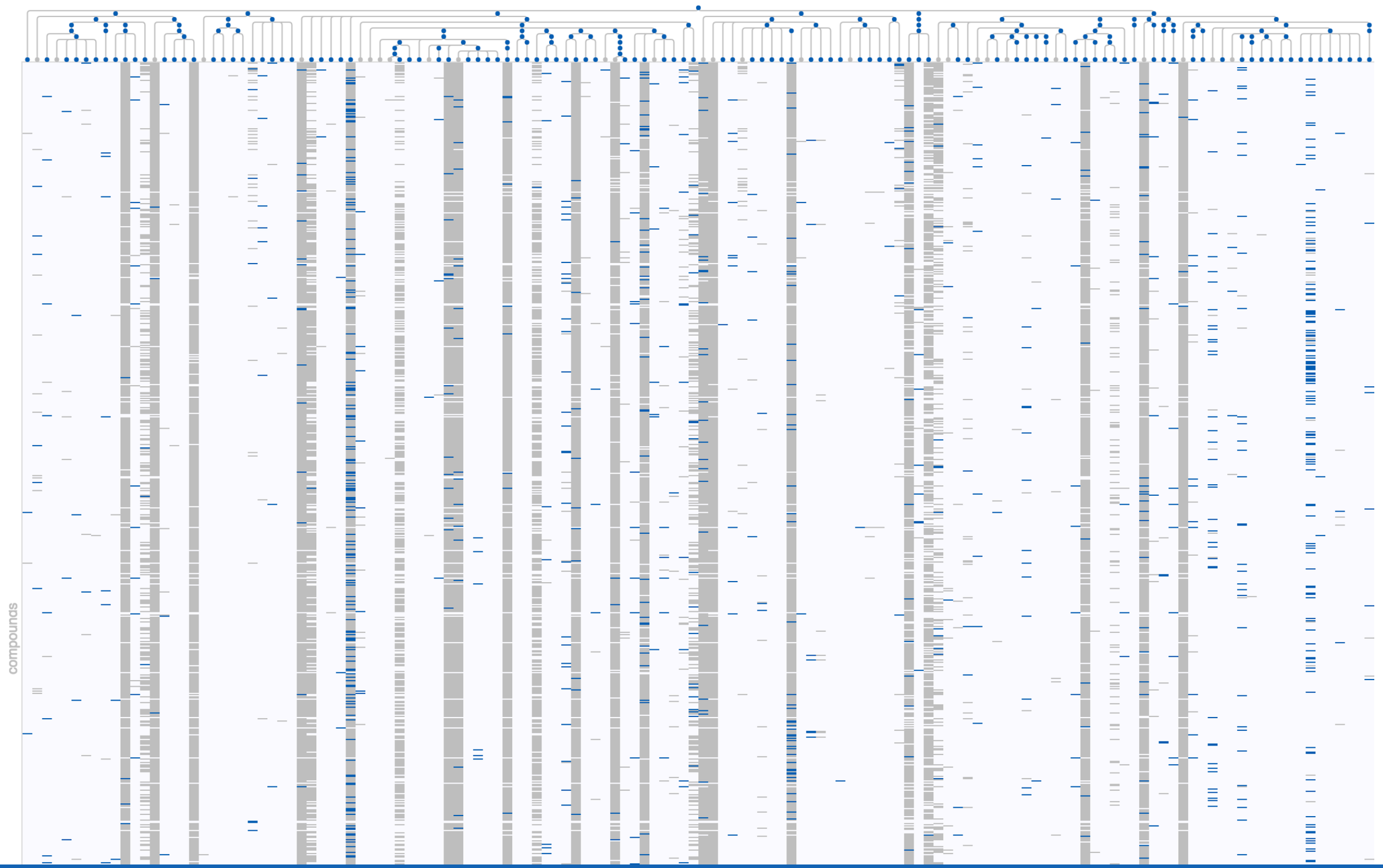
$$\frac{\text{\# inactives}}{\text{\# actives}}$$

