Biomarkers in rare diseases

J. Landman
08-07-2019
Supervisor: dr. E.L. Willighagen

Maastricht University

Introduction

- Inborn errors of Metabolism (IEM) [1]
  - Group of 500 rare genetic diseases
  - Estimated incidence of 1/2500
  - Caused by deficiency or abnormal functioning of an enzyme

- Metabolomic approach [2]

- Biomarkers[3]

1: Tebani A, Abily-Donval L, Afonso C, Marret S, Bekri S. Clinical Metabolomics: The New Metabolic Window for Inborn Errors of Metabolism Investigations in the Post-Genomic Era. Int J Mol Sci. 2016 Jul 20;17(7):E1167
2: El-Hattab AW. Inborn errors of metabolism. Clin Perinatol. 2015 Jun;42(2):413-39
3: Pheng B, Li H, Peng XX. Functional metabolomics: from biomarker discovery to metabolome reprogramming. Protein Cell. 2015 Sep;6(9):628-37

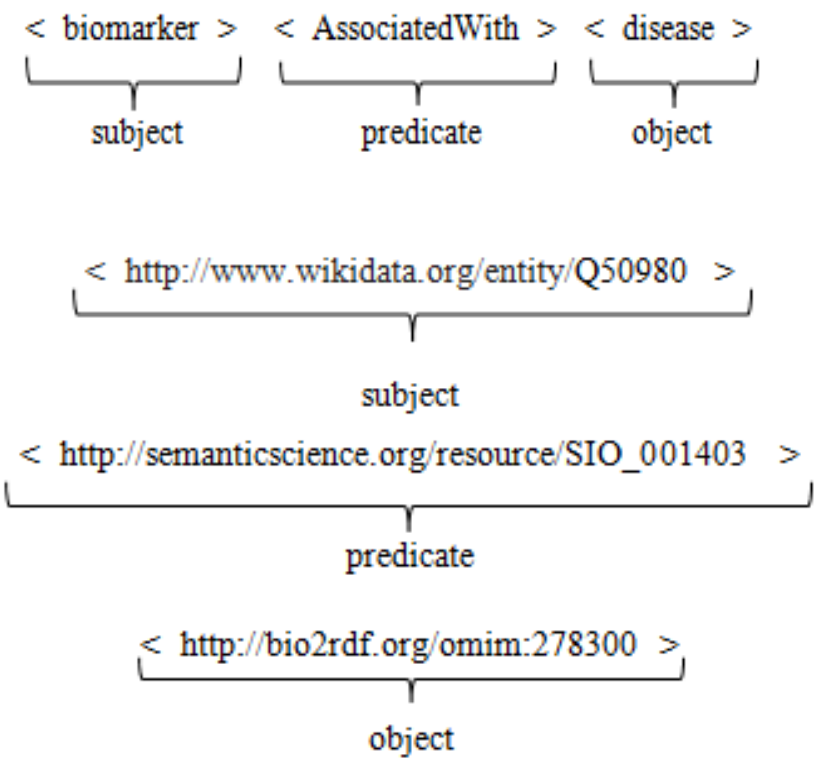
Introduction

- Biological pathways on WikiPathways (WP)
  - Machine readable
  - Resource Description Framework (RDF)
  - Graphical Pathway Markup Language (GPML)

- By using RDF and SPARQL, information from literature and biological pathway information can be combined to examine the link between disease and biomarker on pathway level

Introduction

- Resource Description Framework (RDF) [4]

