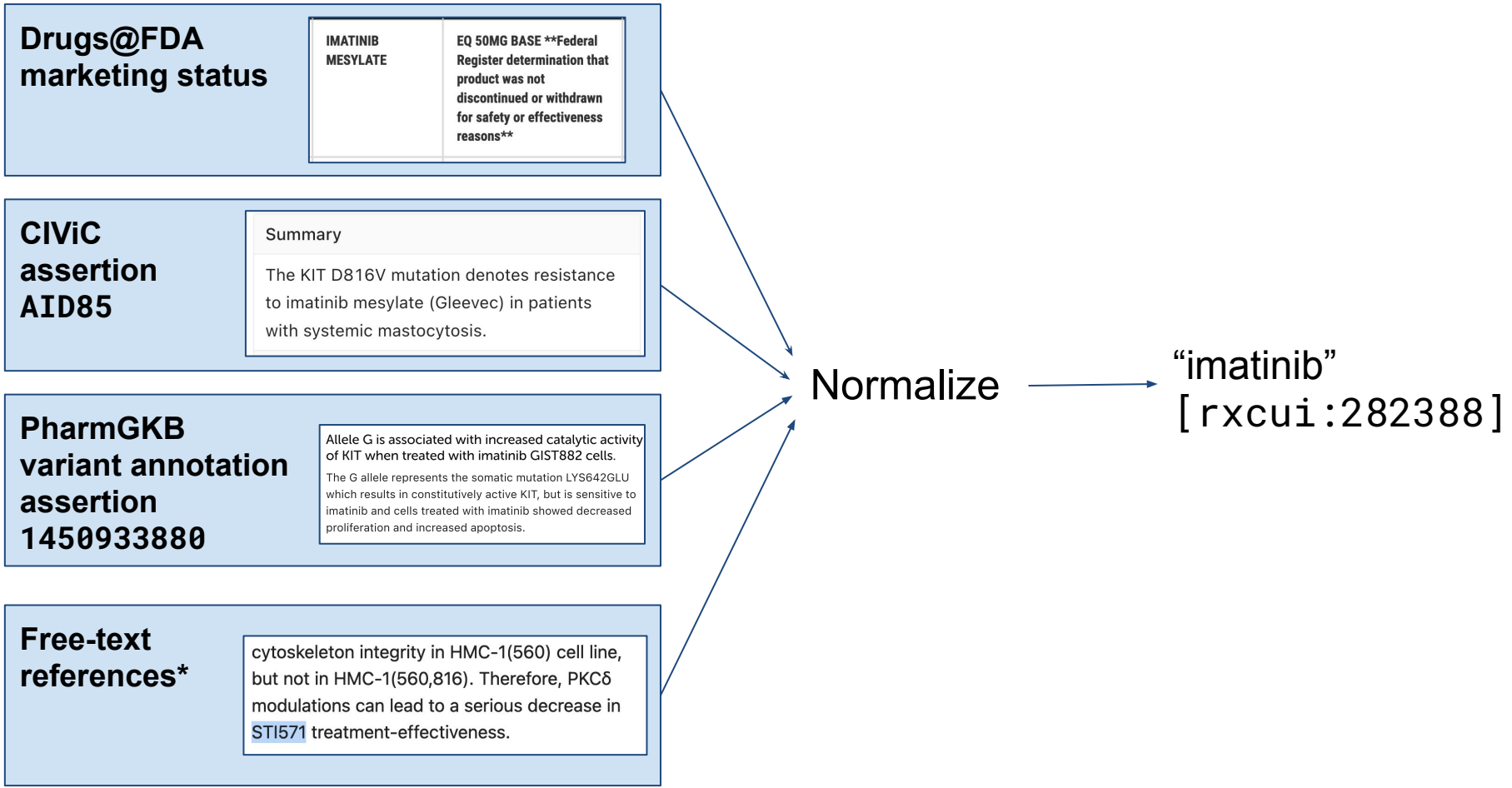
Overcoming challenges in semantic alignment of therapeutics knowledge with TheraPy

James Stevenson

The mission

Align therapeutics data resources using **normalized drug and therapy concepts** to enable clinical data harmonization from open-access knowledgebases.


Drugs@FDA marketing status
IMATINIB MESYLATE
EQ 50MG BASE **Federal Register determination that product was not discontinued or withdrawn for safety or effectiveness reasons**
CIViC assertion AID85
Summary
The KIT D816V mutation denotes resistance to imatinib mesylate (Gleevec) in patients with systemic mastocytosis.
PharmGKB variant annotation assertion 1450933880
Allele G is associated with increased catalytic activity of KIT when treated with imatinib GIST882 cells.
The G allele represents the somatic mutation LYS642GLU which results in constitutively active KIT, but is sensitive to imatinib and cells treated with imatinib showed decreased proliferation and increased apoptosis.
Free-text references*
cytoskeleton integrity in HMC-1(560) cell line, but not in HMC-1(560,816). Therefore, PKCδ modulations can lead to a serious decrease in STI571 treatment-effectiveness.
Normalize
“imatinib”
[rxcui:282388]

