

UNIVERSITY
OF APPLIED
SCIENCES
UTRECHT



Funded by the Horizon 2020
Framework Programme of the
European Union

Building Adverse Outcome Pathways with Natural Language Processing

Marc A.T. Teunis*, marc.teunis@hu.nl

Marie Corradi,

Jan-Willem Lankhaar - University of Applied Sciences, Utrecht

Eefje Poppelaars - Vivaltes B.V.



24 September 2021

OpenTox Virtual Conference

15.00 Session 15. Semantic Interoperability in Data
Annotation and Migration

*presenting author

This project has received funding from
the European Union's Horizon 2020
research and innovation programme
under grant agreement No 963845

This research is part of the Horizon 2020 ONTOX Project
“Ontology-driven and artificial intelligence-based repeated
dose toxicity testing of chemicals for next generation risk
assessment” and the NC3Rs Crack-it project “DARTpaths”

ONTOX; Consortium and budget

<https://ontox-project.eu/>

Consortium

- 18 partners/9 different countries
- 19 teams/60 researchers
- 10 academic institutions
- 5 SMEs/1 large company
- 1 public health institution
- 1 consultant
- Started on 1st May 2021
- Budget: 17 M EURO



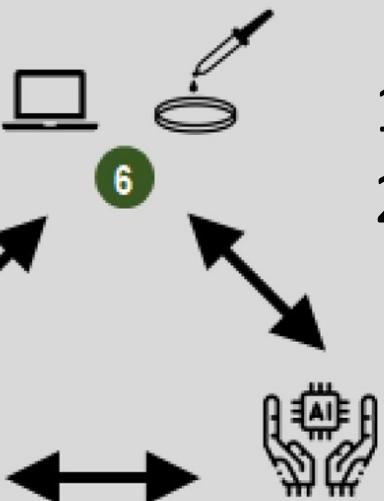
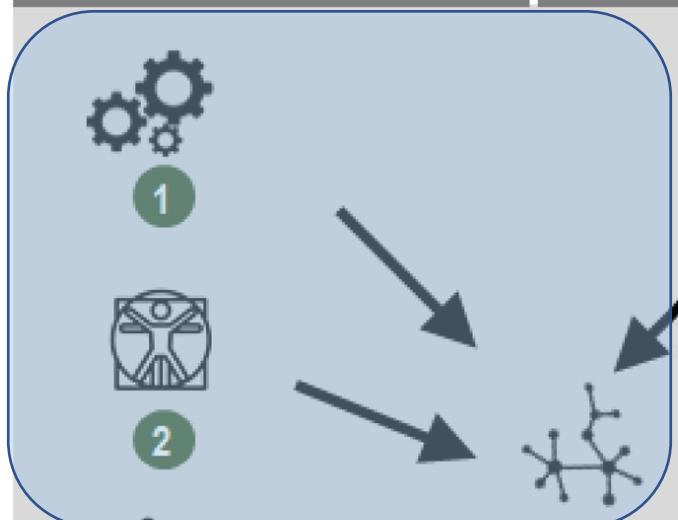
Vinken M, Benfenati E, Busquet F, Castell J, Clevert DA, de Kok TM, Dirven H, Fritsche E, Geris L, Gozalbes R, Hartung T, Jennen D, Jover R, Kandarova H, Kramer N, Krul C, Luechtefeld T, Masereeuw R, Roggen E, Schaller S, Vanhaecke T, Yang C, Piersma AH. Safer chemicals using less animals: kick-off of the European ONTOX project. Toxicology. 2021 Jun 30;458:152846. doi: 10.1016/j.tox.2021.152846. Epub 2021 Jun 30. PMID: 34216698.