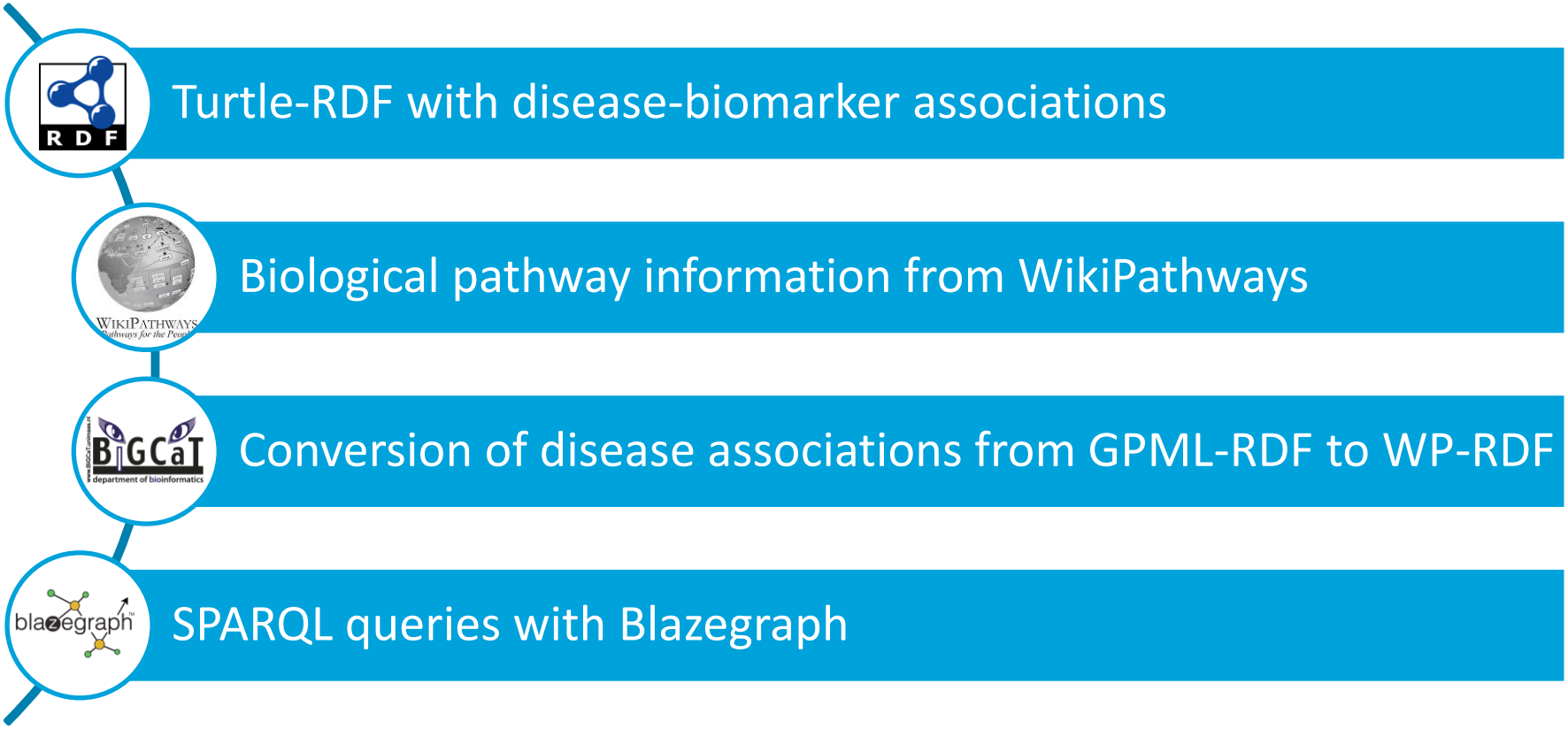


4: Lee J, Kasperovics R, Han WS, Lee JH, Kim MS, Cho H. An efficient algorithm for updating regular expression indexes in RDF databases. Int J Data Min Bioinform. 2015;11(2):205-22

Research questions

- Can we use biological pathway information to check if biomarkers are linked to diseases?
  - Are these biomarkers directly connected to a disease?
  - Can we find the biomarkers in the pathway that is related to the disease?

Workflow


RDF
Turtle-RDF with disease-biomarker associations
WIKIPATHWAYS
Biological pathway information from WikiPathways
BiGCaT
department of bioinformatics
Conversion of disease associations from GPML-RDF to WP-RDF
blazegraph
SPARQL queries with Blazegraph

Workflow - 1

Turtle-RDF with disease-biomarker associations

Input

- "Physician's Guide to the Diagnosis, Treatment and Follow-Up of Inherited Metabolic Disorders" [5]
  - ch 1. Disorders of Phenylalanine and Tetrahydrobiopterin (BH4) Metabolism
  - ch 11. Vitamin B6-Dependent and Responsive Disorders
  - ch 20. Disorders of the Krebs Cycle
  - ch 31. Neurotransmitter Disorders
  - ch 41. Purine and Pyrimidine Disorders

Output

- RDF file with 50 diseases and 58 unique biomarkers
- 103 biomarker-disease associations

5: Blau, N., Duran, M., Gibson, K.M., Dionisi-Vici, C.; ''Physician's Guide to the Diagnosis, Treatment, and Follow-Up of Inherited Metabolic Diseases''; ISBN 978-3-642-40337-8,, 2014

