Increasing the understanding of Metabolomics data with ontologies and network approaches

Denise Slenter

@SMaLLCaT4Sci
http://smallcats4science.blogspot.nl
ORCID: 0000-0001-8449-1318

2018-11-01 BiGCaT Science Cafe

Maastricht University

Introduction

- Metabolomics Researchers gather more and more data
- Using pathway databases (WikiPathways, Reactome and KEGG) can put data in perspective
- However, metabolomics data proves difficult to be linked to biological pathway models
- Sparseness of data hinders pathway interpretation
- Two approaches are being developed @BiGCaT, to increase understanding of metabolomics data

Two approaches:

Ontological approach

Connects more data to pathway knowledge

Will be presented at German Conference of Cheminformatics (2018-11-12)

Network approach

Visualise cause-and-effect of relevant data

Will be presented at first Metabolomics in Maastricht event (2018-11-29)

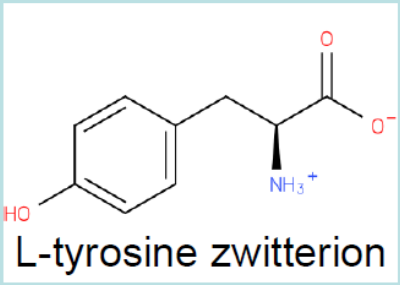
Ontological approach: AIM

- Connect annotated metabolites != in pathway, to related metabolites = in a pathway

- 6 relationship types defined in ChEBI:
  - Child compounds → "Subclass"
  - Parent compounds/groups → "SuperClass"
  - Charged forms → "Charge state"
  - Charges swopped internally → "Tautomers"
  - Diff. stereochemistry → "Enantiomers"

- Poster: https://doi.org/10.6084/m9.figshare.6368921.v1

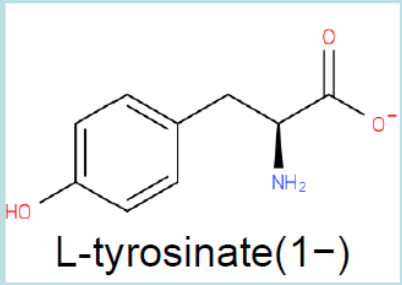
Examples of relationships:



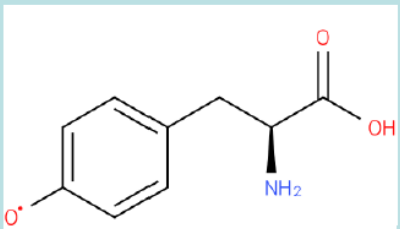
L-tyrosine zwitterion

Tautomer

Charge state

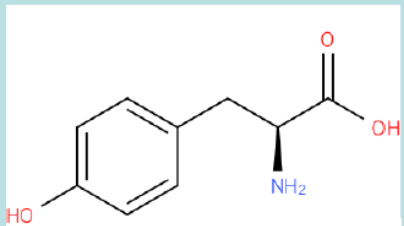


L-tyrosinate(1−)



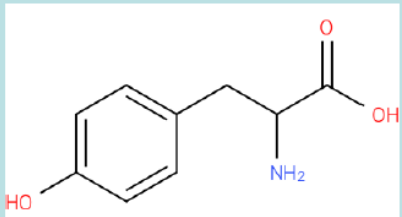
L-tyrosinyl radical

SubClass



L-tyrosine

SuperClass

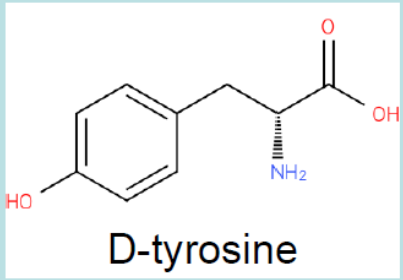


Tyrosine

- EC 1.3.1.43 (arogenate dehydrogenase) inhibitor
- Micronutrient

Roles

Enantiomer

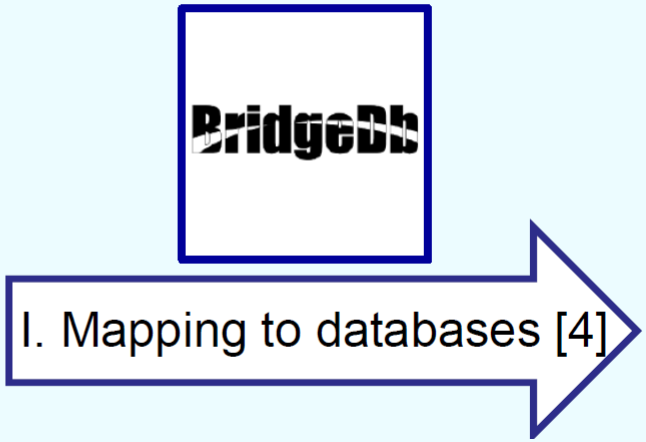


D-tyrosine

Workflow:

Metabolomics Datasets [6-9]

Identifier	Name
HMDB00538	ATP
CHEBI:16761	ADP
CHEBI:16828	L-tryptophan
CHEBI:17895	L-tyrosine



Identifier	Xref	Database
HMDB00538	DB00171	Drugbank
CHEBI:16761	C00019353	KNApSAcK
CHEBI:16828	6305	Pubchem CID
CHEBI:17895	Q188017	Wikidata

Workflow:

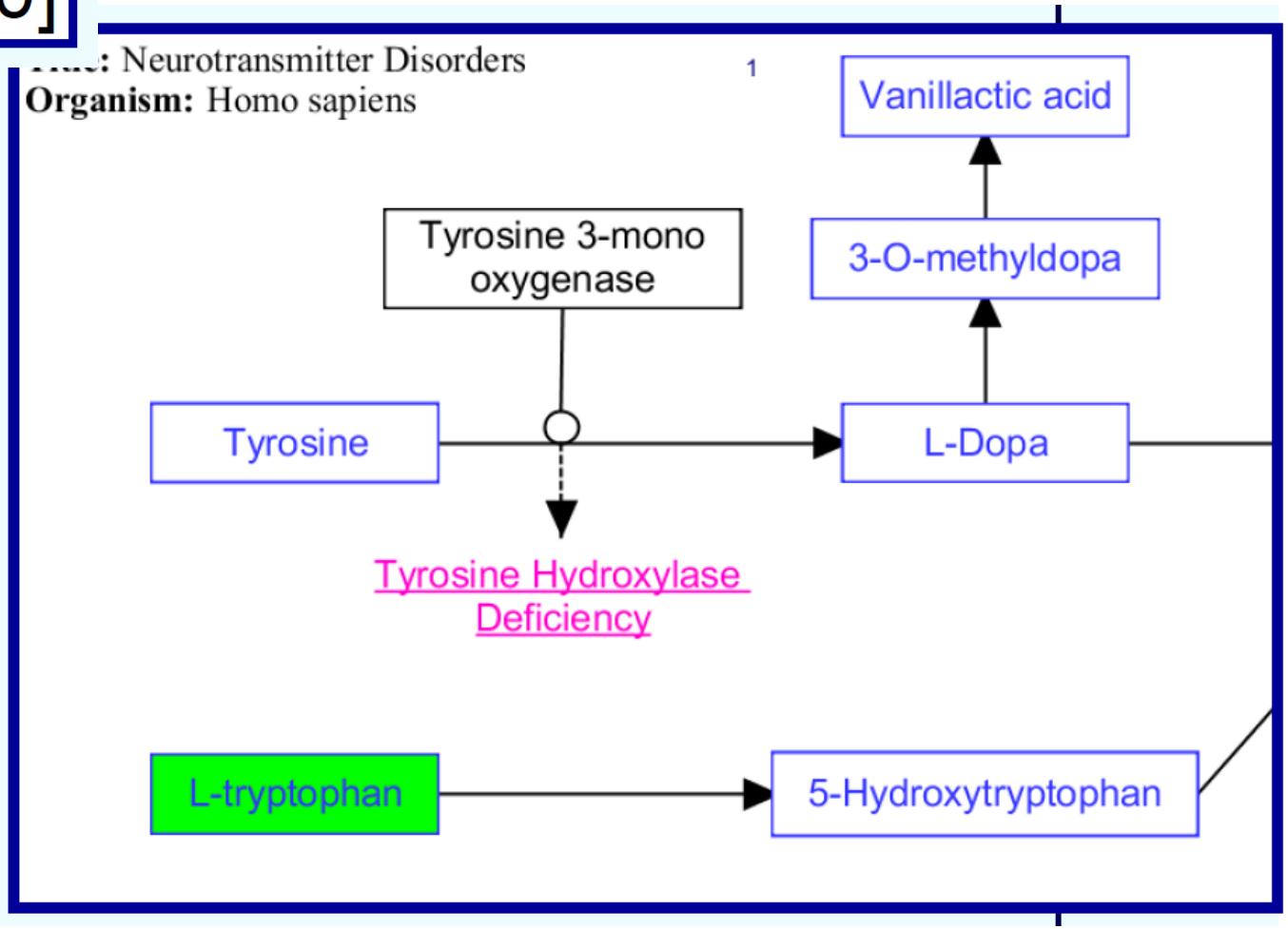
Identifier	Xref	Database
HMDB00538	DB00171	Drugbank
CHEBI:16761	C00019353	KNApSAcK
CHEBI:16828	6305	Pubchem CID
CHEBI:17895	Q188017	Wikidata