PRIMARY **screening campaign stage**

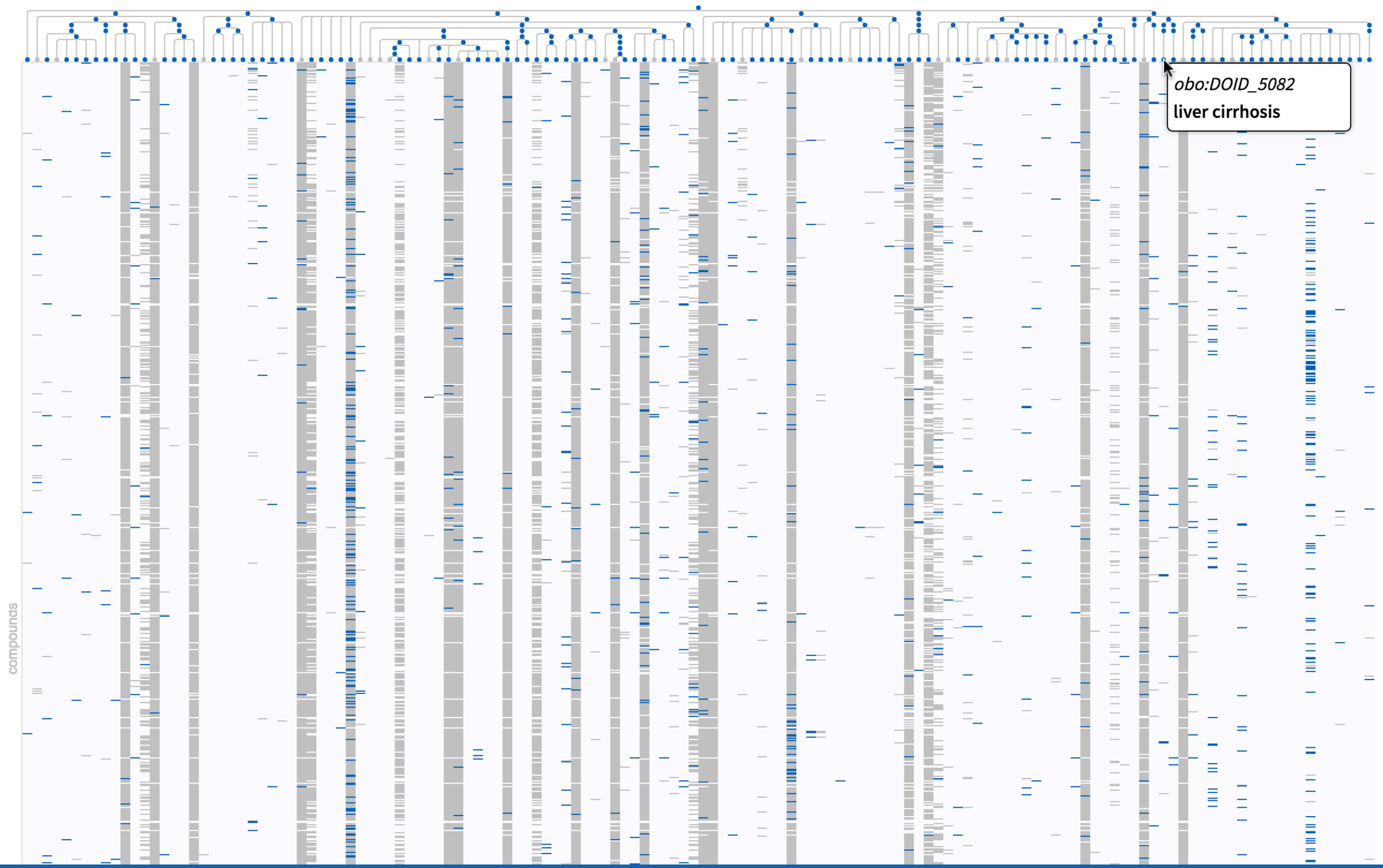
absence
- → **not applicable** (8)

assay screening campaign stage
- → **confirmatory assay** (1106)
- → **lead optimization assay** (269)
- → **preclinical development stage** (5)
- → **primary assay** (349)
- → **secondary assay** (459)
  - → **alternate confirmatory assay** (134)
    - → **alternate assay conditions** (81)
    - → **alternate assay format** (55)
    - → **alternate assay type** (109)
    - → **orthogonal assay design** (107)
    - → **orthogonal assay technology** (33)
  - → **counter screening assay** (693)
    - → **alternate target assay** (250)
    - → **compound toxicity assay** (345)
    - → **parental cell line assay** (20)
    - → **physicochemical profiling assay** (44)
      - → **compound aggregation assay** (3)
      - → **compound fluorescence assay** (27)
      - → **compound redox activity assay** (1)
    - → **variant construct assay** (68)
  - → **selectivity assay** (229)
    - → **alternate cell line assay** (44)
    - → **alternate organism assay** (16)
- → **summary assay** (1)