Workflow - 1

Turtle-RDF with disease-biomarker associations

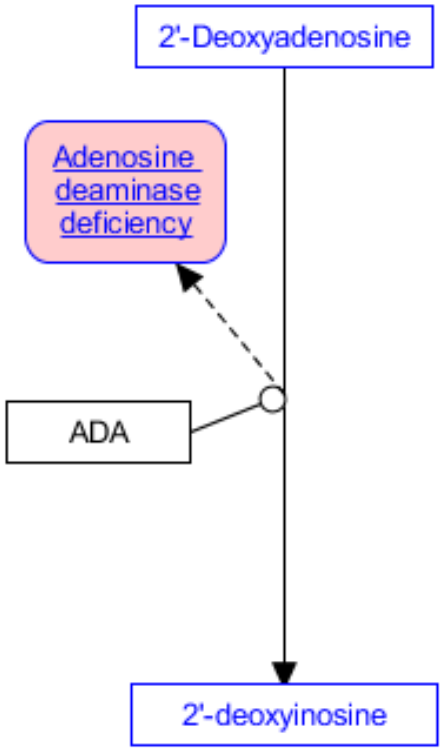
```
# disease labels Chapter41 purine metabolism

omim:311850 a ncit:C7057 ; rdfs:label "Phosphoribolsyl pyrophosphate synthetase 1 defects" .
omim:300661 a ncit:C7057 ; rdfs:label "Phosphoribosyl pyrophosphate synthetase 1 superactivity" .
omim:103050 a ncit:C7057 ; rdfs:label "Adenylosuccinate lyase deficiency" .
omim:608688 a ncit:C7057 ; rdfs:label "AICAR transformylase/IMP cyclohydrolase deficiency" .
omim:102770 a ncit:C7057 ; rdfs:label "Adenosine monophosphate deaminase deficiency" .
omim:146690 a ncit:C7057 ; rdfs:label "Inosine monophosphate dehydrogenase deficiency" .
omim:608958 a ncit:C7057 ; rdfs:label "Adenosine deaminase deficiency " .
omim:613179 a ncit:C7057 ; rdfs:label "Purine nucleoside phosphorylase deficiency" .
omim:278300 a ncit:C7057 ; rdfs:label "Xanthine dehydrogenase deficiency" .
omim:603592 a ncit:C7057 ; rdfs:label "Combined xanthine oxydase and aldehyde oxidase deficiency" .
omim:102600 a ncit:C7057 ; rdfs:label "Adenine phosphoribosyl transferase deficiency" .
omim:300322 a ncit:C7057 ; rdfs:label "Hypoxanthine guanine phosphoribolsyl transferase deficiency" .
```

```
wd:Q105522 a chebi:25367 ; rdfs:label "uric acid" .
:assoc90 a association: ; refersTo: chapter41:, wd:Q105522, omim:311850 .
:assoc91 a association: ; refersTo: chapter41:, wd:Q105522, omim:300661 .
:assoc92 a association: ; refersTo: chapter41:, wd:Q105522, omim:103050 .
:assoc93 a association: ; refersTo: chapter41:, wd:Q105522, omim:613179 .
:assoc94 a association: ; refersTo: chapter41:, wd:Q105522, omim:278300 .
```

Workflow - 2

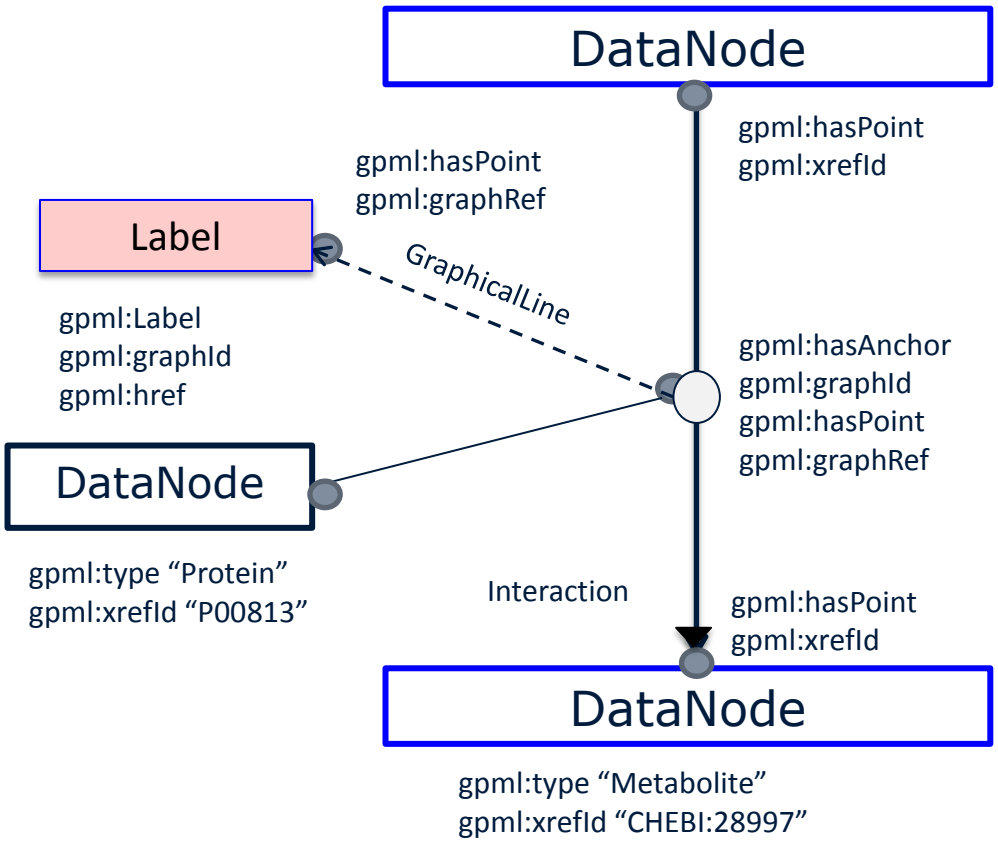
Biological pathway information from WikiPathways