II. IDs in pathway(s) [5]

Pathway Analysis [6,10]

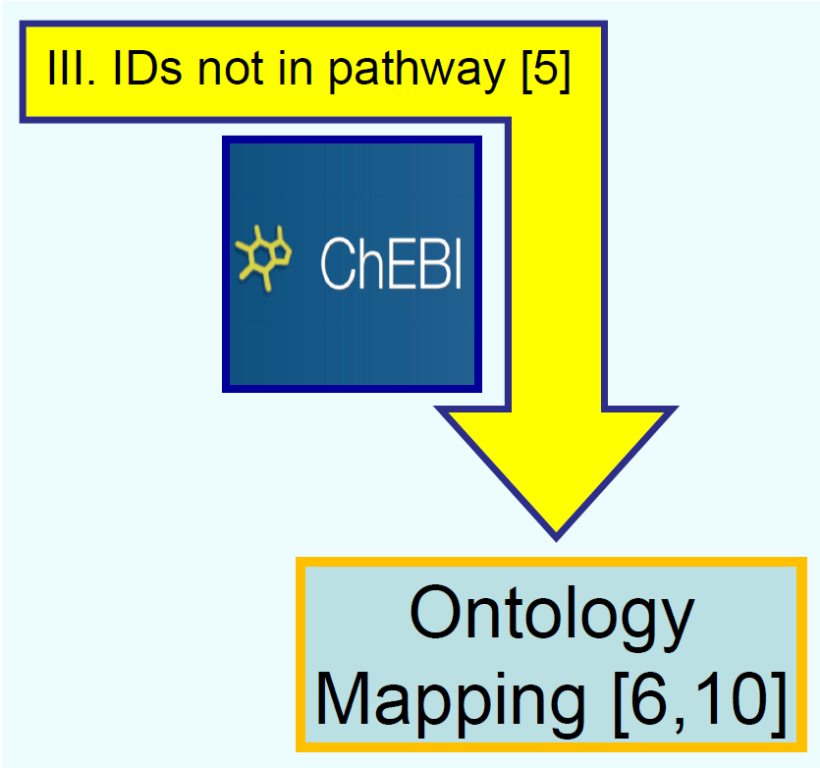
Workflow:

Pathway Analysis [6,10]



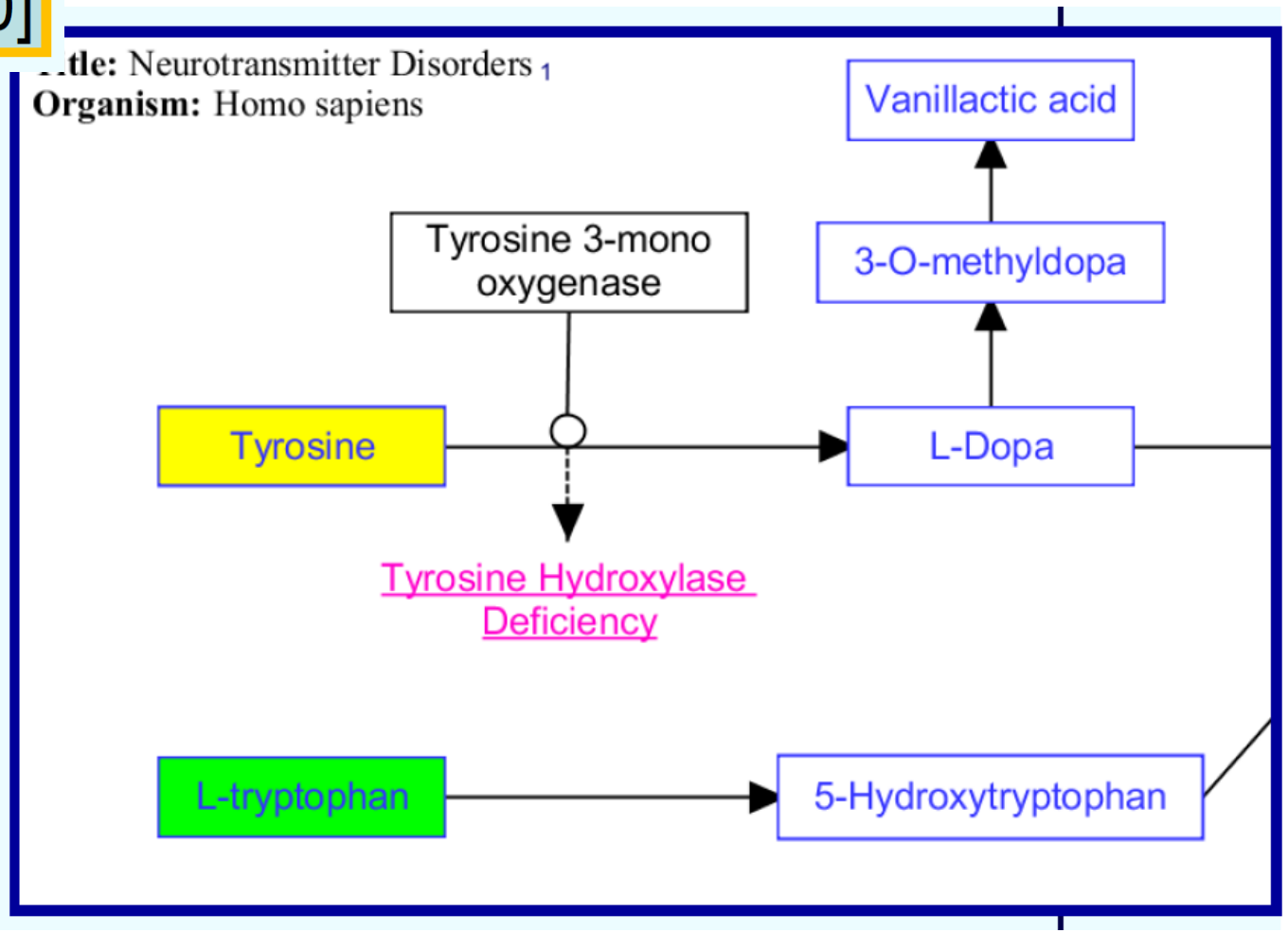
Workflow:

Identifier	Xref	Database
HMDB00538	DB00171	Drugbank
CHEBI:16761	C00019353	KNApSAcK
CHEBI:16828	6305	Pubchem CID
CHEBI:17895	Q188017	Wikidata