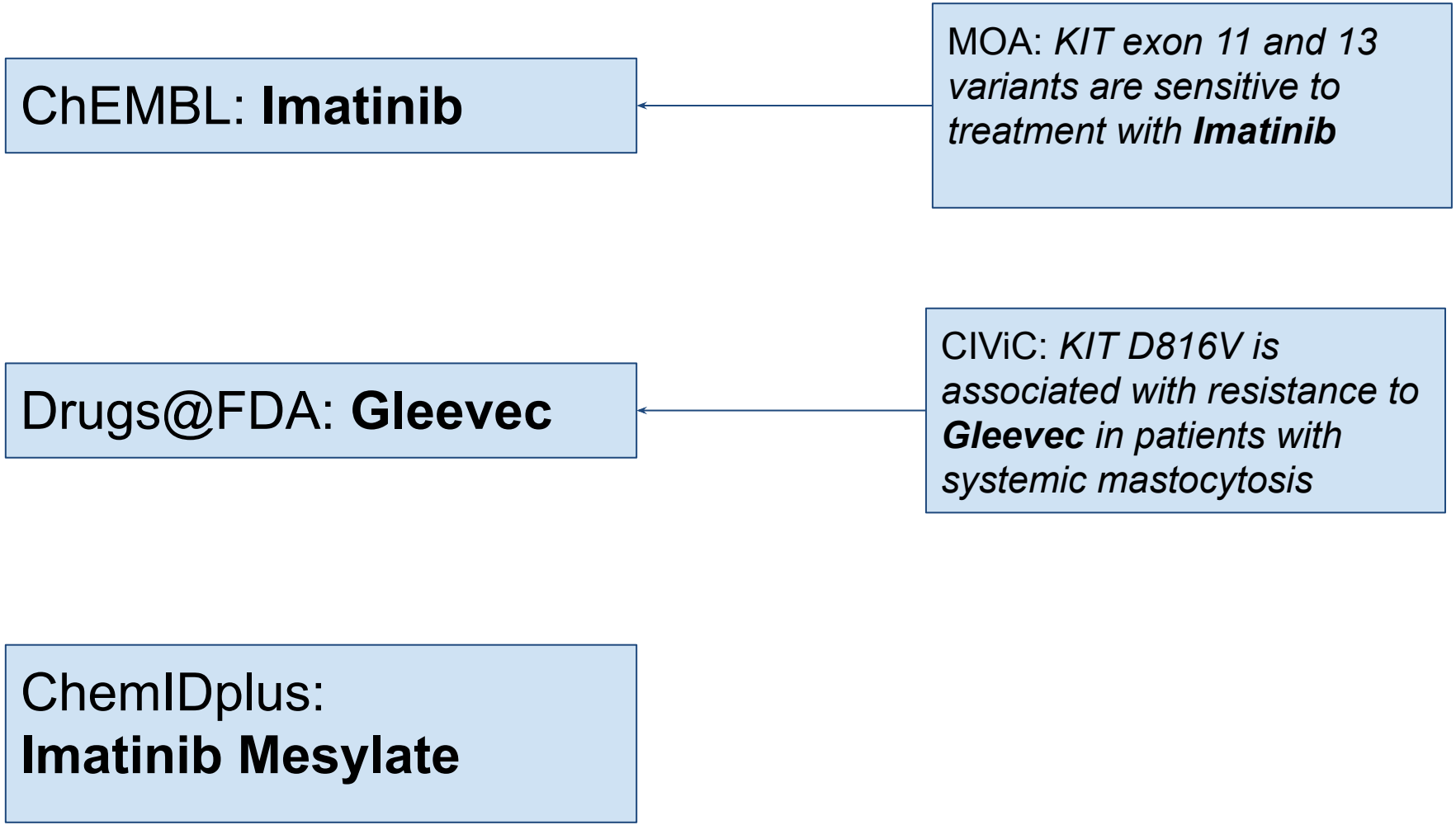
Our problem

Data is

fragmented,

imperfectly structured, and

inconsistently referenced.