ONTOX Workflow

TIER 1: DATA COLLECTION

TIER 2: DATA INTEGRATION AND TESTING

TIER 3: RISK ASSESSMENT AND IMPLEMENTATION

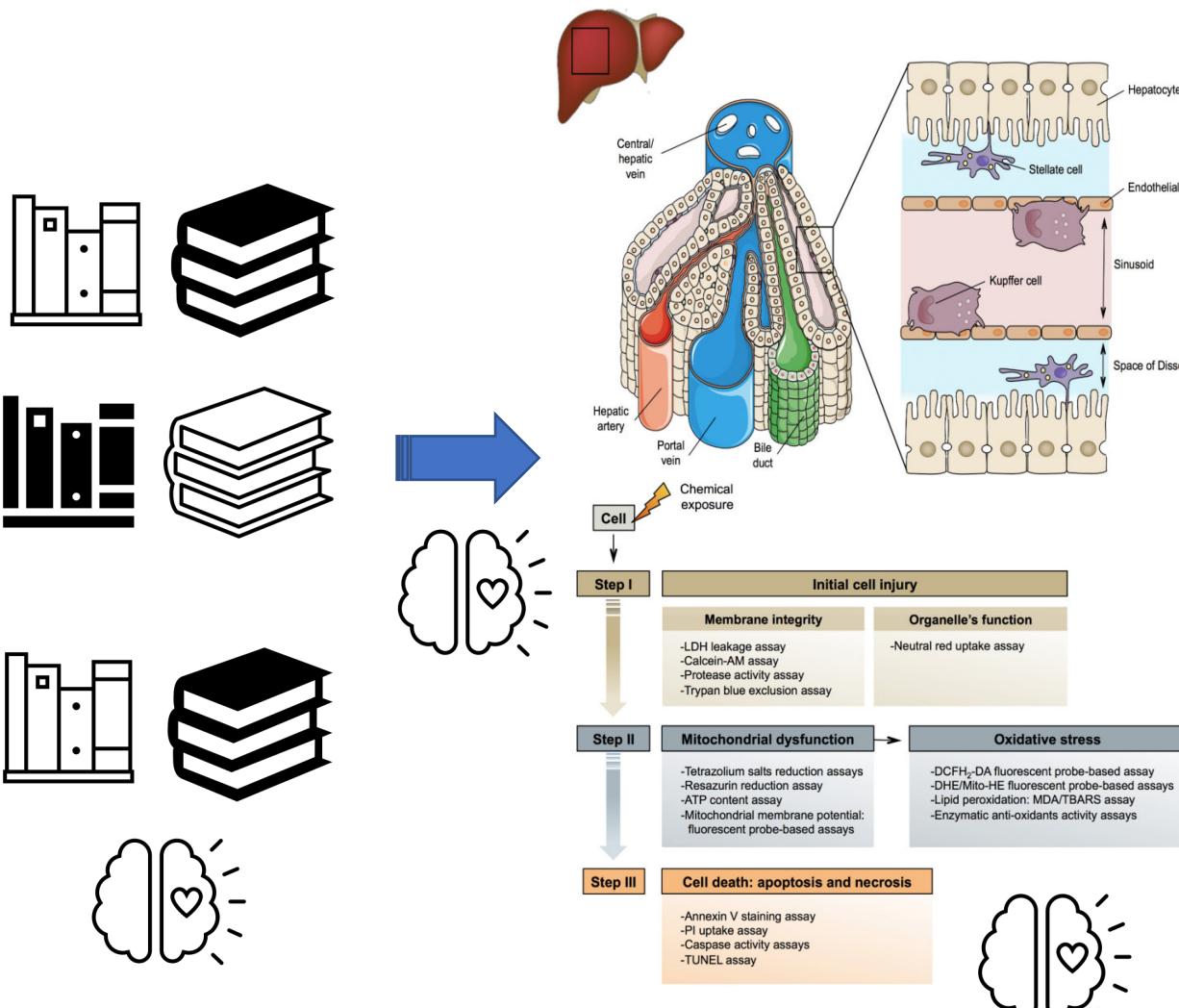


1. SUPPORTING DATA COLLECTION
2. DEVELOP NLP ASSISTED DATA EXTRACTION

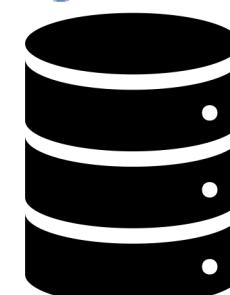


Vinken M, Benfenati E, Busquet F, Castell J, Clevert DA, de Kok TM, Dirven H, Fritzsche E, Geris L, Gozalbes R, Hartung T, Jennen D, Jover R, Kandarova H, Kramer N, Krul C, Luechtelefeld T, Masereeuw R, Roggen E, Schaller S, Vanhaecke T, Yang C, Piersma AH. Safer chemicals using less animals: kick-off of the European ONTOX project. *Toxicology*. 2021 Jun 30;458:152846. doi: 10.1016/j.tox.2021.152846. Epub 2021 Jun 30. PMID: 34216698.

From literature to physiology to AOP to qAOP



sysrev



[Int J Mol Sci. 2021 May; 22\(9\): 5038.](#)

Published online 2021 May 10. doi: [10.3390/ijms22095038](https://doi.org/10.3390/ijms22095038)

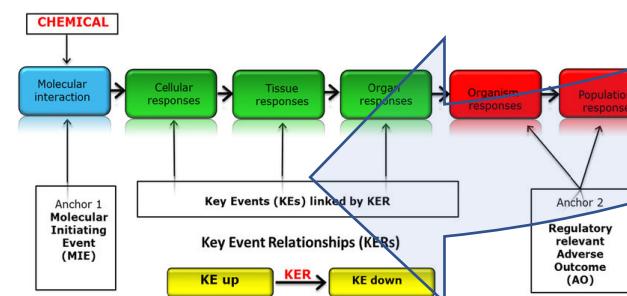
PMCID: PMC8126138

PMID: 34068678

In Vitro Liver Toxicity Testing of Chemicals: A Pragmatic Approach

Andrés Tabernilla,[†] Bruna dos Santos Rodrigues,[†] Alanah Pieters, Anne Caufriez, Kaat Leroy, Raf Van Campenhout, Axelle Cooreman, Ana Rita Gomes, Emma Amesdotter, Eva Gijbels, and Mathieu Vinken^{*}