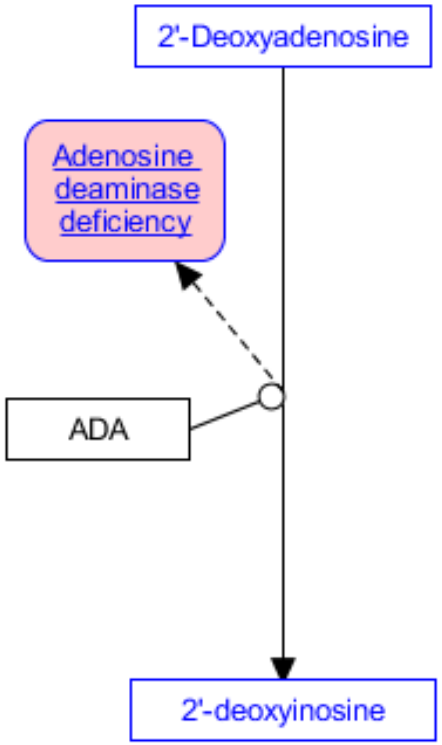


- File with the Pathways corresponding to the Chapters of Blau.
- Convert InteractionLine to disease into GraphicalLine

Workflow - 3

Conversion of disease associations from GPML-RDF to WP-RDF





https://www.wikipathways.org/index.php/Pathway:WP4224

https://jenkins.bigcat.unimaas.nl/job/GPML%20to%20GPML%20+%20WP%20RDF%20(all)/ws/WP2RDF/output/gpml/Human/WP4224.ttl/*view*/

Workflow - 3

Conversion of disease associations from GPML-RDF to WP-RDF

```
<http://rdf.wikipathways.org/Pathway/WP4224_r104773/Label/adf13> ;
a gpml:Label ;
dcterms:isPartOf <http://rdf.wikipathways.org/Pathway/WP4224_r104773> ;
gpml:graphId "adf13"^^xsd:string ;
gpml:href "http://omim.org/102700"^^xsd:string ;
gpml:shapeType "RoundedRectangle"^^xsd:string ;
gpml:textlabel "Adenosine deaminase deficiency"^^xsd:string ;
```

```
<http://rdf.wikipathways.org/Pathway/WP4224_r104773/Label/adf13> rdf:type gpml:Label
<http://rdf.wikipathways.org/Pathway/WP4224_r104773/Label/adf13> gpml:graphId "adf13"
<http://rdf.wikipathways.org/Pathway/WP4224_r104773/Label/adf13> gpml:href "http:omim.org/102700"
<http://rdf.wikipathways.org/Pathway/WP4224_r104773/Label/adf13> rdfs:label "Adenosine deaminase deficiency"
```

Workflow - 3

Conversion of disease associations from GPML-RDF to WP-RDF

- SPARQL query to find all pathways that contain graphical lines connected to a disease label

```
SELECT DISTINCT
    (STR(?wpidLit) AS ?wpid)
    (STR(?labelLit) AS ?label)
    (STR(?hrefLit) AS ?href)
    (STR(?arrow1) AS ?int)

WHERE
{
  FILTER (CONTAINS(STR(?hrefLit), "omim"))
  ?arrow1 a wp:Interaction ; dcterms:isPartOf ?wppathway .
  ?arrow2 a gpml:Interaction ; ^wp:isAbout ?arrow1 .
  ?anchor dcterms:isPartOf ?arrow2 .
  ?anchor a gpml:Anchor ; gpml:graphId ?href3 .
  ?point3 gpml:graphRef ?href3 .
  ?point dcterms:isPartOf ?gpmlpathway ;
              gpml:graphRef ?id ;
         dcterms:isPartOf ?interaction .
  ?interaction gpml:hasPoint ?point3 .
  ?labelNode gpml:textlabel ?labelLit ; gpml:href ?hrefLit .
  ?labelNode gpml:graphId ?id .
  ?gpmlpathway gpml:hasLabel ?labelNode .
  ?wppathway a wp:Pathway ;
             wp:isAbout ?gpmlpathway  ;
             dcterms:identifier ?wpidLit .
}
```