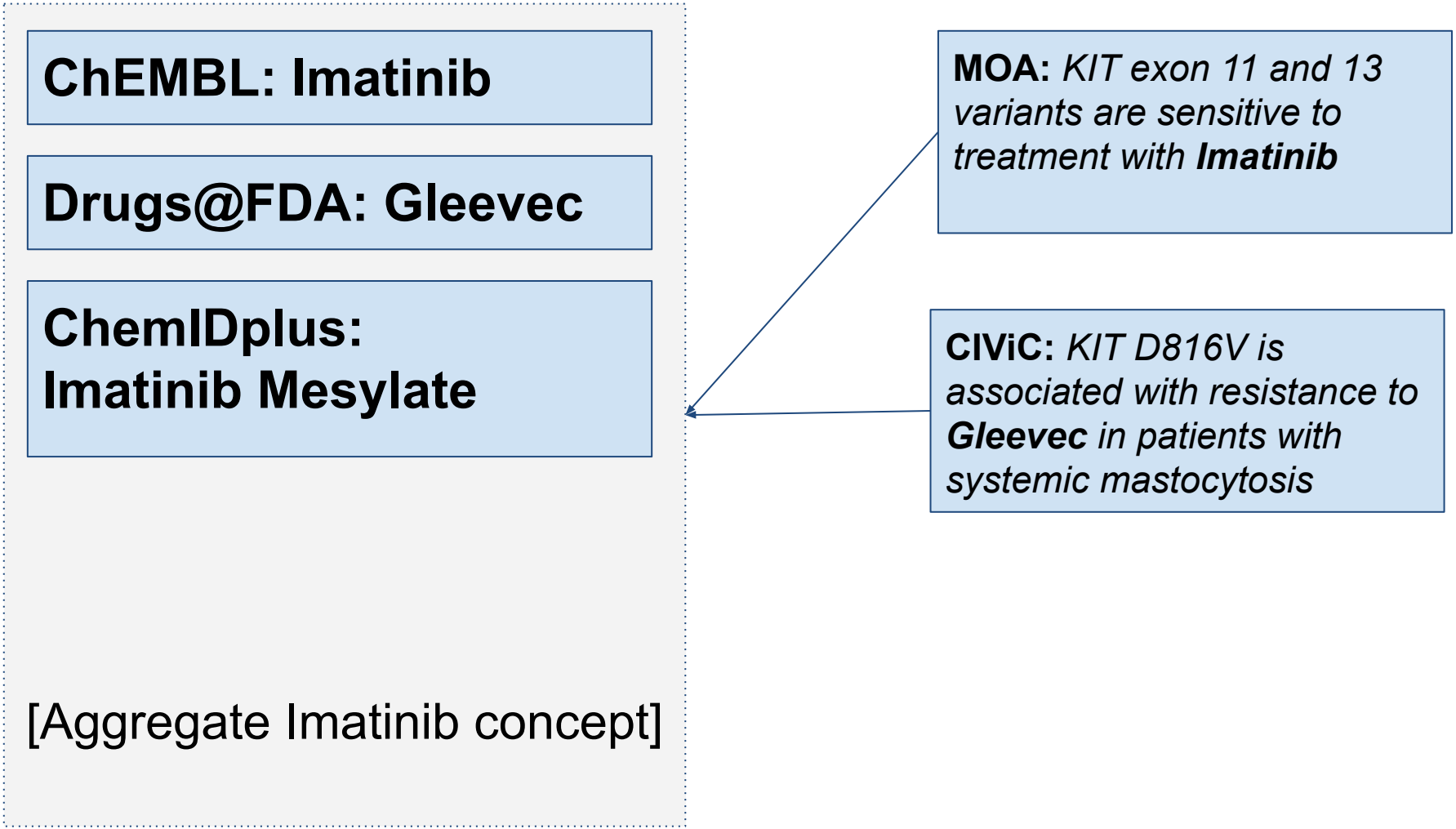

ChEMBL: **Imatinib**
MOA: *KIT exon 11 and 13 variants are sensitive to treatment with **Imatinib***
Drugs@FDA: **Gleevec**
CIViC: *KIT D816V is associated with resistance to **Gleevec** in patients with systemic mastocytosis*
ChemIDplus: **Imatinib Mesylate**

ChEMBL: Imatinib

Drugs@FDA: Gleevec

ChemIDplus:
Imatinib Mesylate

[Aggregate Imatinib concept]


ChEMBL: Imatinib
Drugs@FDA: Gleevec
ChemIDplus:
Imatinib Mesylate
[Aggregate Imatinib concept]
MOA: KIT exon 11 and 13 variants are sensitive to treatment with Imatinib
CIViC: KIT D816V is associated with resistance to Gleevec in patients with systemic mastocytosis

Referential ambiguity

Generic name	Sunitinib
Brand name	Sutent, Lucisun, Sunitix
Active ingredient	Sunitinib Malate
Development ID	SU-11248
IUPAC chemical structure	5-(5-Fluoro-2-oxo-1,2-dihydro…
Database ID	CHEMBL535, D0R0MW

Referential ambiguity

Generic name	Sunitinib
Brand name	Sutent, Lucisun, Sunitix
Active ingredient	Sunitinib Malate
Development ID	SU-11248
IUPAC chemical structure	5-(5-Fluoro-2-oxo-1,2-dihydro…
Database ID	CHEMBL535, D0R0MW

Referential ambiguity

Generic name	Sunitinib
Brand name	Sutent, Lucisun, Sunitix
Active ingredient	Sunitinib Malate
Development ID	SU-11248
IUPAC chemical structure	5-(5-Fluoro-2-oxo-1,2-dihydro…
Database ID	CHEMBL535, D0R0MW

Referential overlap

Aliases may be:

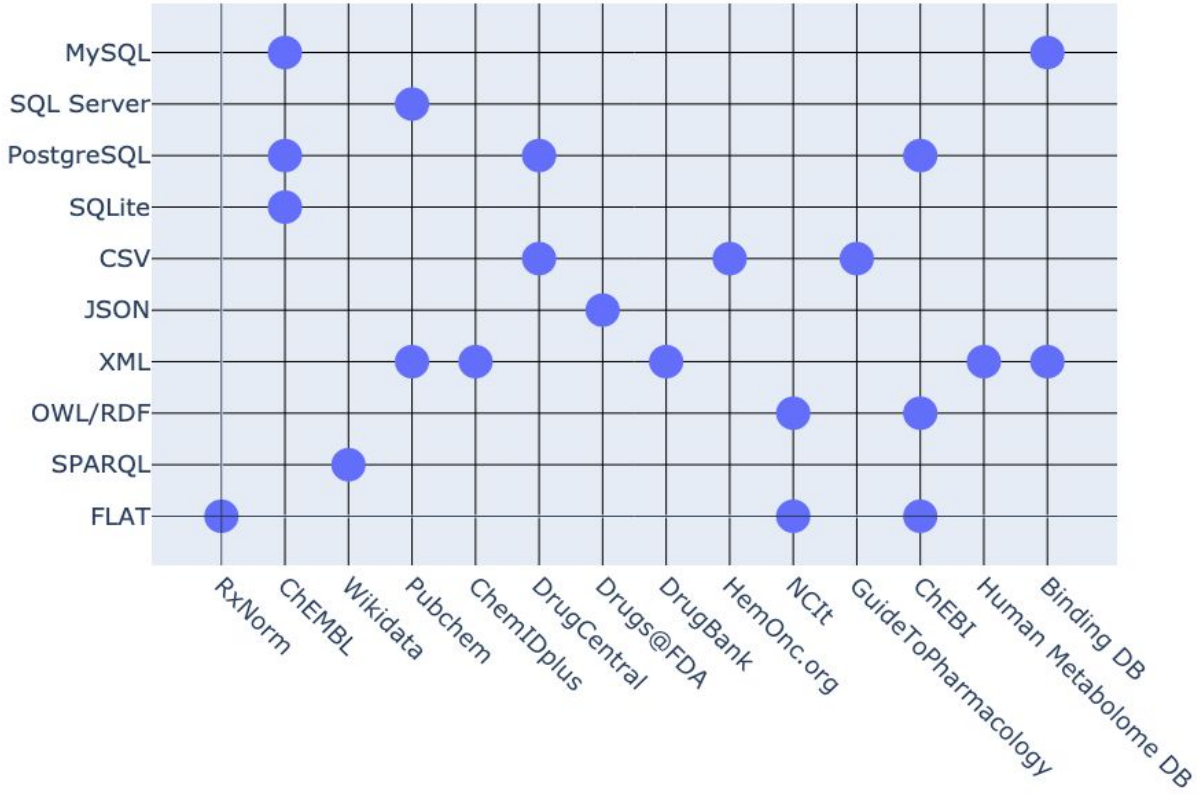
- Overly broad
  - “Ig gamma-1 chain C region”
  - “Anti-HER2”
- Indeterminate
  - “A6”