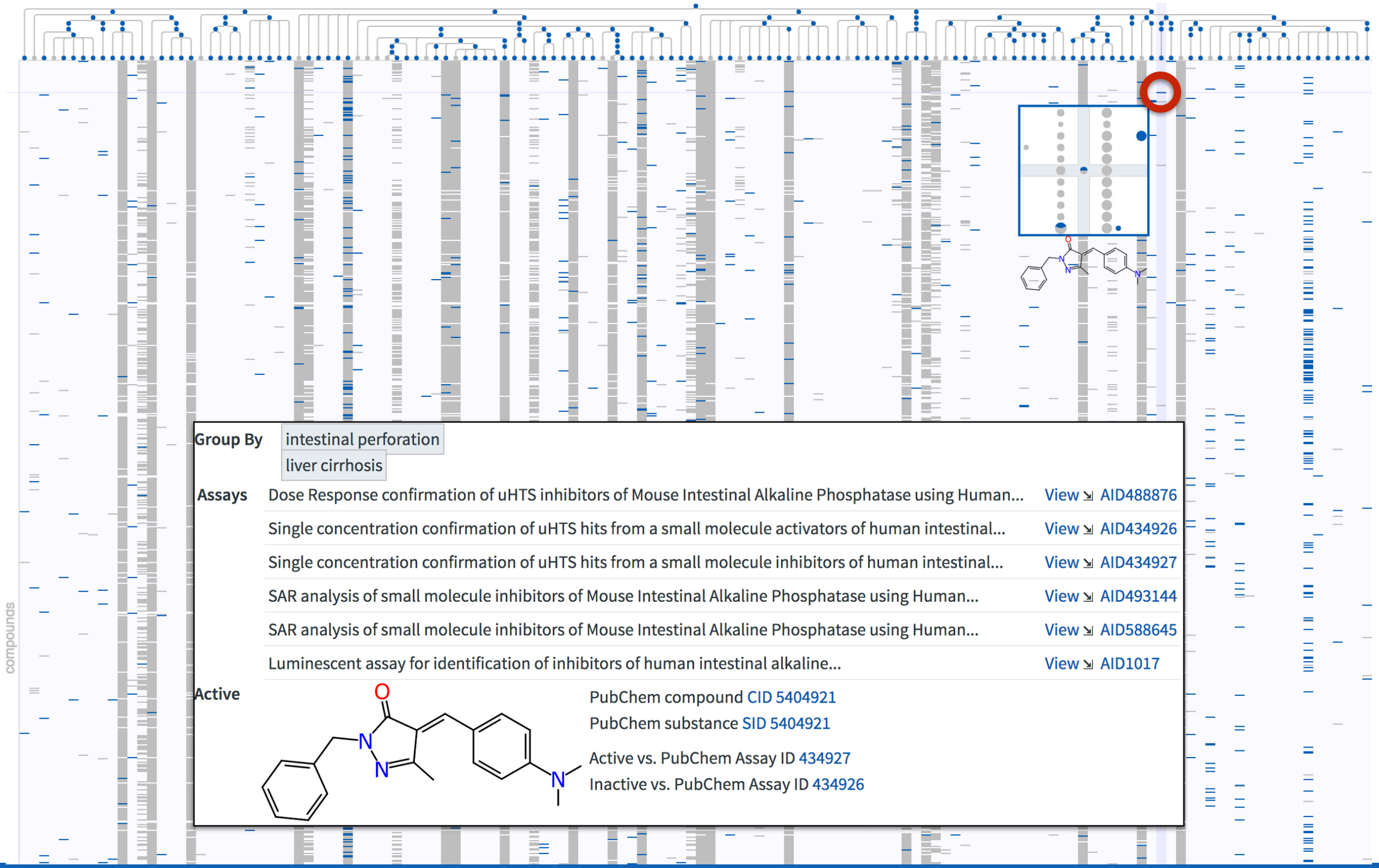
Assay Grid (2°)



Assay Grid (2°)