Lina Ghibelli, Academic Editor



Collecting data for AOPs

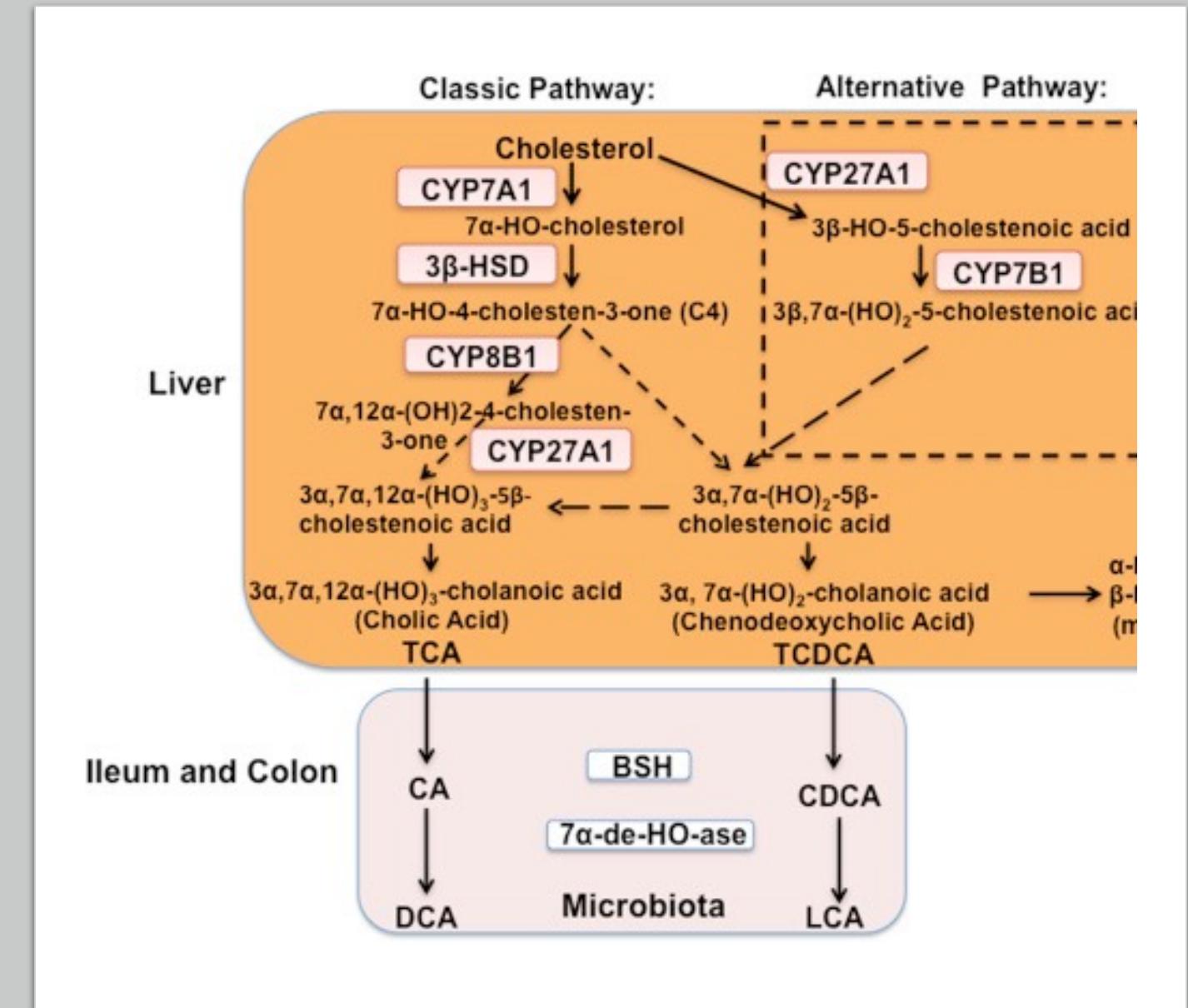
- Many 'domain experts' working together
- Needs to be standardized
- Need for clear templates
- Collection of 'reactions', 'proteins', 'compounds', 'enzymes', 'compartments', 'quantities', 'regulators', 'genes', 'relations', 'references' and 'definitions'
- Need for an interoperable user-friendly format
- Sbtab -> SBML

Sbtab: <https://github.com/tlubitz/SBtab>

Liver bile synthesis

- proof of concept -

- <https://doi.org/10.1016/j.livres.2017.05.001>
- Proof of concept for collecting data from literature
- Casting information into SBtab format, using Shiny app
- Export Sbtab to SBML – xml
- Import SBML – xml into CellDesigner