Workflow - 3

Conversion of disease associations from GPML-RDF to WP-RDF

wpid	label	href	int
WP4224	Purine nucleoside phosphorylase deficiency	https://omim.org/entry/613179	http://rdf.wikipathways.org/Pathway/WP4224_r104773/WP/Interaction/id88da3202
WP4224	Adenine phospho-ribosyltransferase deficiency	https://omim.org/entry/614723	http://rdf.wikipathways.org/Pathway/WP4224_r104773/WP/Interaction/id455e3501
WP4224	Purine nucleoside phosphorylase deficiency	http://omim.org/613179	http://rdf.wikipathways.org/Pathway/WP4224_r104773/WP/Interaction/id905b1edd
WP4224	AICAr transformylase/ IMP cyclohydrolase deficiency	http://omim.org/608688	http://rdf.wikipathways.org/Pathway/WP4224_r104773/WP/Interaction/id8d646469
WP4224	Purine nucleoside phosphorylase deficiency	https://omim.org/entry/613179	http://rdf.wikipathways.org/Pathway/WP4224_r104773/WP/Interaction/id488c9916
WP4224	IMP dehydrogenase deficiency	http://omim.org/146690	http://rdf.wikipathways.org/Pathway/WP4224_r104773/WP/Interaction/id4effd6ce

Workflow - 3

Conversion of disease associations from GPML-RDF to WP-RDF

New ttl file with interactions with their associated diseases

```
<http://rdf.wikipathways.org/Pathway/WP4224_r104773/WP/Interaction/idafa86019> associatedWith: omim:102700 .
omim:102700 a obo:DOID_4 ; rdfs:label "Adenosine  deaminase deficiency" .
<http://rdf.wikipathways.org/Pathway/WP4523_r104251/WP/Interaction/idb880c49> associatedWith: omim:124080 .
omim:124080 a obo:DOID_4 ; rdfs:label "Corticosterone methyl oxidase deficiency" .
<http://rdf.wikipathways.org/Pathway/WP4224_r104773/WP/Interaction/id2eb2cadd> associatedWith: omim:278300 .
omim:278300 a obo:DOID_4 ; rdfs:label "Xanthinuria, Type I" .
<http://rdf.wikipathways.org/Pathway/WP4271_r104475/WP/Interaction/idbc82c43> associatedWith: omim:250940 .
omim:250940 a obo:DOID_4 ; rdfs:label "Methionine Synthase Deficiency cblG" .
<http://rdf.wikipathways.org/Pathway/WP4224_r104773/WP/Interaction/id905b1edd> associatedWith: omim:613179 .
omim:613179 a obo:DOID_4 ; rdfs:label "Purine nucleoside phosphorylase deficiency" .
<http://rdf.wikipathways.org/Pathway/WP4518_r104483/WP/Interaction/dc7ef> associatedWith: omim:260005 .
omim:260005 a obo:DOID_4 ; rdfs:label "Oxoprolinuria" .
<http://rdf.wikipathways.org/Pathway/WP4518_r104483/WP/Interaction/c6654> associatedWith: omim:230450 .
omim:230450 a obo:DOID_4 ; rdfs:label "Gamma-Glutamylcysteine Synthetase Deficiency" .
```

Workflow - 4

SPARQL queries with Blazegraph

- 3 files were loaded in Blazegraph

  - Turtle-RDF
  - WP-RDF of 6 WikiPathways
  - Datadiseases ttl file

Workflow - 4

SPARQL queries with Blazegraph

- Can we find the biomarker in the pathway that is related to the disease?
  - In case the biomarker is not in the related pathway, can we find the biomarker in another pathway on WikiPathways?

1. ***Query for co-occurrence of disease and biomarker in the same biological pathway***

Workflow - 4

SPARQL queries with Blazegraph

- Are the biomarkers directly connected to a disease?
  - Is the disease associated with an interaction that involves the biomarker?
  - Is the disease associated with an interaction directly linked to an interaction that involves the biomarker?