Assay Grid (2°)


compounds
Group By
intestinal perforation
liver cirrhosis
Assays
Dose Response confirmation of uHTS inhibitors of Mouse Intestinal Alkaline Phosphatase using Human... View AID488876
Single concentration confirmation of uHTS hits from a small molecule activators of human intestinal... View AID434926
Single concentration confirmation of uHTS hits from a small molecule inhibitors of human intestinal... View AID434927
SAR analysis of small molecule inhibitors of Mouse Intestinal Alkaline Phosphatase using Human... View AID493144
SAR analysis of small molecule inhibitors of Mouse Intestinal Alkaline Phosphatase using Human... View AID588645
Luminescent assay for identification of inhibitors of human intestinal alkaline... View AID1017
Active
PubChem compound CID 5404921
PubChem substance SID 5404921
Active vs. PubChem Assay ID 434927
Inactive vs. PubChem Assay ID 434926

Similarity

- Can select compounds by similarity to a reference structure (ECFP6/Tanimoto):

Assays

Query: (bao:BAX_0000002=@bao:BAO_0002989,@bao:BAO_0000590,@bao:BAO_0003011,@b

Search

Assays: **3510**

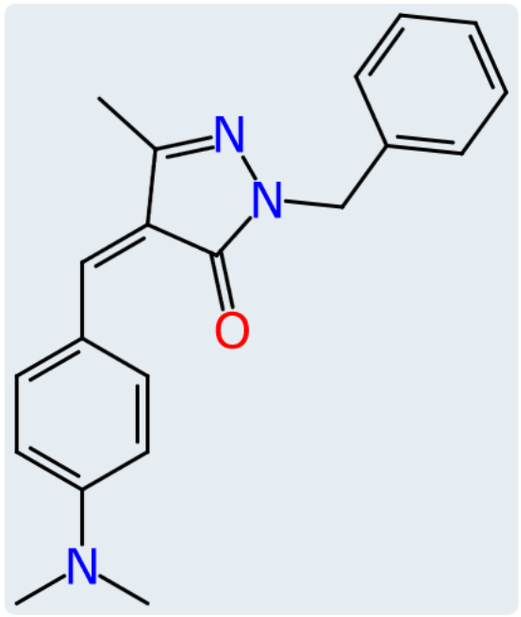
Compounds

Frequent Hitters | Selectivity | Query Structures

Maximum # of compounds: 1000

☐ Only select probe compounds

Rank by similarity to:



Similarity

- Can select compounds by similarity to a reference structure (ECFP6/Tanimoto):