Entering data in a Shiny app

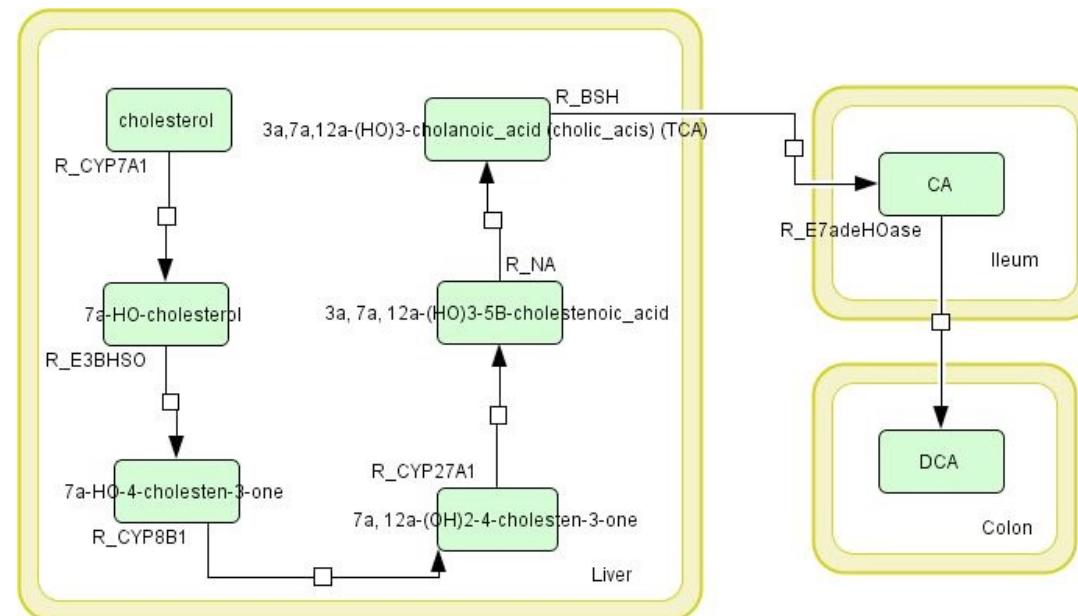
- <https://ontox-pmdep.shinyapps.io/ontox-pmdep/>
- Data is entered in Sbtab (tabular SBML) format
- Export to SBML and BioPAX (in the future)

The screenshot shows a web-based application titled "Data Entry Portal". At the top, there is a blue header bar with the title and a menu icon. Below the header, a message says "and discover how it works (default)". A table is displayed with columns labeled: ComponentName, Description, and several other columns that are mostly empty or have placeholder text like "and discover how it works (default)".

ComponentName	Description
BuildEnzyme	Include enzyme in SBML model
BuildEnzymeProduction	Describe enzyme production in SBML model
BuildReaction	Include reaction in SBML model
Comment	Comment
Description	Description of the row element
Enzyme	Enzyme catalysing the reaction
Enzyme:SBML:parameter:id	SBML ID of enzyme
Enzyme:SBML:species:id	SBML ID of enzyme