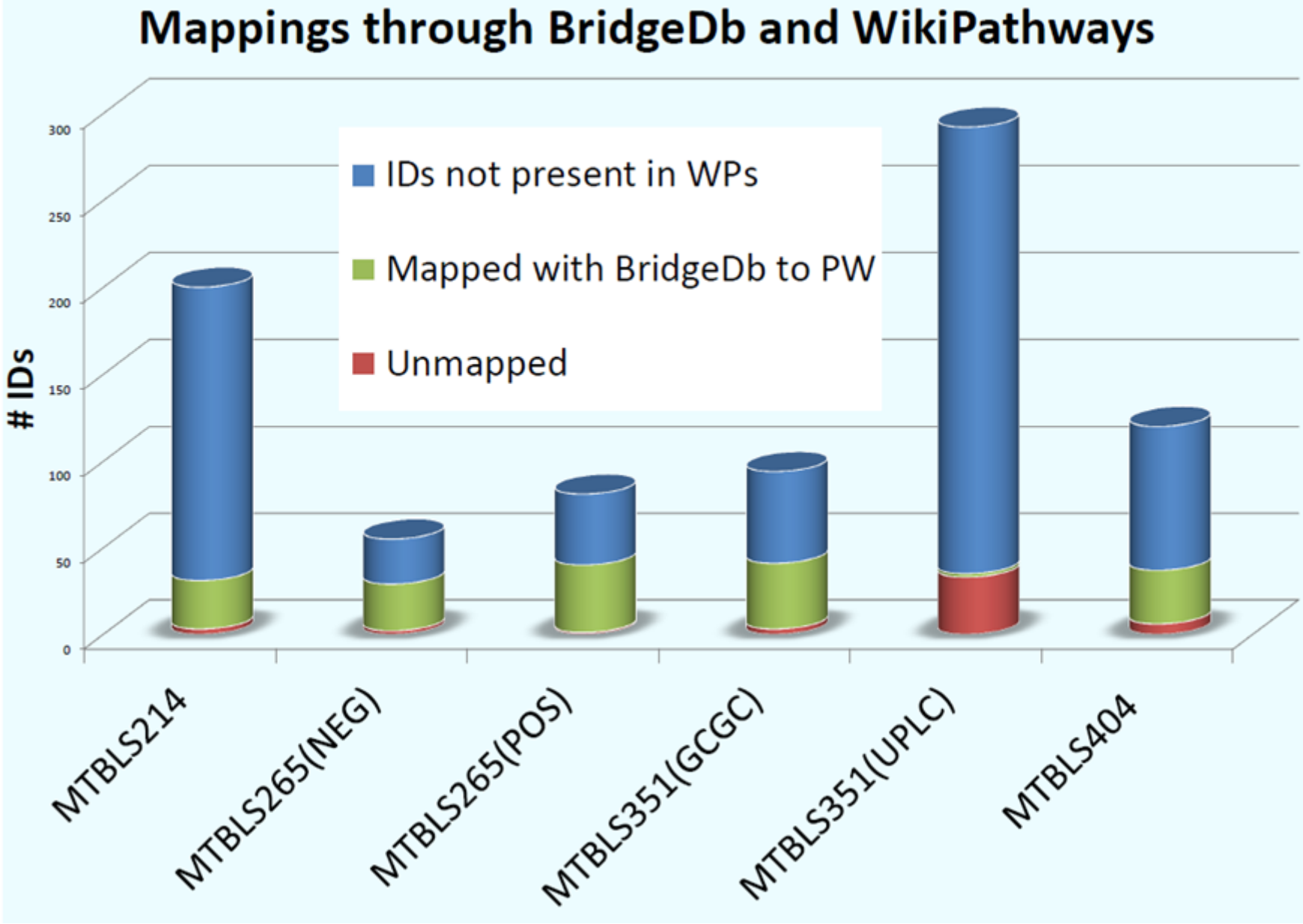
Workflow:

Ontology Mapping [6,10]