Varying data structures

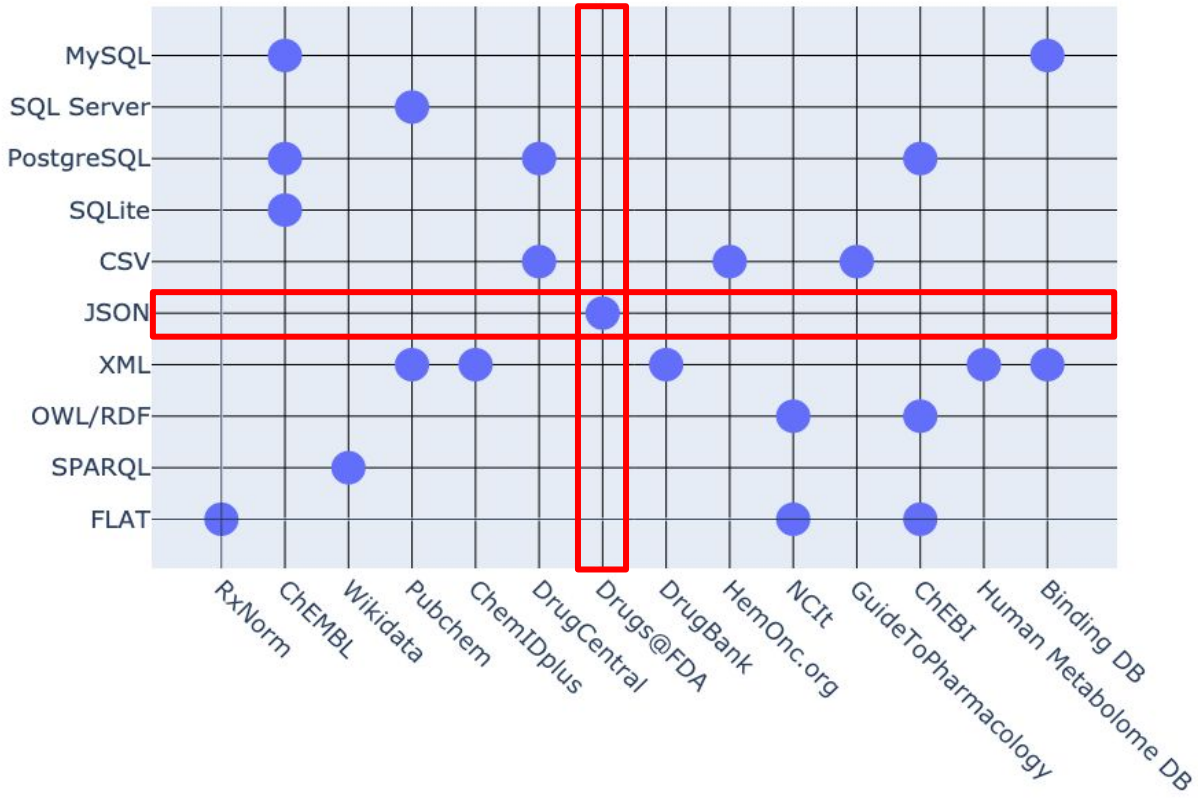
Data exports may be:

- **Relational** (SQL)
- **Key-value** (JSON)
- **Markup trees** (XML)
- **Semantic triples** (RDF, OWL)