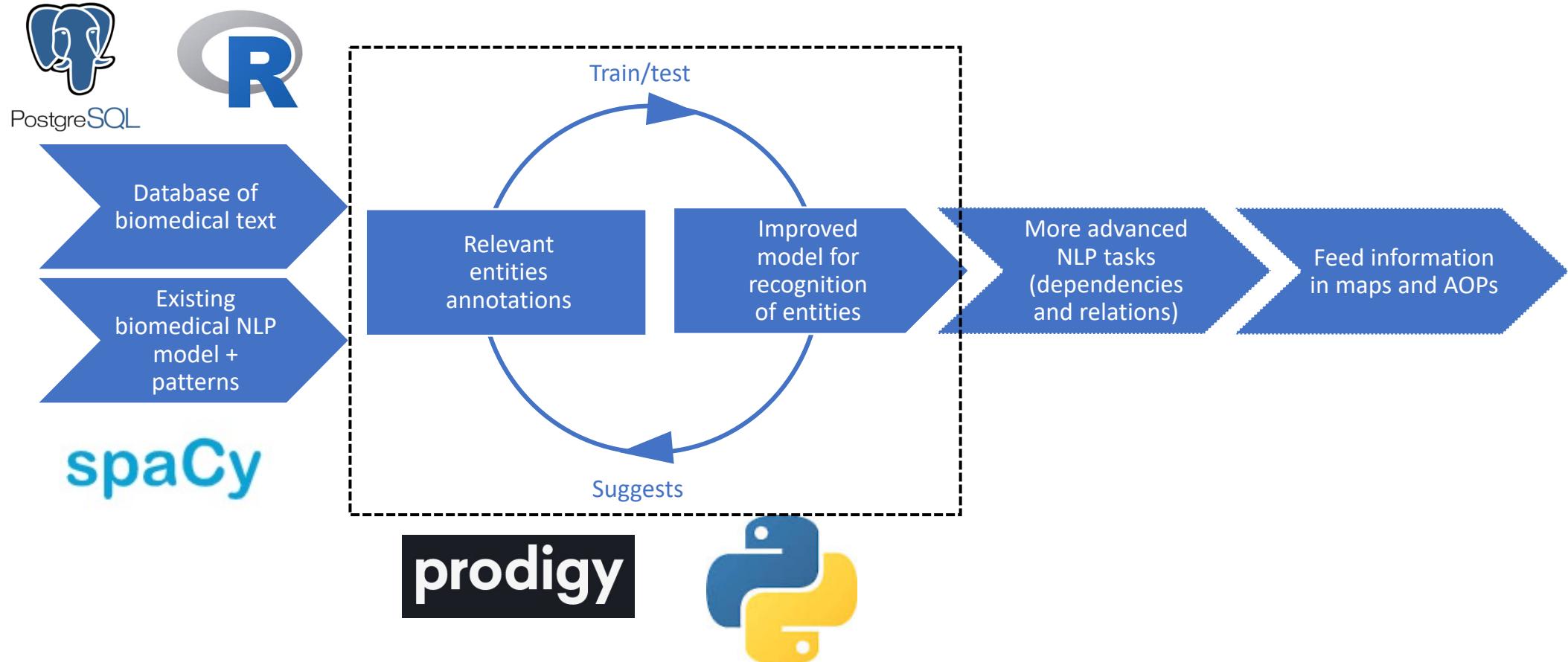
Crude CellDesigner map from SBML

- <http://www.celldesigner.org/>
- Data is imported into CellDesigner from SBML – xml format
- Result is a crude map – needs manual editing