2. ***Query for disease is associated with an interaction that involves the biomarker***
3. ***Query for disease is assocaited with an interaction directly linked to an interaction that involves the biomarker***

Results

1. *Query for co-occurence of disease and biomarker in the same biological pathway*

```
SELECT DISTINCT ?wpid ?disease ?diseasename ?biomarker
WHERE
{
  ?metabolite a wp:Metabolite , wp:DataNode .
#metabolite is a metabolite and a DataNode
  ?metabolite rdfs:label ?biomarker .
#the predicate rdfs:label gives the name of the metabolite
  ?metabolite dcterms:isPartOf ?pathway .
#metabolite had to be part of a pathway
  ?pathway a wp:Pathway ; dcterms:identifier ?wpid .
#this pathway is a pathway and has an identifier called wpid
  ?metabolite dcterms:isPartOf ?interaction .
#the metaboltie is part of an interaction
  ?interaction a wp:Interaction .
#this interaction is an wikipathways interaction
  ?interaction associatedWith: ?disease .
#the interaction has to be association with a disease
  ?disease a obo:DOID_4 ; rdfs:label ?diseasename .
#disease is a disease and has a label which is the disease name

}

ORDER BY ?disease
```

Results

1. *Query for co-occurence of disease and biomarker in the same biological pathway*

- 34 of the 50 diseases present in the WikiPathways connected to a biomarker
- 5 biomarkers for Adenosine deaminase deficiency

wpid	disease	diseasename	biomarker
WP4224	omim/102700	Adenosine deaminase deficiency	Adenosine
WP4224	omim/102700	Adenosine deaminase deficiency	Inosine
WP4224	omim/102700	Adenosine deaminase deficiency	Adenine
WP4224	omim/102700	Adenosine deaminase deficiency	**2'-Deoxyadenosine**
WP4224	omim/102700	Adenosine deaminase deficiency	2'-deoxyinosine

```
wd:Q422457 a chebi:25367 ; rdfs:label "2'-deoxyadenosine" .
:assoc87 a association: ; refersTo: chapter41:, wd:Q190012, omim:102700 .
```

Results

1. Query for co-occurence of disease and biomarker in the same biological pathway

- WP4236 “Disorders of the Krebs cycle”
- SPARQL query showed only two diseases that were present in WP4236

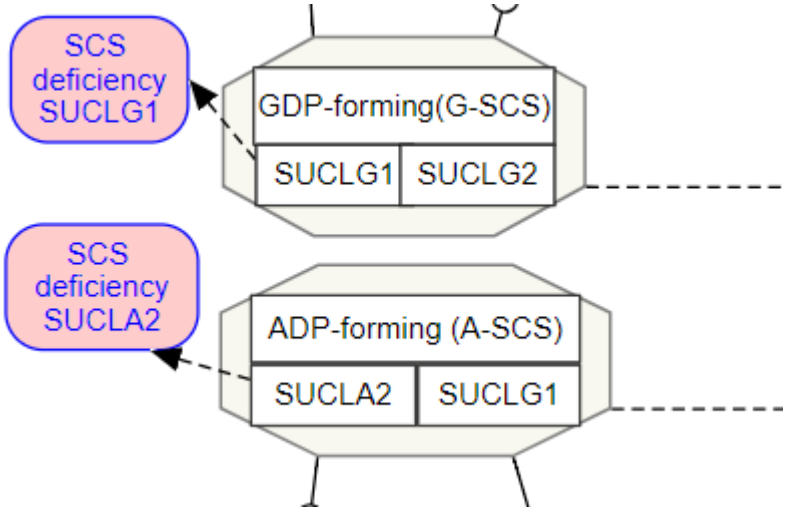
wpid	disease	diseasename	biomarker
WP4236	<http://identifiers.org/omim/136850>	Fumarase deficiency	(S)-malate
WP4236	<http://identifiers.org/omim/136850>	Fumarase deficiency	Fumarate
WP4236	<http://identifiers.org/omim/2030740>	2-oxoglutaric aciduria	Succinyl coenzyme A
WP4236	<http://identifiers.org/omim/2030740>	2-oxoglutaric aciduria	alpha-ketoglutarate

Results

1. *Query for co-occurence of disease and biomarker in the same biological pathway*

- WP4236 “Disorders of the Krebs cycle”
- SPARQL query showed only two diseases that were present in WP4236

```
# Disease labels chapter20
omim:203740 a obo:DOID_4 ; rdfs:label "2-oxoglutaric aciduria" .
omim:606812 a obo:DOID_4 ; rdfs:label "fumarase deficiency" .
omim:603921 a obo:DOID_4 ; rdfs:label "SCS deficiencies SUCLA2" .
omim:611224 a obo:DOID_4 ; rdfs:label "SCS deficiency SUCLG1" .
```