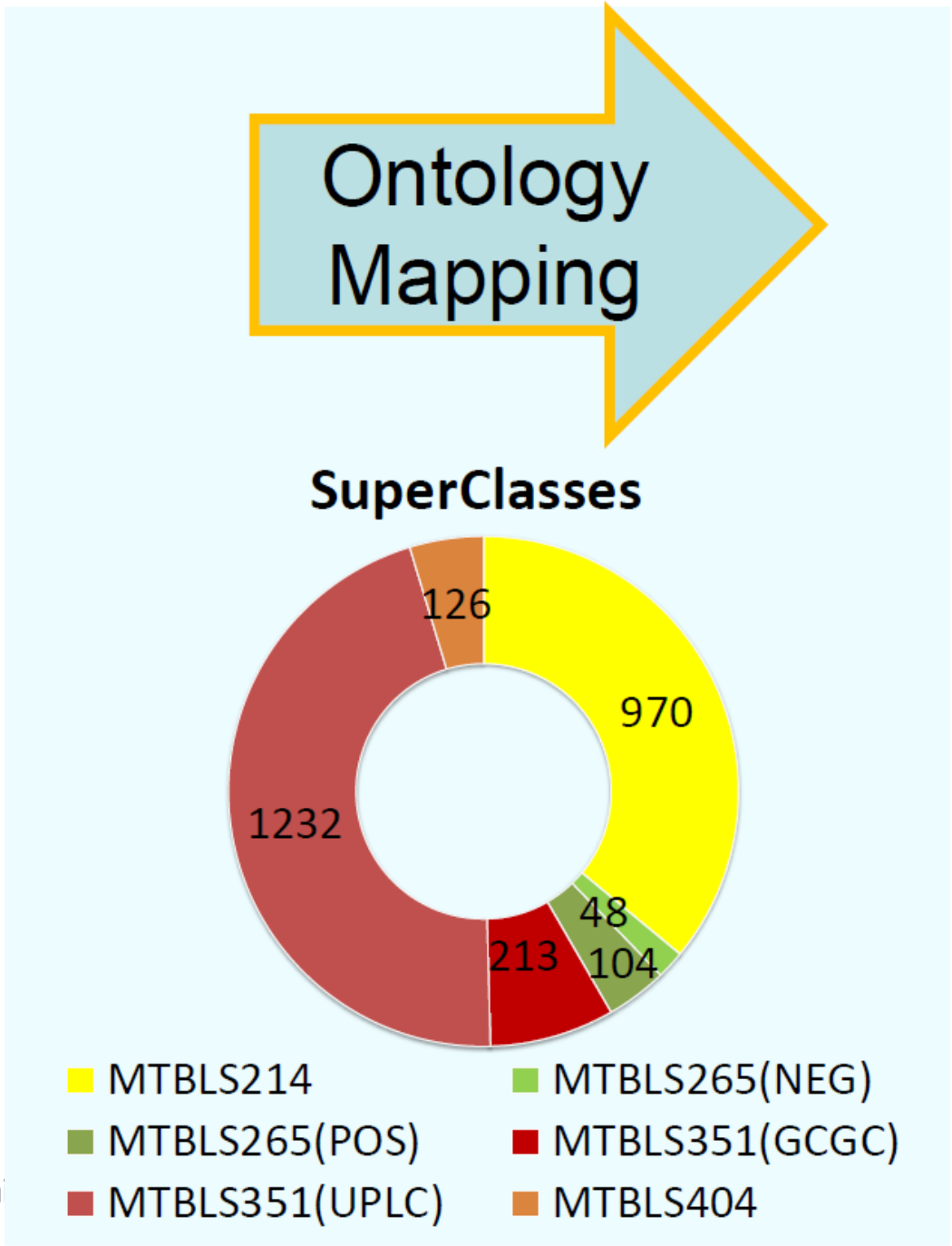


Part 1: first results

Results I + II

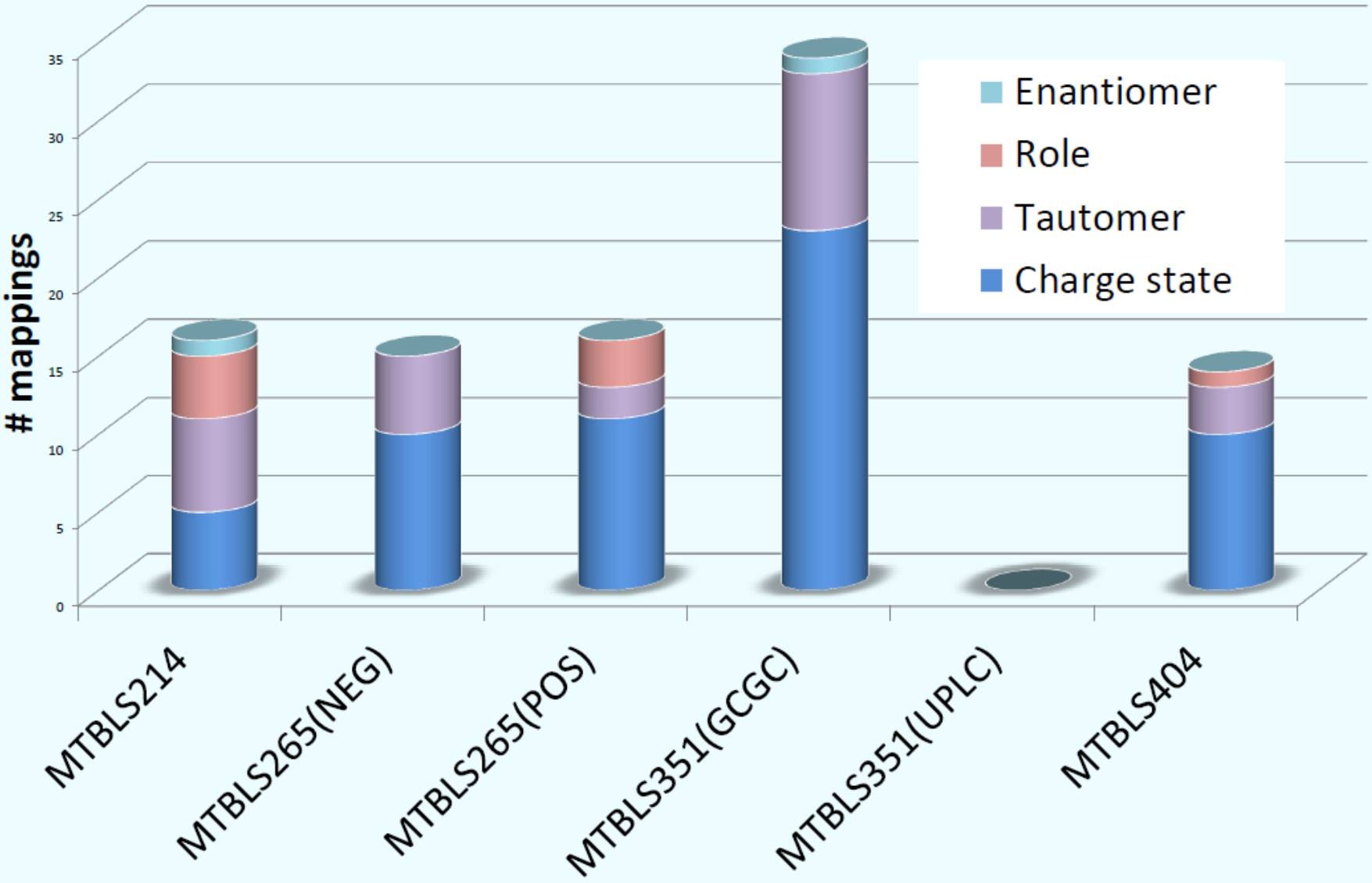

Mappings through BridgeDb and WikiPathways
IDs
300
250
200
150
100
50
0
IDs not present in WPs
Mapped with BridgeDb to PW
Unmapped
MTBLS214
MTBLS265(NEG)
MTBLS265(POS)
MTBLS351(GCGC)
MTBLS351(UPLC)
MTBLS404

Results


Ontology Mapping
SuperClasses
126
970
1232
48
213
104
MTBLS214
MTBLS265(NEG)
MTBLS265(POS)
MTBLS351(GCGC)
MTBLS351(UPLC)
MTBLS404

Results III

Additional Mappings through Ontologies



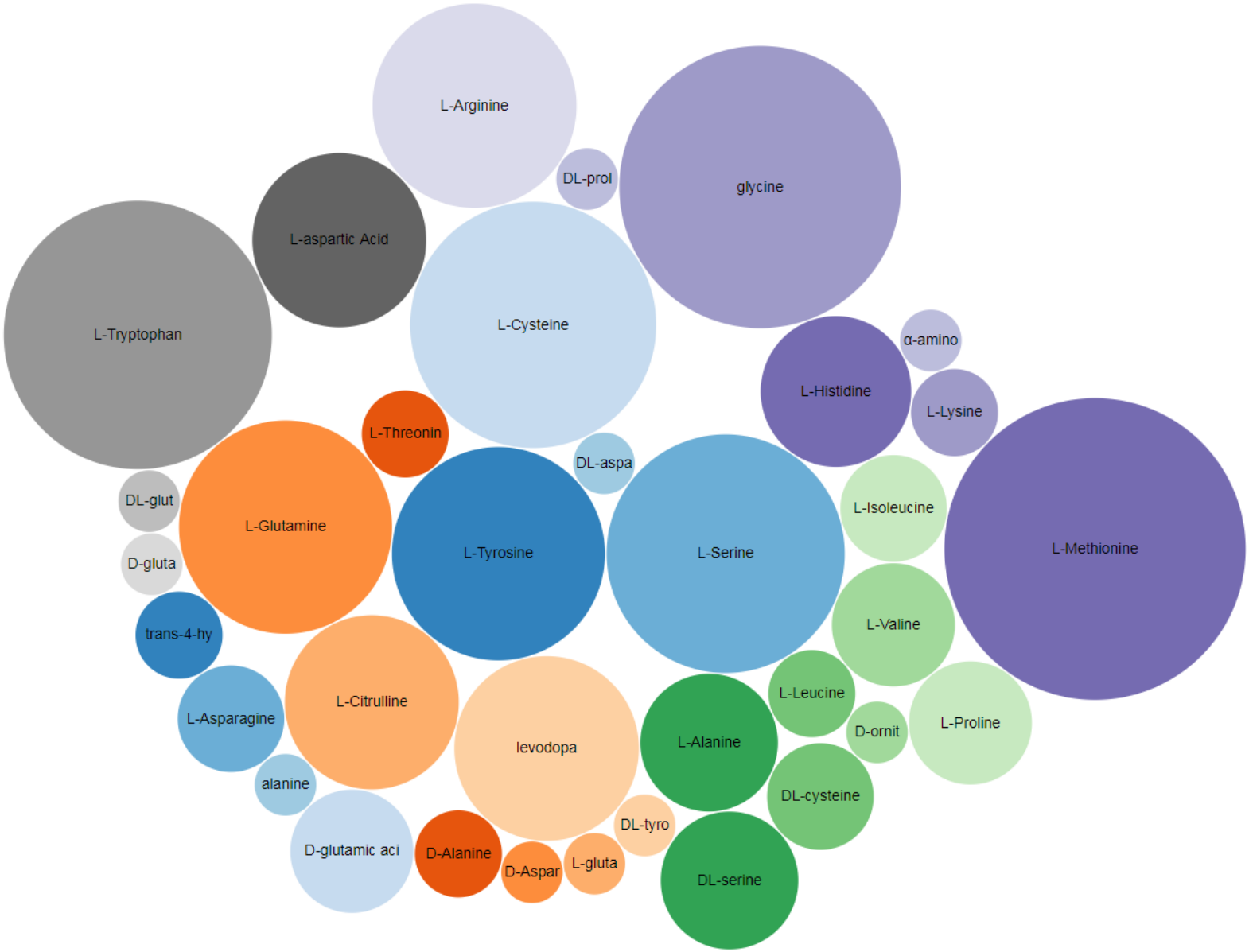
Part 2: testing on compound classes

Tested approach on several major compound classes from Biocrates Assays [1] :

- Amino acids & Biogenic amines
- Acylcarnitines
- Lysophosphatidylcholines
- Sphingomyelins
- Hexoses
- Steroids

https://www.wikipathways.org/index.php/Pathway:WP4234

[1]: Griffiths, William J., et al. "Targeted metabolomics for biomarker discovery." *Angewandte Chemie International Edition* 49.32 (2010)
Biocrates Patent:
https://patents.google.com/patent/US8265877B2/en



http://tinyurl.com/y93cvlxr
http://smallcats4science.blogspot.com/2017/12/more-details-on-amino-acids-from.html

Maastricht University

Results part 2:

Group:	# IDs	# Mapped to PW
Amino acids & Biogenic amines	72	50 (70%)
Acylcarnitines	50	4 (8%)
Lysophosphatidyl-cholines	19	2 (10%)
Sphingomyelins	15	0 (0%)
Hexoses	1	1 (100%)
Steroids	17	15 (88%)