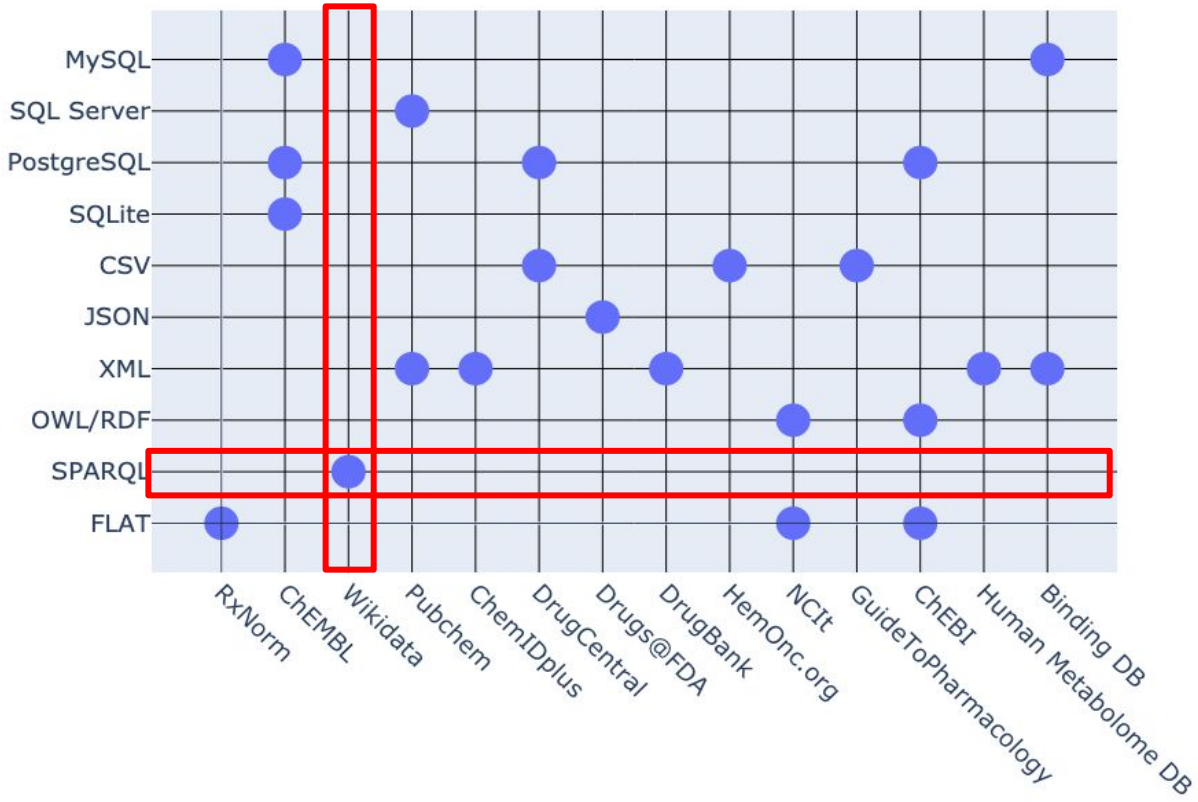
Varying data structures



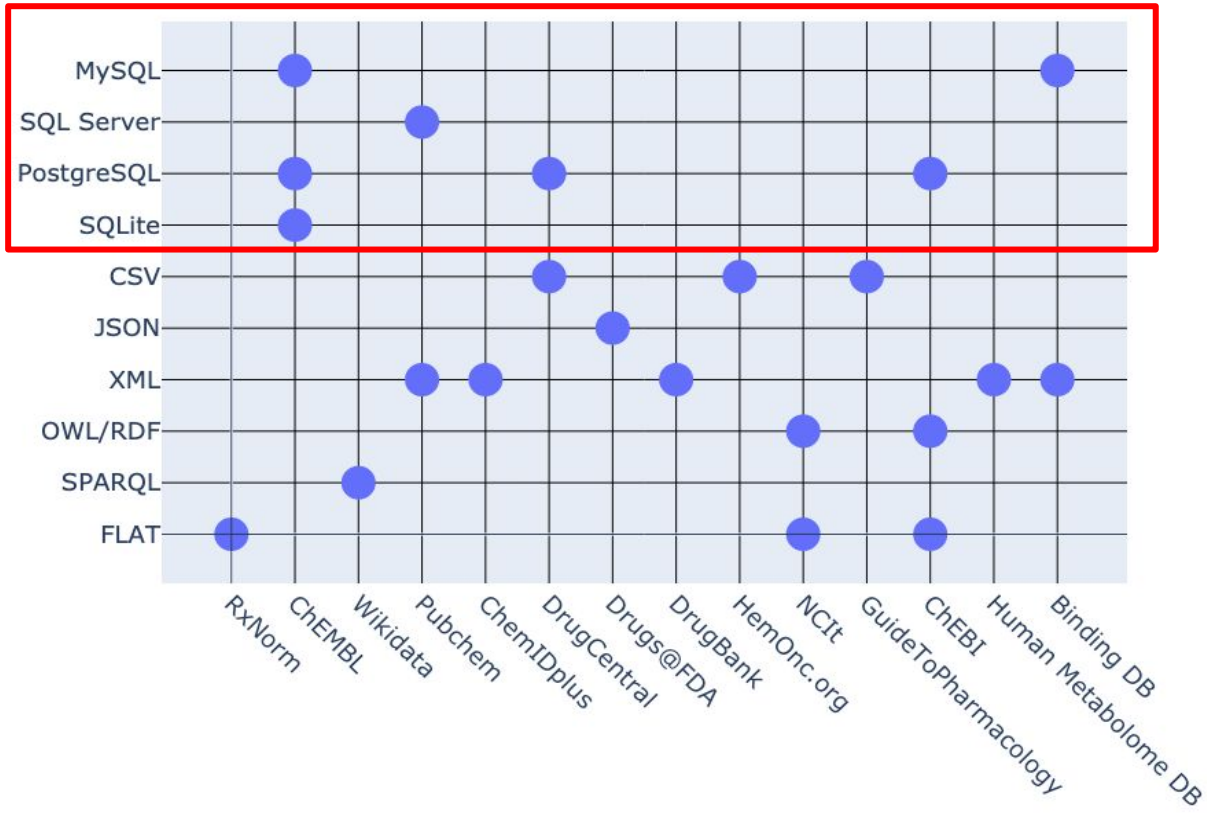
Varying data structures



Varying data structures



Varying data structures



Incongruous conceptual structures

- “Chemicals”, “compounds”, “small molecules”
- “Ligands”
- “Drugs”
- “Drug products”
- Combination products, combination therapies

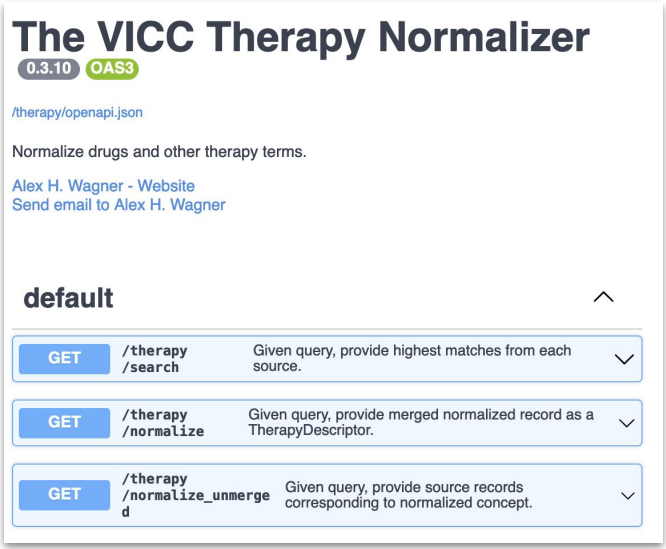
Cross-referential ambiguity

- Is an xref to another source intended to convey…
  - A is equivalent to B?
  - A is contained by B?
  - A contains B?
  - A is related to B?