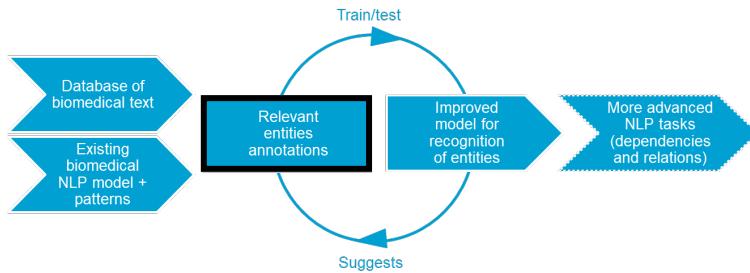


Supporting information extraction with Natural Language Processing

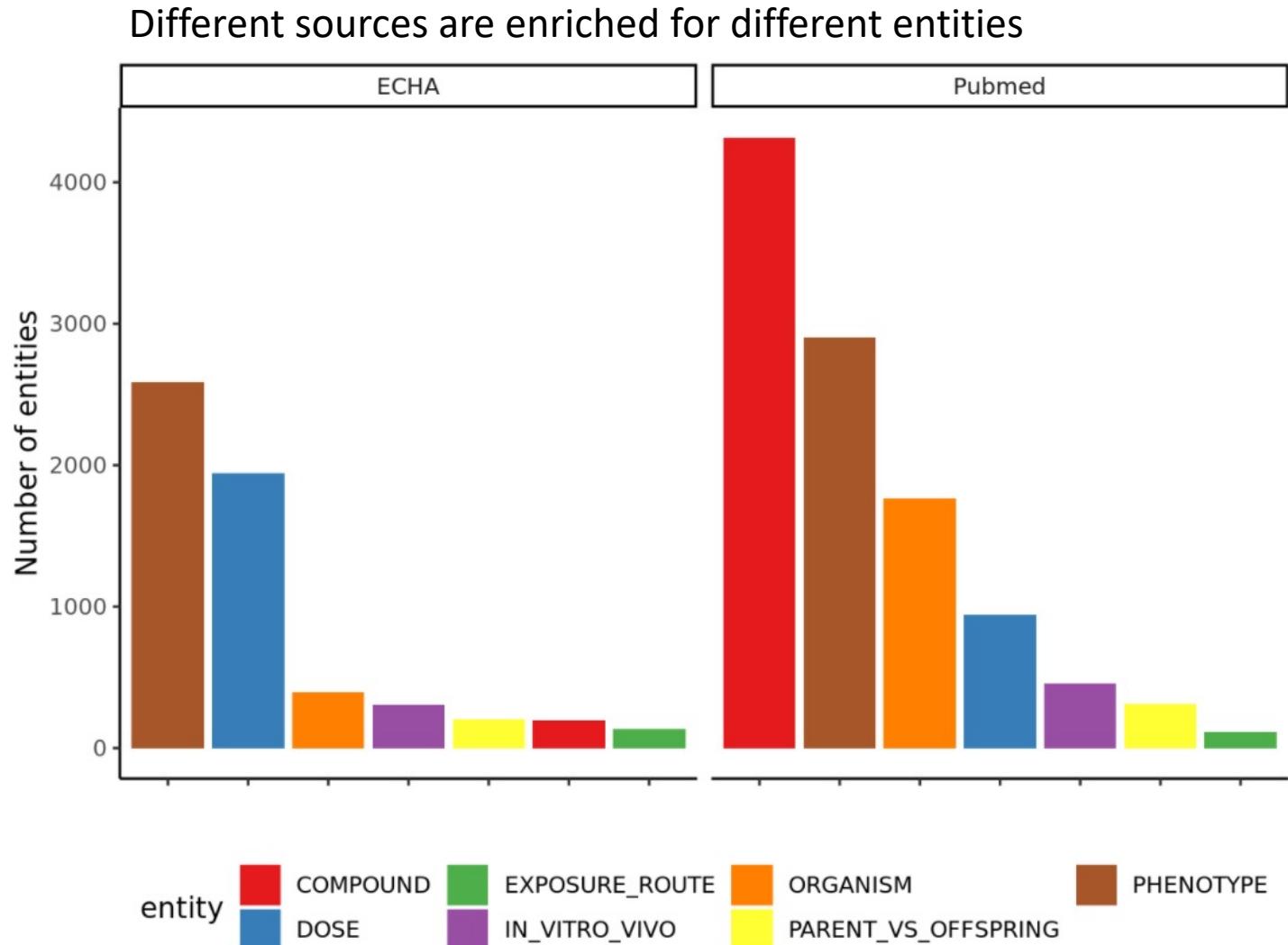
This work has been supported by the NC3Rs,
CRACK-it project
[DARTPaths](#)