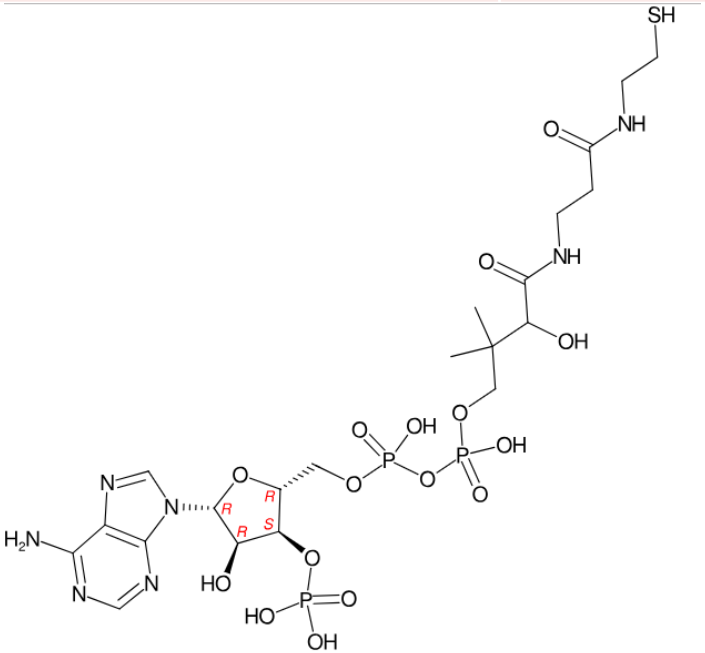
Part 3: Mapping derivative compounds

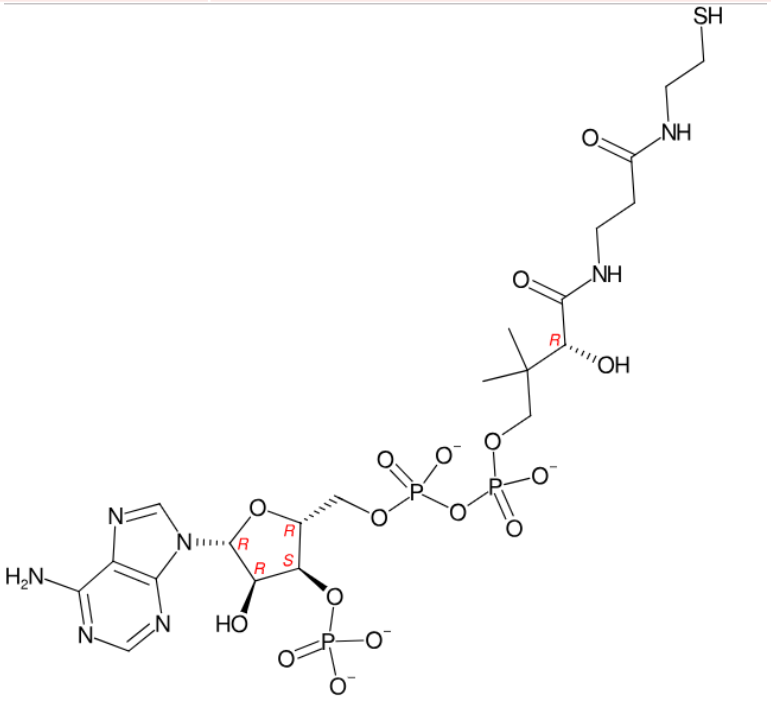
- Tested approach for:
  - CoA-derivatives: substrates for various enzymes
  - glucuronide derivatives: related to the excretion of toxic compounds.

- Find compounds structurally related to original via:
  - InCHIKey
  - PubChem
  - Fragmentation

Results: InCHIKey CoA

mol	molLabel	InChIKey
wd:Q407635	coenzyme A	RGJOEKWQDUBAIZ-DRCCLKDXSA-N
wd:Q27124396	coenzyme A(4-)	RGJOEKWQDUBAIZ-IBOSZNHHSA-J





Results: InCHIKey glucuronide

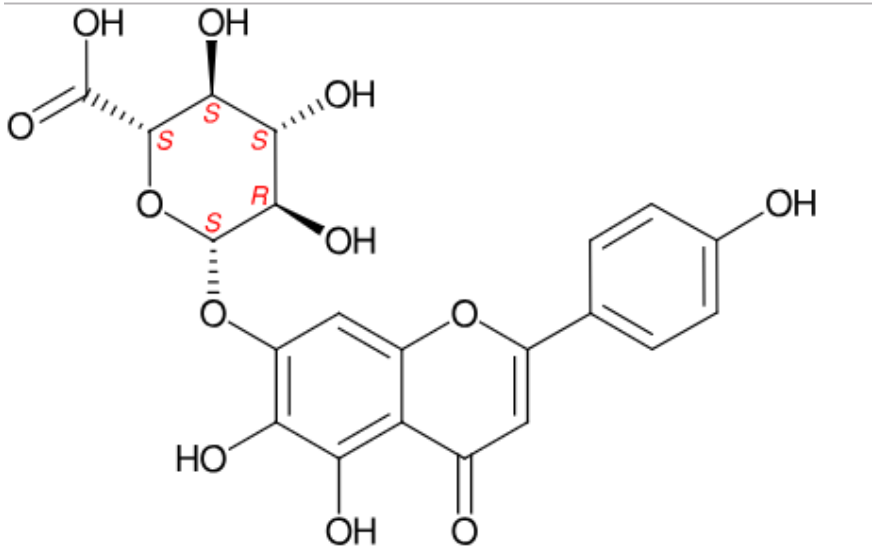
- No matches via InCHIKey for:
  - retinoyl β-glucoronide Q2713934
  - Querciturone Q6871627
  - Morphine-6-glucuronide Q4303469

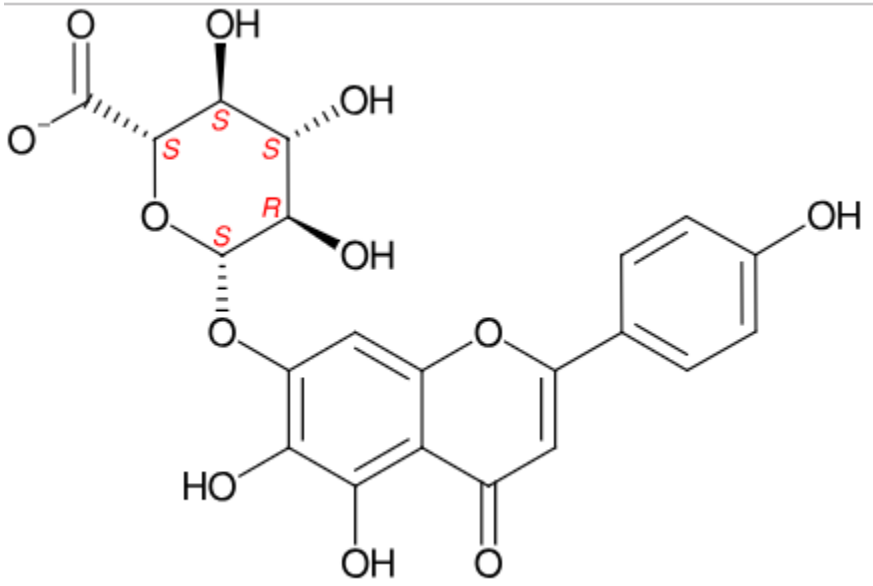
- No link between child-parent compounds in Wd (before I looked at these compounds).

https://en.wikipedia.org/wiki/Glucuronide

Results: InCHIKey Scutellarin

mol	molLabel	InChIKey
wd:Q410712	scutellarin	DJSISFGPUUYILV-ZFORQUDYSA-N
wd:Q27130982	scutellarin(1-)	DJSISFGPUUYILV-ZFORQUDYSA-M





Results: PubChem: CoA(top) and scutellarin (bottom)

5.2 Related Compounds

Same Connectivity	33 records
Same Parent, Connectivity	114 records
Same Parent, Exact	55 records
Mixtures, Components, and Neutralized Forms	58 records
Similar Compounds	3306 records

5.2 Related Compounds

Same Connectivity	17 records
Same Parent, Connectivity	23 records
Same Parent, Exact	7 records
Mixtures, Components, and Neutralized Forms	7 records
Similar Compounds	15167 records
Similar Conformers	225 records

https://pubchem.ncbi.nlm.nih.gov/compound/87642 CoA
https://pubchem.ncbi.nlm.nih.gov/compound/185617 Scutellarin

Results: Fragmentation

Still in progress ☺

Future Directions:

- Import ChEBI database into Neo4j
- Add mappings to pathway information
- Query metabolites in dataset (with set length for child-parent relationships)
- Visualise mappings automatically in Cytoscape

Question, discussion

References:
[1] Hastings J. *et al.* Nucl. Acids Res. 2016
[2] Slenter D.; *et al.* Nucl. Acids Res. 2018
[3] Haug K; *et al.* Nucl. Acids Res. 2013
Data for workflow:
[4] BridgeDb metabolites mapping file DOI: 10.6084/m9.figshare.6260003.v1
[5] WikiPathways pathways accession number: http://data.wikipathways.org/20180410/