Results

2. *Query for disease is associated with an interaction that involves the biomarker*

```
SELECT DISTINCT ?wpid ?disease ?interaction ?wikidataid ?biomarker
WHERE
{
  ?interaction associatedWith: ?disease .
  ?disease a obo:DOID_4 ; rdfs:label ?diseasename .
  ?interaction dcterms:isPartOf ?pathway .
  ?pathway a wp:Pathway ; dcterms:identifier ?wpid .
  ?interaction a wp:DirectedInteraction .
#interaction has to be a directed interaction
  ?metabolite dcterms:isPartOf ?interaction .
  ?metabolite dcterms:isPartOf ?pathway .
  ?metabolite a wp:Metabolite .
  ?metabolite wp:bdbWikidata ?wikidataid .
#match the ChEBI identifiers to Wikidata identifiers
  ?association refersTo: ?wikidataid .
  ?wikidataid a chebi:25367 .
  ?wikidataid rdfs:label ?biomarker .
}
ORDER by ?wpid
```

Results

2. *Query for disease is associated with an interaction that involves the biomarker*

- Biomarker “hypoxanthine” associated with three diseases

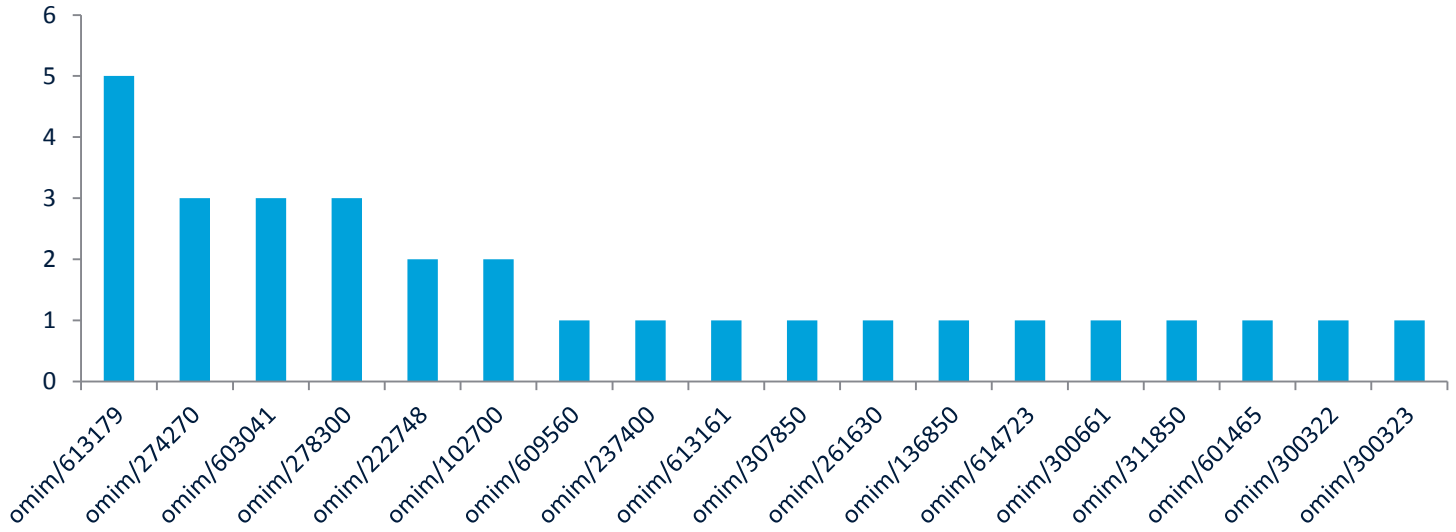
wpid	disease	interaction	wikidataid	biomarker
WP4224	omim/278300	id2eb2cadd	Q410305	hypoxanthine
WP4224	omim/613179	id6fb2fa37	Q410305	hypoxanthine
WP4224	omim/613179	id88da3202	Q410305	hypoxanthine
WP4224	omim/300322	idce0beeb9	Q410305	hypoxanthine
WP4224	omim/300323	idce0beeb9	Q410305	hypoxanthine

```
wd:Q410305 a chebi:25367 ; rdfs:label "hypoxanthine" .
:assoc98 a association: ; refersTo: chapter41:, wd:Q410305, omim:278300 .
:assoc99 a association: ; refersTo: chapter41:, wd:Q410305, omim:603592 .
:assoc100 a association: ; refersTo: chapter41:, wd:Q410305, omim:300322 .
:assoc101 a association: ; refersTo: chapter41:, wd:Q410305, omim:300661 .
```

Results

2. Query for disease is associated with an interaction that involves the biomarker

- Disease purine nucleoside phosphorylase (PNP) deficiency omim (613179) associated with 5 unique biomarkers



Results

3. *Query for disease is associated with an interaction directly linked to an interaction that involves the biomarker*