Assays

Query: (bao:BAX_0000002=@bao:BAO_0002989,@bao:BAO_0000590,@bao:BAO_0003011,@b

Search

Assays: **3510**

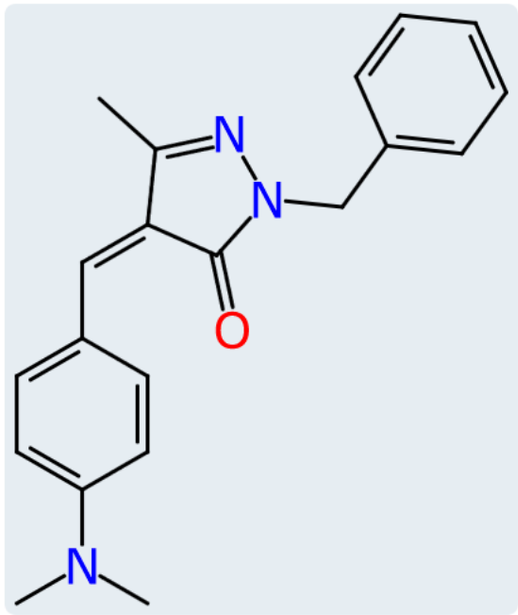
Compounds

Frequent Hitters | Selectivity | Query Structures

Maximum # of compounds: 1000

Only select probe compounds

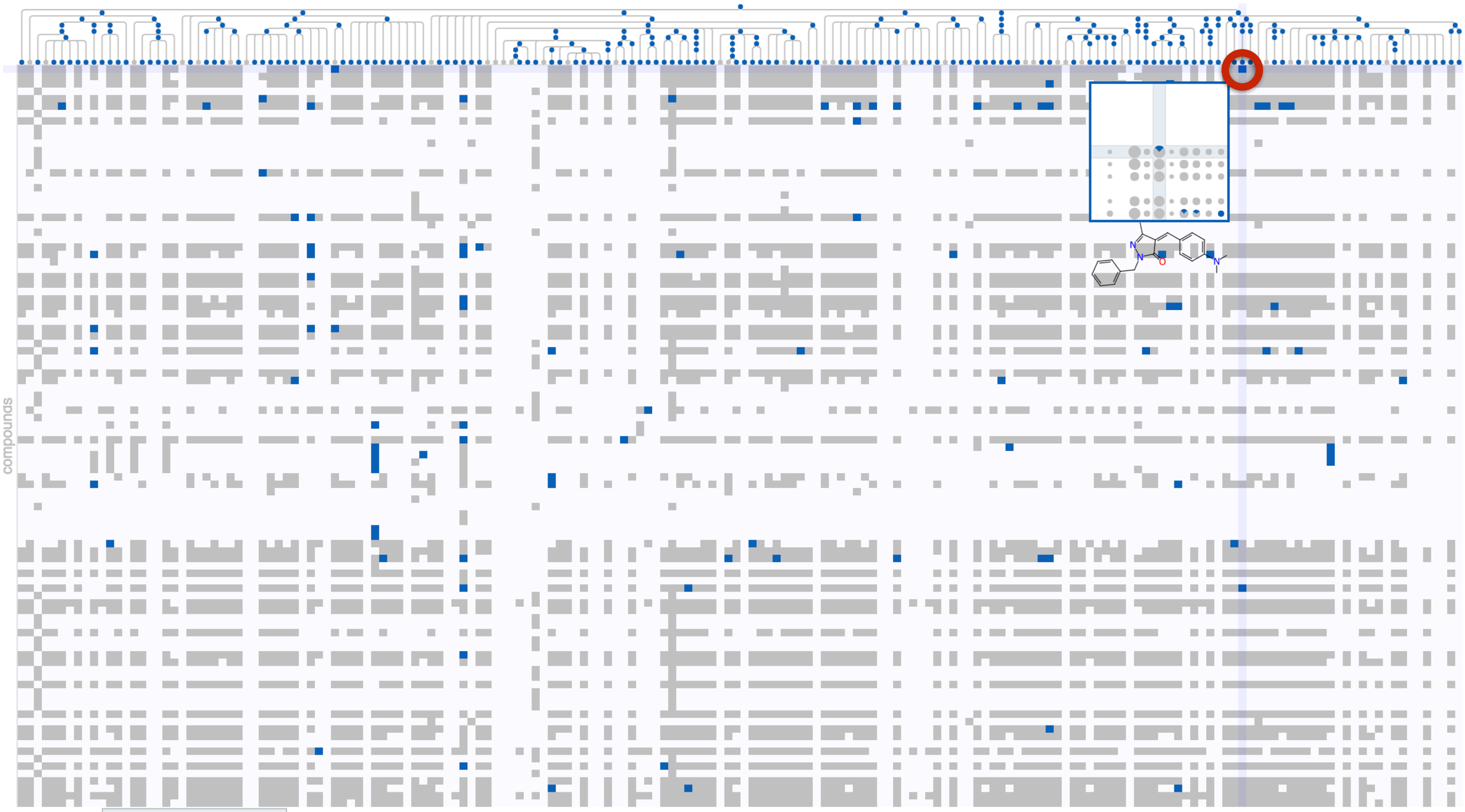
Rank by similarity to:



← paste/drag structure

Assay Grid

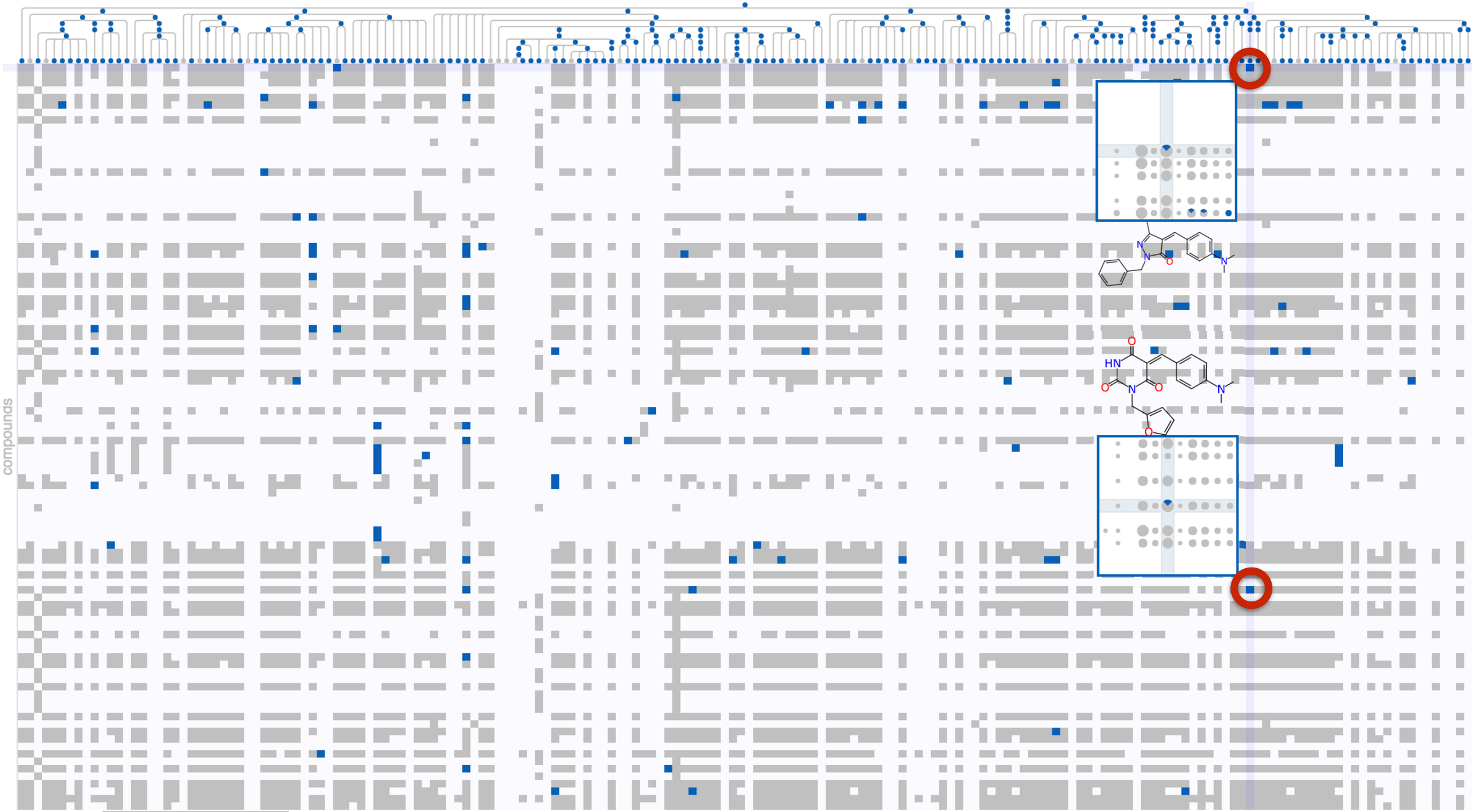
liver cirrhosis



Group By intestinal perforation
liver cirrhosis

Assay Grid

liver cirrhosis



Group By intestinal perforation
liver cirrhosis

Assay Grid

liver cirrhosis

parasitic helminthiasis

compounds

Group By intestinal perforation
liver cirrhosis

Conclusion

- *Cheminformatics* & *bioinformatics* are well established, but **assay informatics** is new
- Data curation has begun: much yet remains (>1M)
- Content **curation** is the main bottleneck
  - *BioAssay Express* is key to solving this problem
- Content **consumption** is at proof of concept stage
  - ... and this is the fun part.

Future Work

- Crowd curation of public data: you should want **your** assays to be machine readable
- Model building: create from structure-activity-assay, apply all existing knowledge to prospective discovery
- ELN context: use semantic annotations *instead of* text, describe assays *comprehensively*
  - convert annotations to text, for communication
  - translate protocols to import formats for robots

Assay Data

- All features, and curated data, available at:
  - http://**www.bioassayexpress.com**
  - http://**beta.bioassayexpress.com**
- Ready for crowd curation - free to use in public
- Commercial product: private installations, with custom settings, are being deployed

Acknowledgments

- Biologists
  - Janice Kranz
  - Haifa Ghandour
  - Karen Featherstone
- Collaborative Drug Discovery
  - Barry Bunin
  - Hande Kücük
  - and the rest of the team

- More information

http://www.bioassayexpress.com

http://github.com/cdd/bioassay-template

http://collaborativedrug.com

PeerJ Comp Sci **2**:e61 (2016)