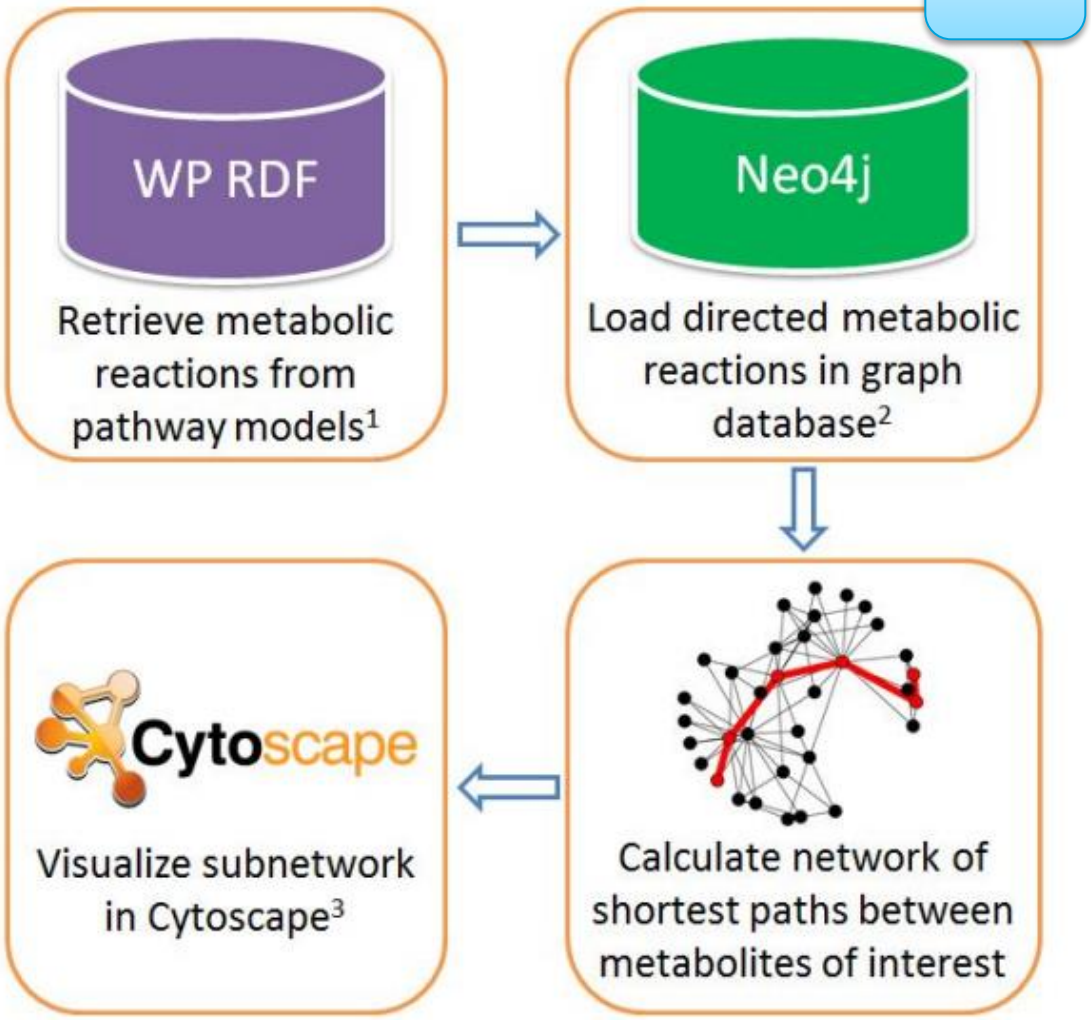
Metabolomics datasets from Metabolights:
[6] MTBLS124, DOI: 10.1371/journal.pgen.1004801
[7] MTBLS265, DOI: 10.1073/pnas.1603023113
[8] MTBLS351, DOI: 10.1038/nature18646
[9] MTBLS404, DOI: 10.1021/acs.jproteome.5b00354
Pathway analysis for MTBLS214 on:
[10] WP4220_r97366, neurotransmitter disorders; Friesacher A., Slenter D., Willighagen E.

Network approach: AIM

- Directed network of metabolites from pathway knowledge bases
- Calculate sub-network between active metabolites
- Visualise directed paths
- Interpret metabolomics datasets

Workflow

1. Directed metabolic reactions in human pathway models are retrieved from the WikiPathways RDF[1].
2. Those interactions are stored in the graph database Neo4j[2].
3. Using the Cypher query language the shortest paths between metabolites of interests are calculated.
4. Finally, the resulting subnetwork is visualized in Cytoscape[3].



Maastricht University

[1] **WikiPathways RDF:** http://sparql.wikipathways.org (released July 10, 2017)
Kutmon *et al* (2016) doi:10.1093/nar/gkv1024, Waagmeester *et al.* (2016) doi: 10.1371/journal.pcbi.1004989

[2] **Neo4j**: https://neo4j.com/

[3] **Cytoscape**: http://cytoscape.org/
Shannon *et al.* (2003) doi: 10.1101/gr.1239303

Building up the network (HS)

Database:	#nodes	#edges	#nodes – side	#edges – side	Reduced nodes (%)	Reduces edges (%)
WikiPathways & Reactome (Wikidata)	1667	16216	1607	3027	3.6	81.3
WikiPathways & Reactome (ChEBI)	2395	28196	2329	6746	2.8	76.1
WikiPathways & Reactome (HMDB)	2156	23576	2071	6365	3.9	73

What about other species?
WikiPathways & Reactome (Wikidata)

Species:	#nodes – side	#edges – side
Homo sapiens	1607	3027
Bos taurus	525	769
Danio rerio	104	140
Mus musculus	340	491
Arabidopsis thaliana	93	103
Saccharomyces cerevisiae	212	370
E.coli	12	10

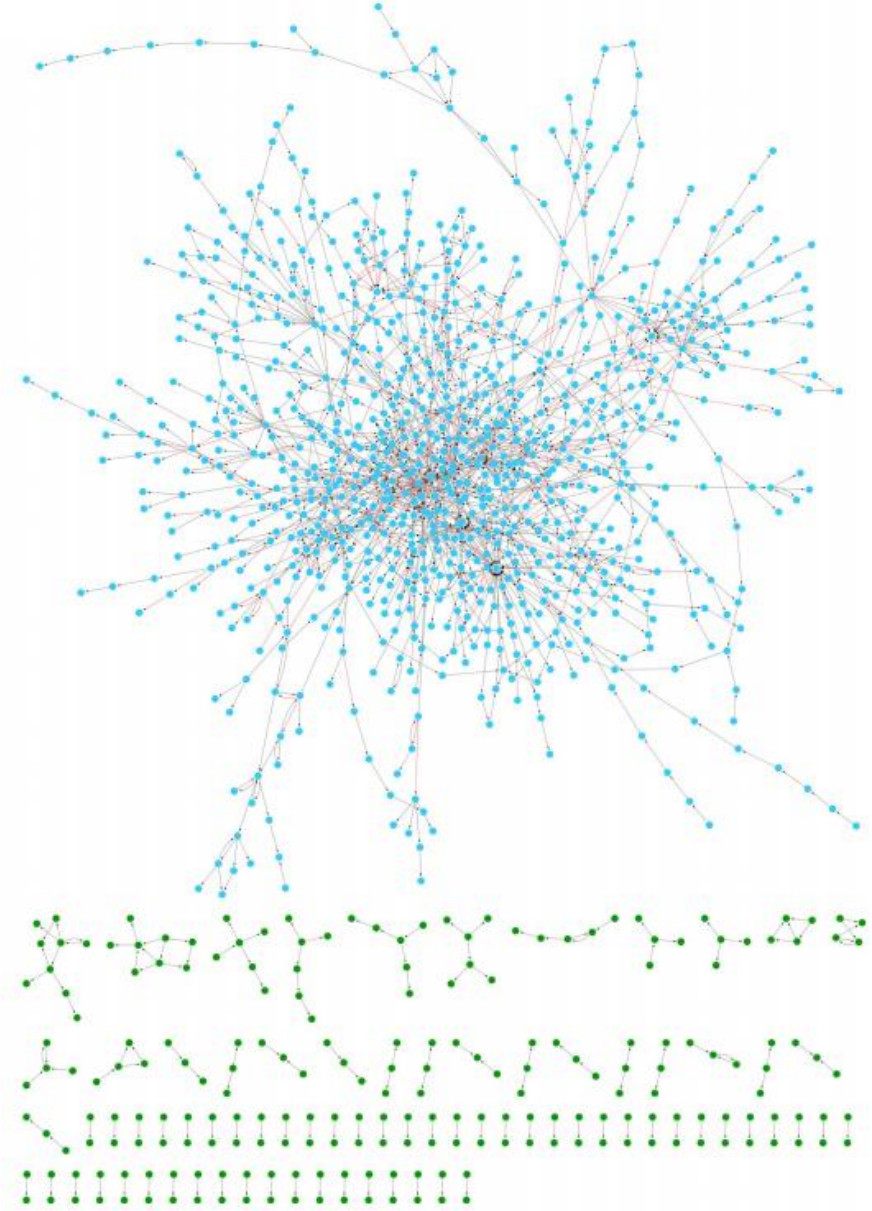
Problems that needed fixing:

- Use one unique identifier per node
- Annotate side metabolic reactions within the network
- Annotate “unapproved” pathway reactions within the network
- Add more IDs to one node for easy querying
- Update Reactome data for final visualisation

Biological role					
Electron donor/receiver.		Energy donor/receiver.		Miscellaneous, relevant for various metabolic reactions.	
Identifier	**Name**	**Identifier**	**Name**	**Identifier**	**Name**
Q5203615	O2	Q80863	ATP	Q307434	S-adenosyl-L-homocysteine
Q506710	H+	Q185253	ADP	Q201312	S-adenosyl-L-methioninate
Q20856948	Na+ (redirected to Q3154110)	Q318369	AMP	Q407635	Coenzyme-A
Q3154110	Na+	Q422582	GDP	Q715317	Acetyl coenzyme a
Q283	H2O	Q392227	GTP		
Q1997	CO2	Q26987754	NADP+		
Q177811	PO4 3-	Q26841327	NADPH		
Q411092	Pyrophos-phoric acid	Q26987253	NAD+		
Q190901	ammonium cation	Q26987453	NADH		
		Q27102690	FADH2		

Side metabolites

DIRECTED NETWORK VISUALISATION

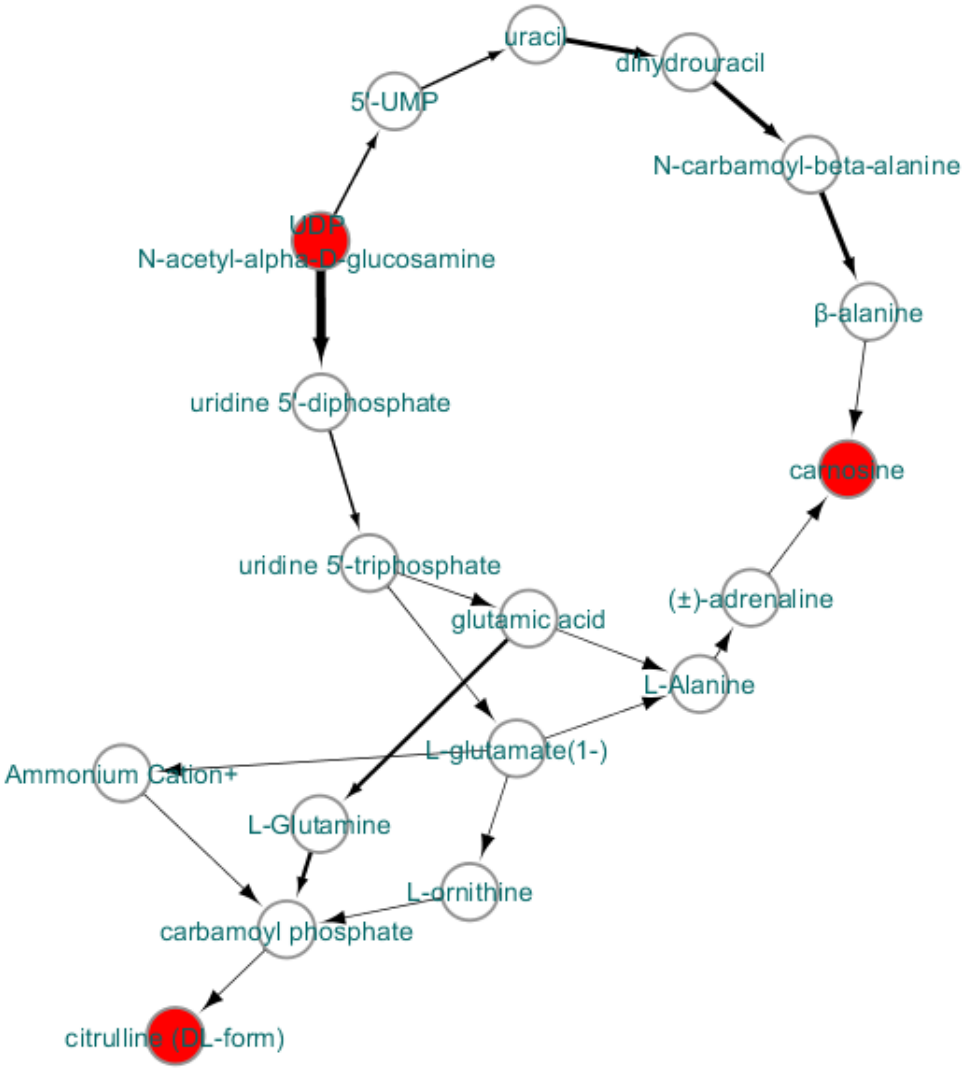


Dataset 1:

- MetaboLights dataset (MTBLS265) [4]
- Metabolic profiles (LC-MS) in **blood** samples of 15 young (29 ± 4 y of age) and 15 elderly (81 ± 7 y of age) individuals.

[4] **MetaboLights dataset:** http://www.ebi.ac.uk/metabolights/MTBLS265, Chaleckis *et al.* (2016) doi: 10.1073/pnas.1603023113

Dataset 1:

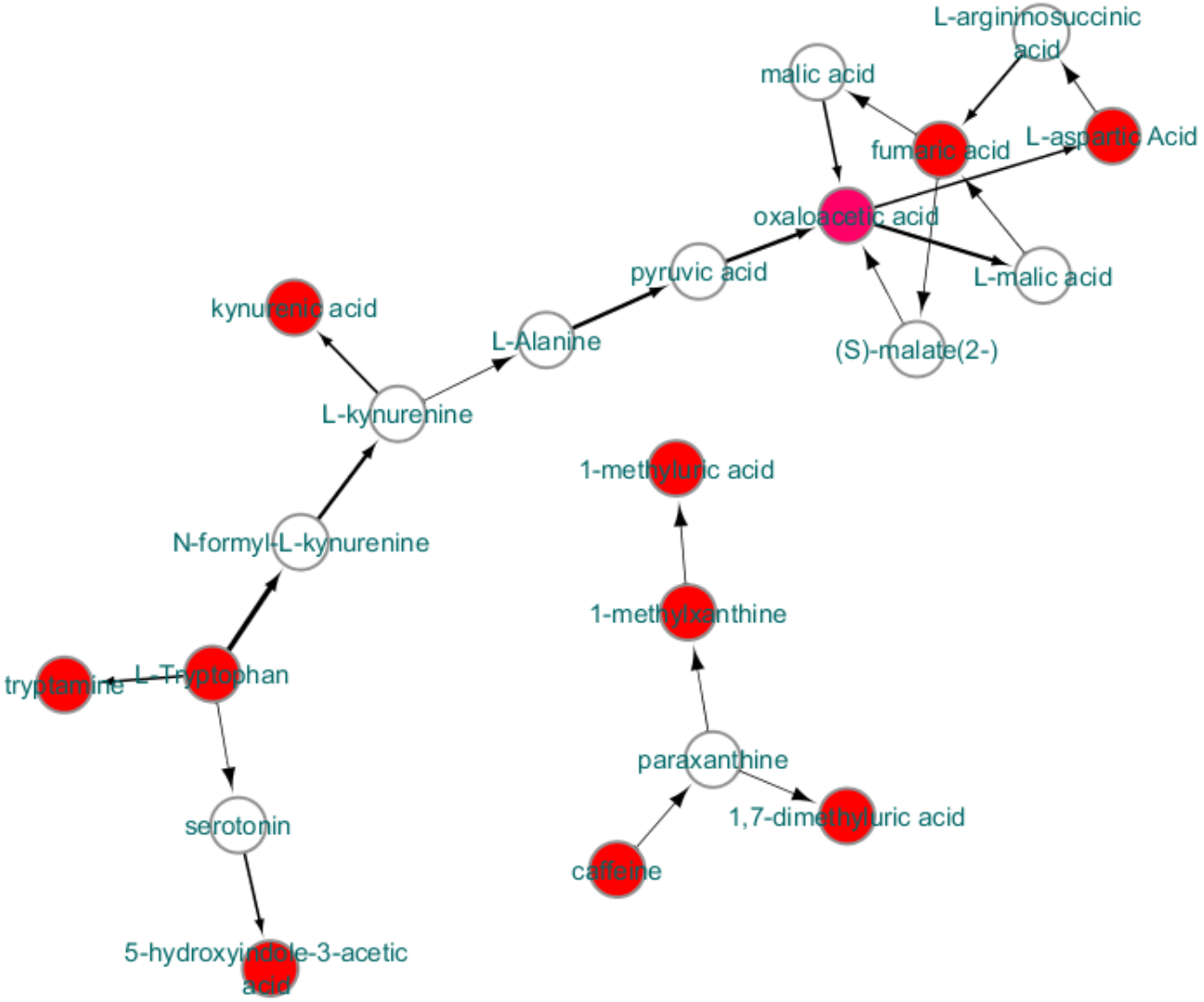

uracil
dihydrouracil
5'-UMP
N-carbamoyl-beta-alanine
UDP
N-acetyl-alpha-D-glucosamine
β-alanine
uridine 5'-diphosphate
carnosine
uridine 5'-triphosphate
(±)-adrenaline
glutamic acid
L-Alanine
L-glutamate(1-)
Ammonium Cation+
L-Glutamine
L-ornithine
carbamoyl phosphate
citrulline (DL-form)

Dataset 2:

- MetaboLights dataset (MTBLS404) [5]
- Metabolic profiles (LC-HRMS) in **urine** samples of 183 individuals, ages 40.9 ± 10.3 years.

Metabolights dataset: https://www.ebi.ac.uk/metabolights/MTBLS404
Thévenot et al. (2015) 10.1021/acs.jproteome.5b00354

Dataset 2:



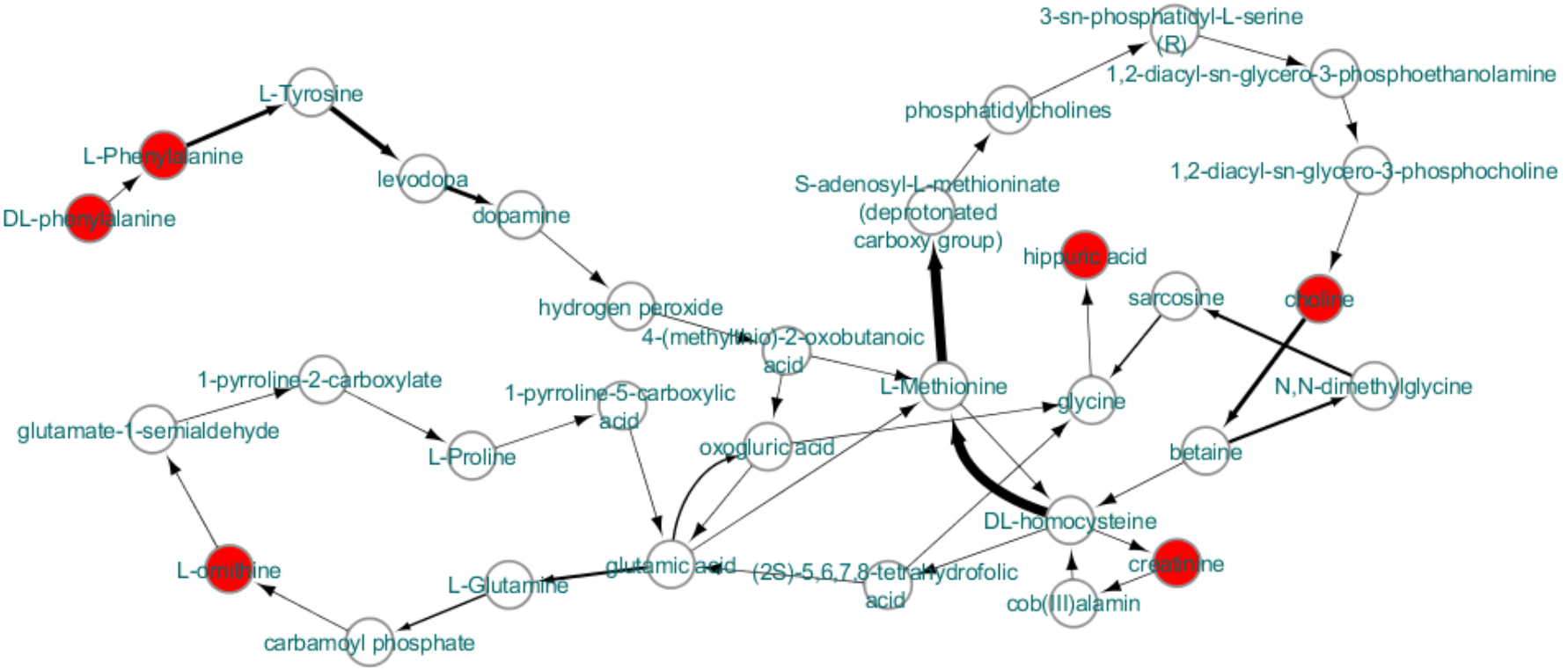
Thévenot et al. (2015) 10.1021/acs.jproteome.5b00354

Dataset 3:

- Dataset from Rist et al. [6]
- Metabolic profiles ((GC-GC)-MS, GC-MS, LC-MS/MS, 1H_NMR) in **urine** and blood samples of 301 individuals, ages 47.5 ± 17.1 years.

[6] Rist et al. (2017) https://doi.org/10.1371/journal.pone.0183228

Dataset 3:


3-sn-phosphatidyl-L-serine
(R)
1,2-diacyl-sn-glycero-3-phosphoethanolamine
phosphatidylcholines
L-Tyrosine
L-Phenylalanine
levodopa
1,2-diacyl-sn-glycero-3-phosphocholine
DL-phenylalanine
dopamine
S-adenosyl-L-methioninate
(deprotonated
carboxy group)
hippuric acid
sarcosine
choline
hydrogen peroxide
4-(methylthio)-2-oxobutanoic
acid
L-Methionine
1-pyrroline-2-carboxylate
1-pyrroline-5-carboxylic
acid
glycine
N,N-dimethylglycine
glutamate-1-semialdehyde
L-Proline
oxoglutaric acid
betaine
DL-homocysteine
L-ornithine
glutamic acid
(2S)-5,6,7,8-tetrahydrofolic
acid
creatinine
L-Glutamine
cob(III)alamin
carbamoyl phosphate

CONCLUSION

- Calculation of directed subnetwork connecting active metabolites is possible with presented workflow

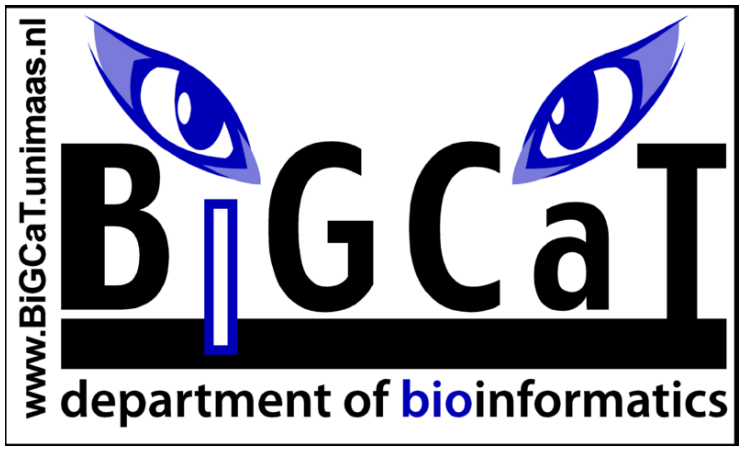
- Neo4j and Cytoscape allow computational calculation for larger networks and advanced visualisation

FUTURE PERSPECTIVE

- Add more pathway knowledge bases (now WikiPathways and Reactome, could add KEGG in the future?)
- Create app for direct visualisation with Cytoscape from Neo4j (first tests have been run)
- Allow for integration with other omics data sources, such as proteomics and transcriptomics

Acknowledgements, questions, discussion

- Martina Kutmon
- Jonathan Melius
- Ryan Miller
- Georg Summer
- Chris T Evelo
- Egon L Willighagen



Maastricht University

Find the original poster @
https://figshare.com/articles/Poster_Visualizing_metabolomics_data_in_directed_biological_networks/5234851