- Biomarker should be part of interaction2 which is linked to interaction1

```
SELECT DISTINCT ?disease (count(distinct ?biomarker) as ?biomarkercount)
WHERE
{
  ?interaction1 associatedWith: ?disease .
  ?disease a obo:DOID_4 ; rdfs:label ?diseasename .
  ?interaction1 dcterms:isPartOf ?pathway .
  ?pathway a wp:Pathway ; dcterms:identifier ?wpid .
  ?interaction1 a wp:DirectedInteraction .
  ?metabolite a wp:Metabolite .
  ?metabolite dcterms:isPartOf ?pathway .
  ?metabolite wp:bdbWikidata ?wikidataid .
  ?association refersTo: ?wikidataid .
  ?wikidataid a chebi:25367 .
  ?wikidataid rdfs:label ?biomarker .
  ?interaction2 a wp:Interaction .
  ?interaction2 a wp:DirectedInteraction .
  ?interaction2 dcterms:isPartOf ?pathway .
  ?interaction2 dcterms:isPartOf ?metabolite .

FILTER (?interaction != ?interaction2)
}

GROUP by ?disease ORDER by desc(?biomarkercount)
```

Discussion

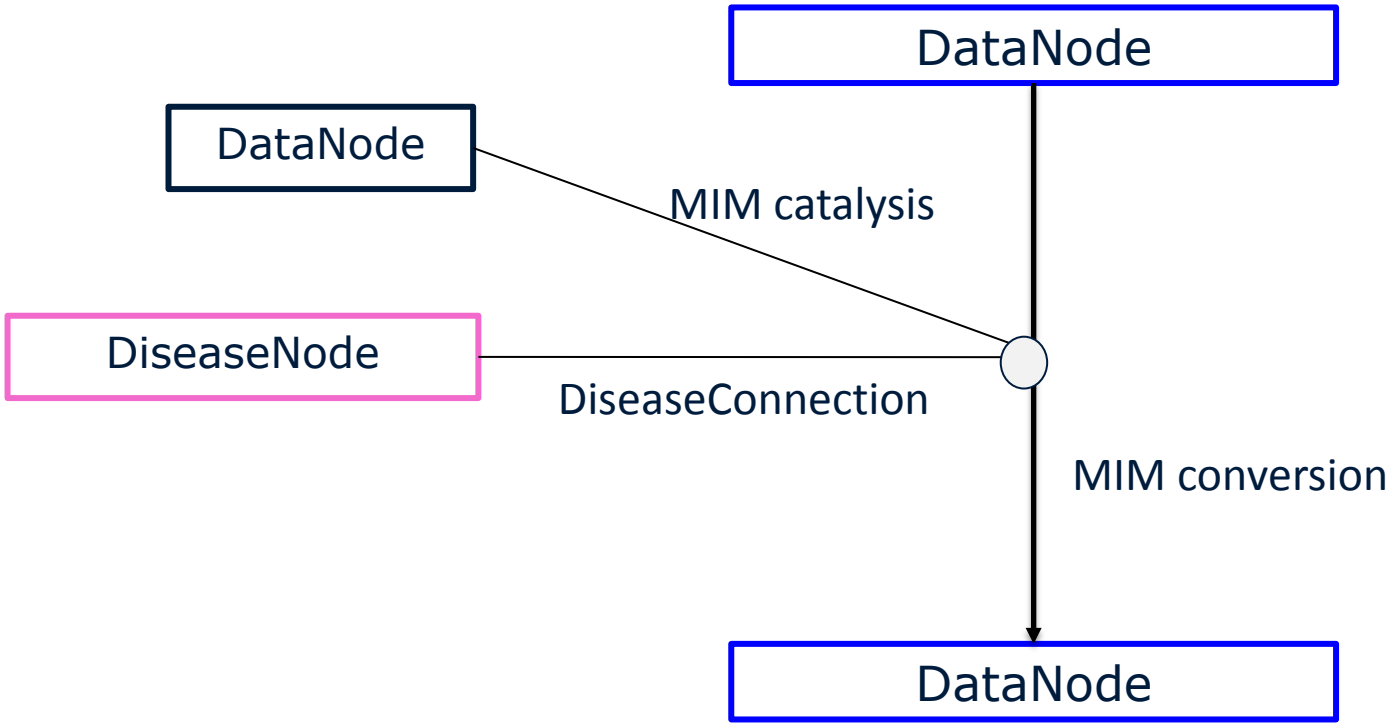
- RDF with biomarker-disease associations annotated with Wikidata identifiers, metabolites in WP with ChEBI identifiers
- Not all diseases in pathways are connected to the interaction in the right way
- Biomarkers present in pathway but not all biomarkers are linked to diseases

Conclusion

- Biological pathways can be used to check for **direct** disease-biomarker associations
- Turtle-RDF should use different predicates which match to WP-RDF

Conclusion

- New design for PathVisio



Questions?