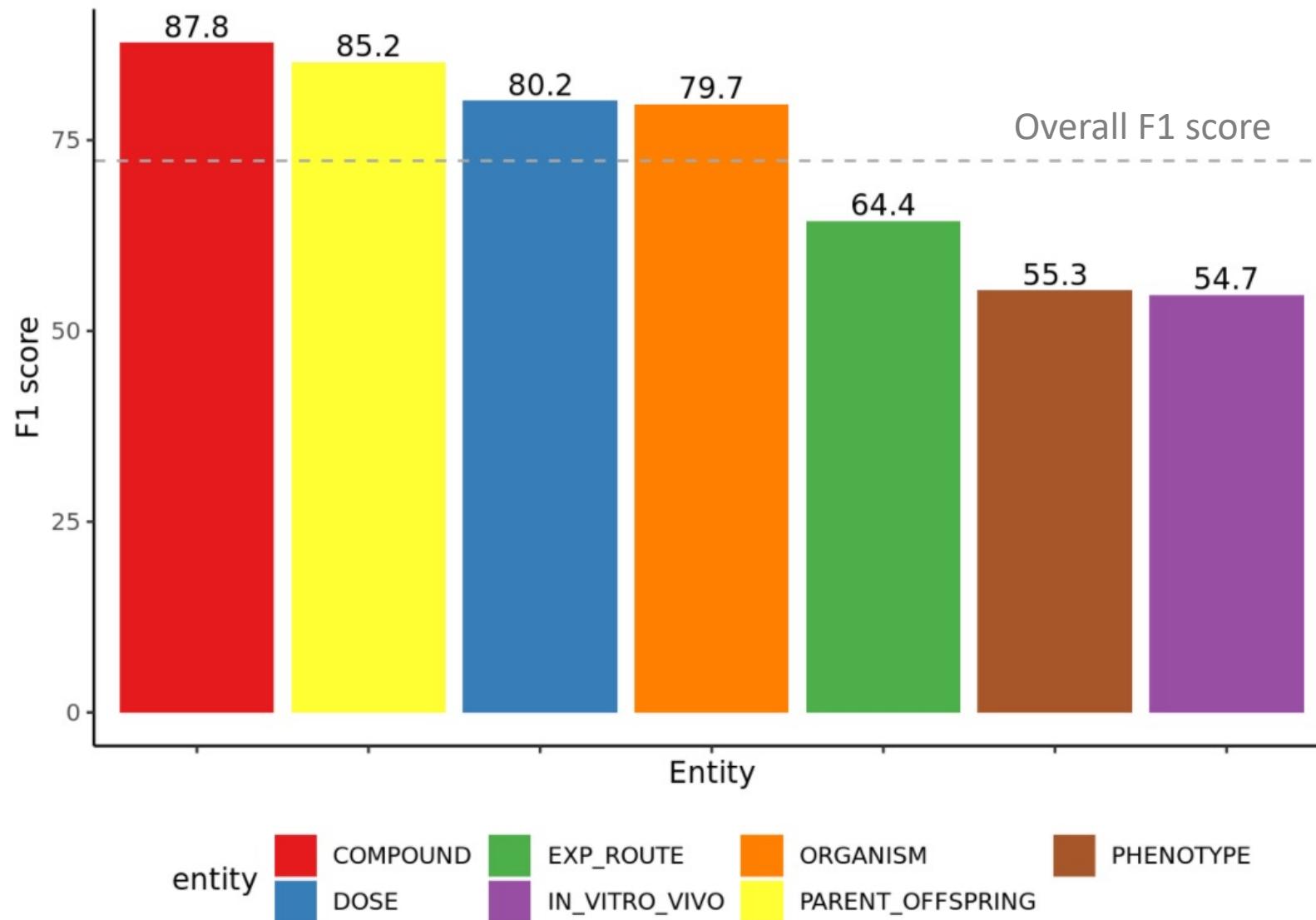
Entities & annotation process



Around **8000** sentences annotated, corresponding to **5000** sentences from **100 Pubmed articles** and **3000** sentences from **200 ECHA reports**



Model performance – metrics on training text



- F1 = weighted average of Precision and Recall
- “Easy” entities are well recognized
- Phenotype is still difficult
- EXP_ROUTE and IN_VITRO_VIVO possibly lack annotations

Drug-induced liver injury

- proof of concept - retroviral inhibitors: ritonavir nevirapine

```
from metapub import PubMedFetcher
fetch = PubMedFetcher()

# get PMIDs matching query for antiviral agents and DILI
# We do not set an API key here as this is for ~10 articles, but with more we would need one to be more efficient
pmids = fetch.pmids_for_query('DILI AND (ritonavir OR nevirapine)', retmax=1000)

# get abstract for each article:
abstracts = {}
for pmid in pmids:
    abstracts[pmid] = fetch.article_by_pmid(pmid).abstract
```

2021-09-08 12:36:25 8205769c2db2 metapub.config[1222] WARNING NCBI_API_KEY was not set.

```
#print(abstracts)
```

13 hits in the PubMed database



Review > Liver Int. 2019 May;39(5):802-811. doi: 10.1111/liv.14004. Epub 2018 Dec 4.

Drug-induced liver injury with skin reactions: Drugs and host risk factors, clinical phenotypes and prognosis