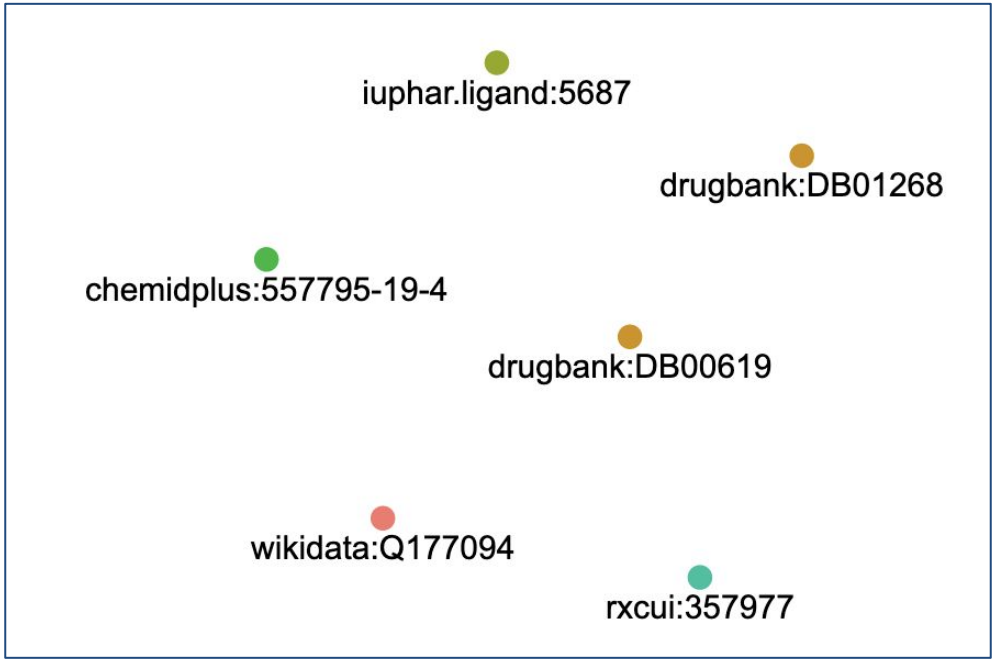
Alignment framework and implementation: TheraPy

TheraPy

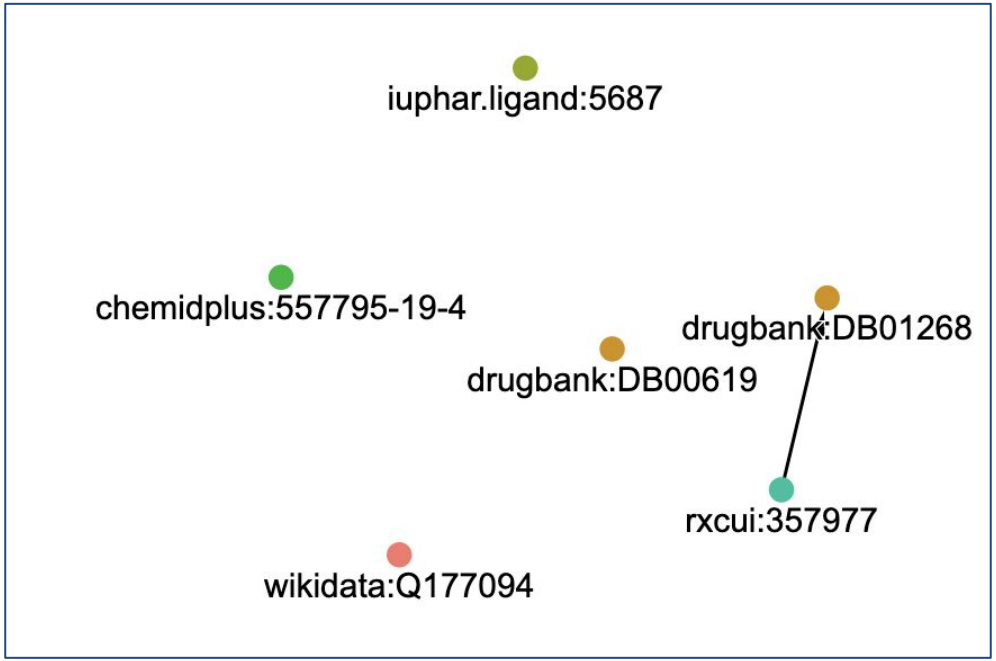
Produces ~16,000 multi-source, merged therapy concepts