Harshad Devarbhavi ¹, Sujata Raj ²

Affiliations + expand

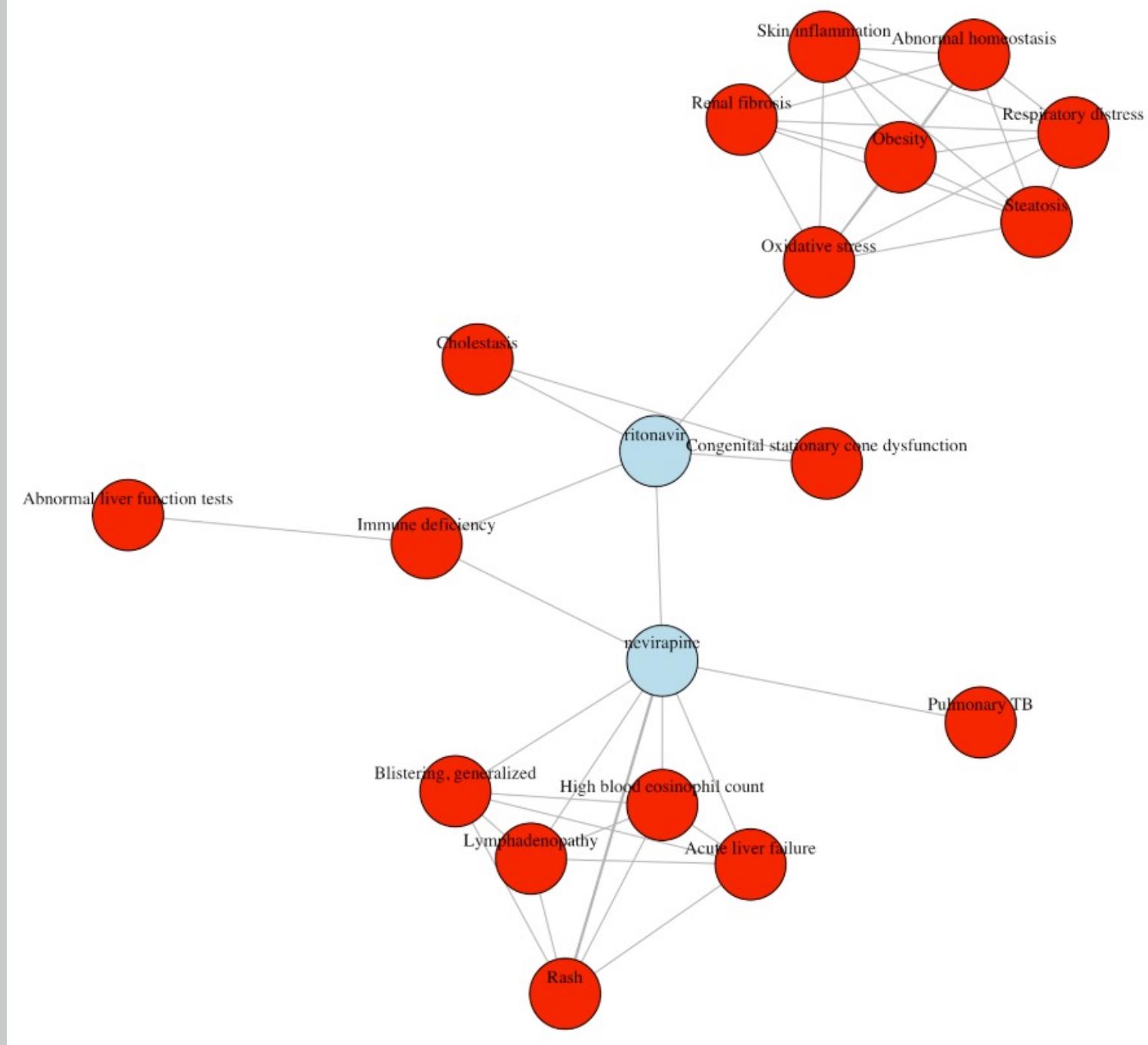
PMID: 30515930 DOI: 10.1111/liv.14004

Named Entity Recognition (NER) using the trained model

cytotoxicity to minimize false positive assay signals and other complications in data interpretation. Based on the experimental approaches employed and corresponding data, the prevalence of FXR COMPOUND antagonism was relatively low across this broad DILI PHENOTYPE test set, with 16-24% prevalence based on individual assay results or combined signals in both assays. Moreover, FXR antagonism was not highly predictive for identifying clinically relevant hepatotoxicants retrospectively, where FXR antagonist classification alone had minimal to moderate predictive value as represented by positive and negative likelihood ratios of 2.24-3.84 and 0.72-0.85, respectively. The predictivity did not increase significantly when considering only compounds with high clinical exposure (maximal or efficacious plasma exposures > 1.0 μ M DOSE). In contrast, modest gains in predictive value PHENOTYPE of FXR COMPOUND antagonism were observed considering compounds that also inhibit bile salt export pump PHENOTYPE. In addition, we have identified novel FXR antagonistic effects of well-studied hepatotoxic drugs, including bosentan COMPOUND, tolcapone COMPOUND and ritonavir COMPOUND. In conclusion, this work represents a comprehensive evaluation of FXR antagonism in the context of DILI PHENOTYPE, including its overall predictivity and challenges associated with

Mapping a co-occurrence graph from NLP results

- Phenotypic entities were linked to HPO terms via EntityLinker in spaCy
- Filtered results only when linkage was found
- Co-occurrence on the ‘corpus’ level



Future directions

Motto: "We do everything in code and will share as much as possible"

<https://github.com/ncbi-nlp/PhenoTagger>



Expand the
number of
chemicals



Expand and
improve
entity
linkage to
molecular
ontologies



Find
communities
with similar
chemicals or
similar
adversity



Describe
pathways and
mechanisms
from entity
linkage



Construct
AOP and
knowledge
graphs



Share
methods,
code and
results in R
package -
{ontoxr}



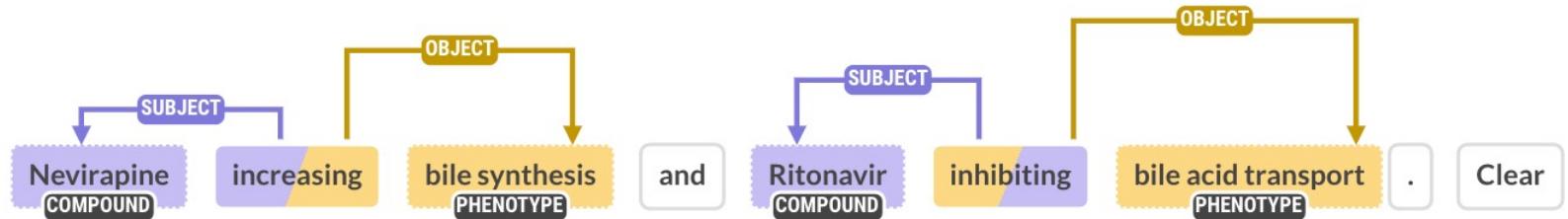
Cross-
validate
approach
to
annotated
sources
(such as
CTD)

<https://ontox-project.github.io/>

Semantic dependency parsing

At the toxic doses, both drugs showed direct cholestatic potential with

Improving biological context: chemical phenotype relations



Nevirapine
COMPOUND is associated with a 6%-10% risk of developing a hypersensitivity reaction

, with different phenotypes, including the blistering conditions Stevens-Johnson syndrome

(SJS
PHEN...) and toxic epidermal necrolysis (TEN
PHENO...)



Sharing our work:
-- UNDER CONSTRUCTION --

<https://ontox-project.github.io/>

For R packages, App source, and spaCy
NLP models