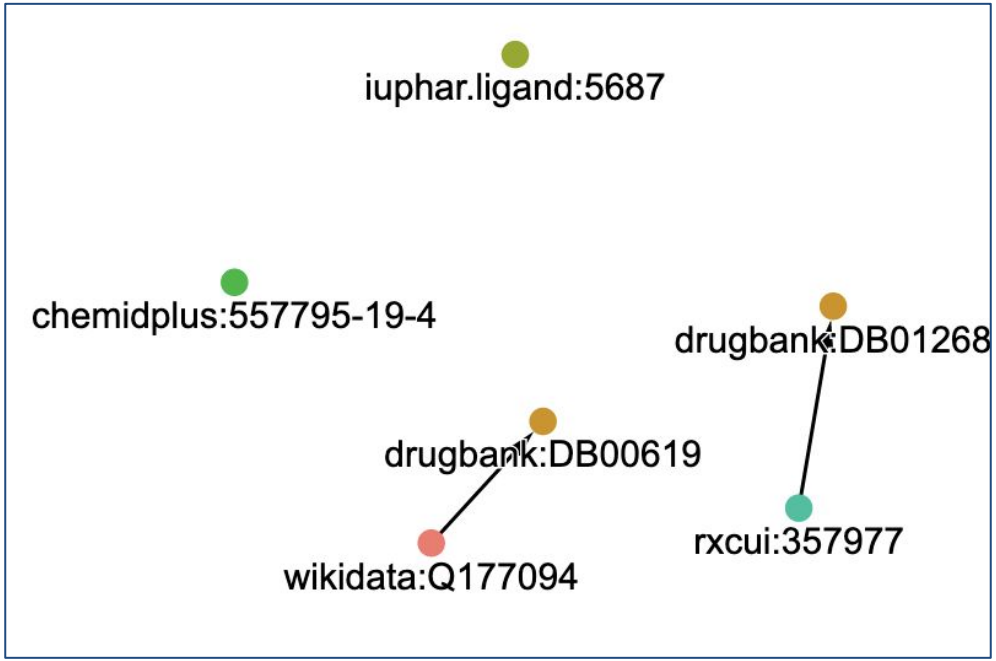
Concept generation



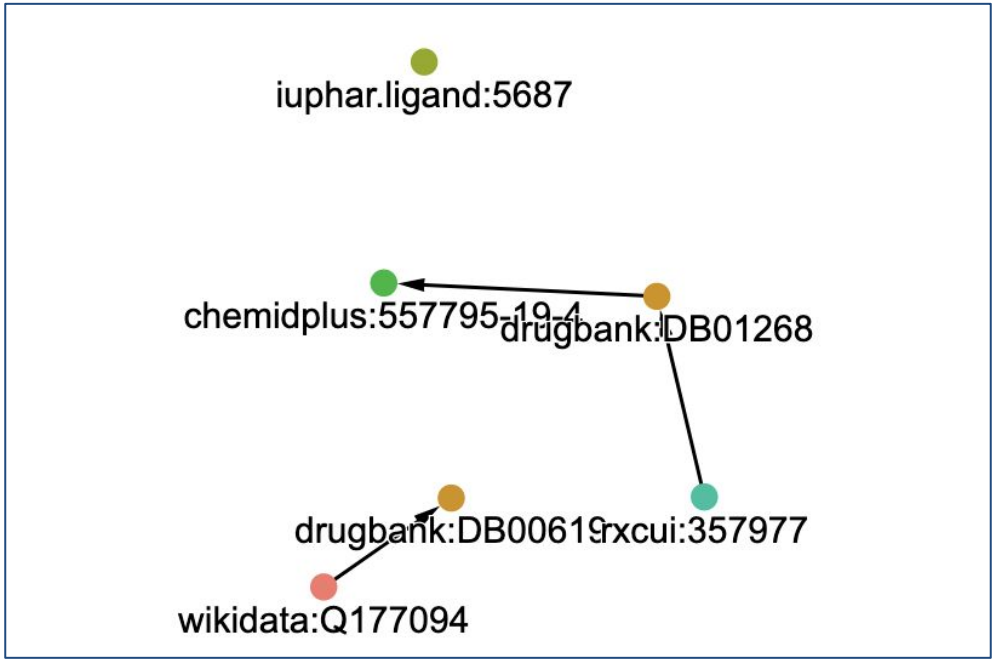
Concept generation



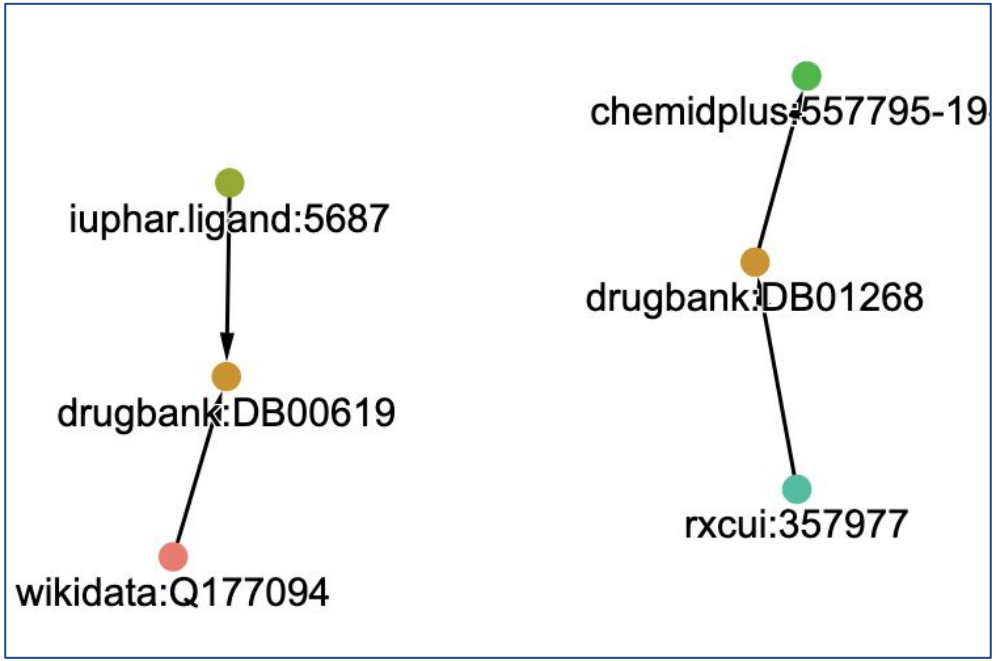
Concept generation



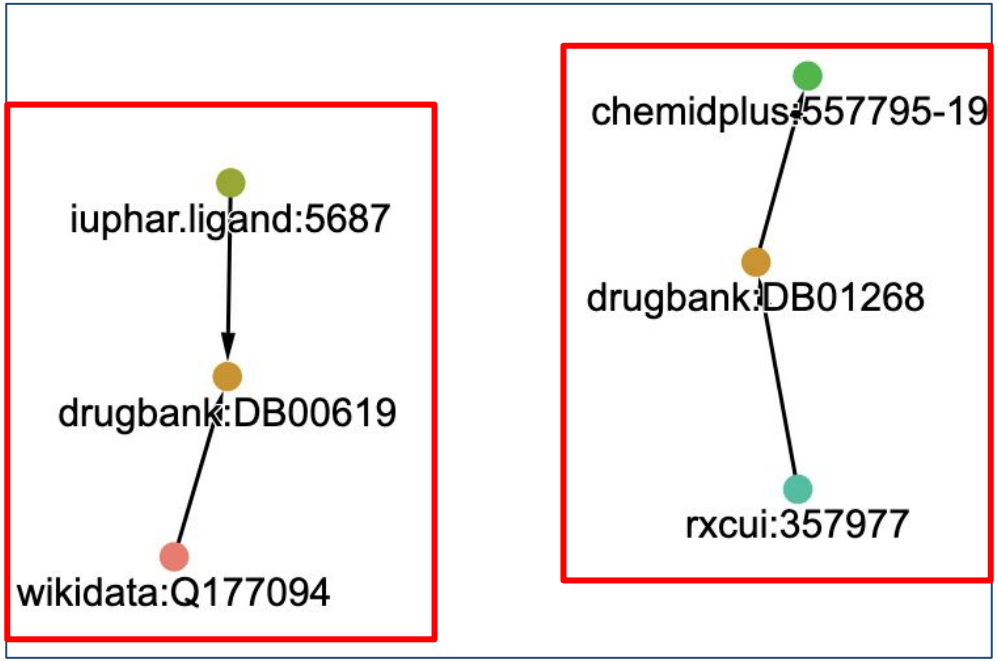
Concept generation



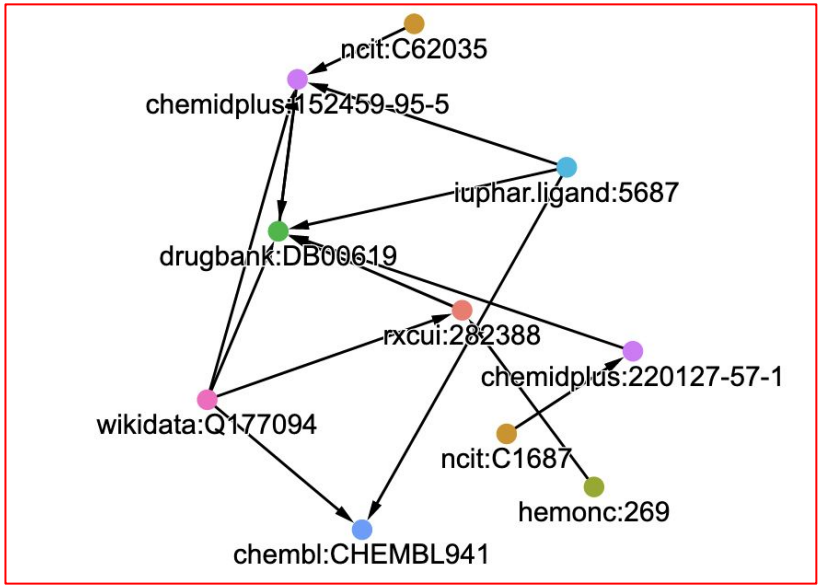
Concept generation



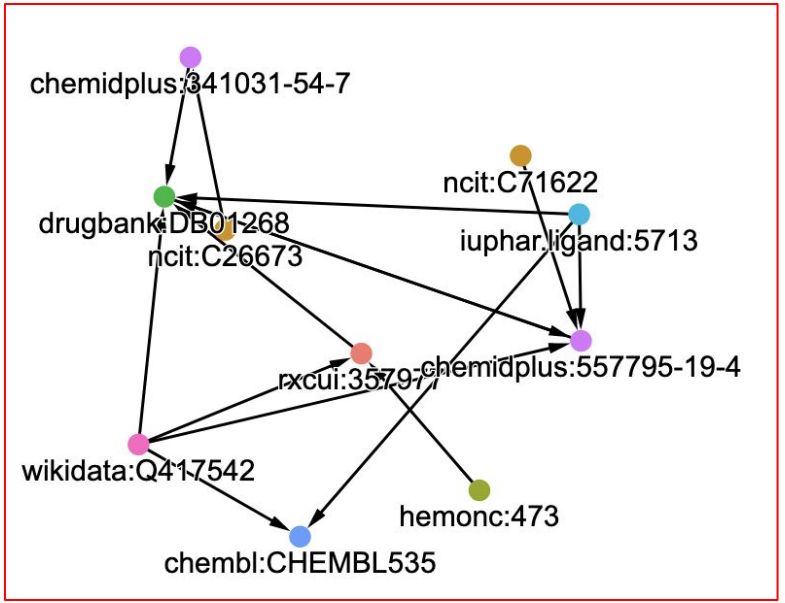
Concept generation



Concept generation



Imatinib [`rxcui:282388`]



Sunitinib [`rxcui:357977`]

Monday, August 14, 2023 @ 11:45 AM

Improving interoperability of therapeutics and their targets for clinical and precision medicine applications

Matthew Cannon, *Nationwide Children's Hospital*

Normalization of Drug and Therapeutic Concepts with TheraPy

Matthew Cannon, James Stevenson, Kori Kuzma, Susanna Kiwala, Jeremy L Warner, Obi L Griffith, Malachi Griffith, Alex H Wagner

doi: https://doi.org/10.1101/2023.07.27.23293245

Future: Refining grouping

- Probabilistic mapping (Bayesian methods)
- Increased manual curation
- Additional data sources
- Applications in NLP

Conclusion

- Open-source drug data is plentiful, but **structural inconsistencies** hamper inter-source alignment.
- TheraPy **pools drug descriptions** to define a normalized concept and mappings in support of data harmonization.

Acknowledgements

Nationwide Children's Hospital

- Alex Wagner, PhD
- Matthew Cannon, PhD
- Kori Kuzma

Brown University

- Jeremy Warner, MD

Washington University, St. Louis

- Obi Griffith, PhD
- Malachi Griffith, PhD
- Susanna Kiwala
- Adam Coffman

Funding via NHGRI #R00HG010157

Get in touch

- james.stevenson@nationwidechildrens.org
- https://go.osu.edu/tpy
- https://cancervariants.org/projects/integration/