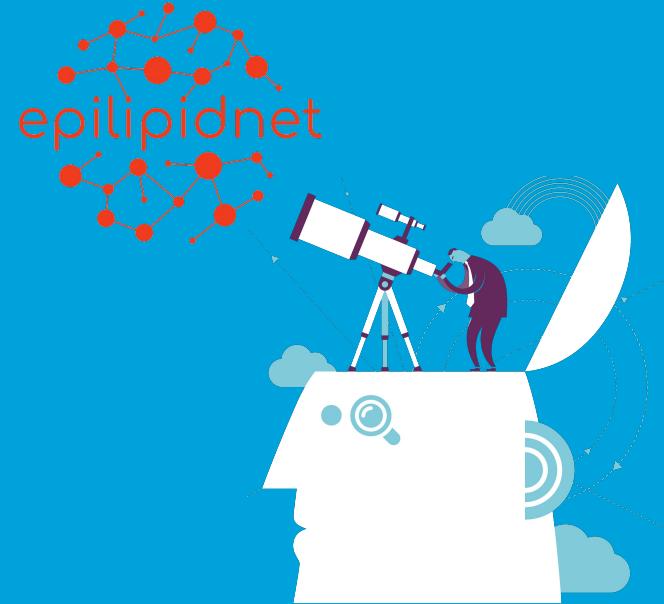


# Metabolomics data analysis

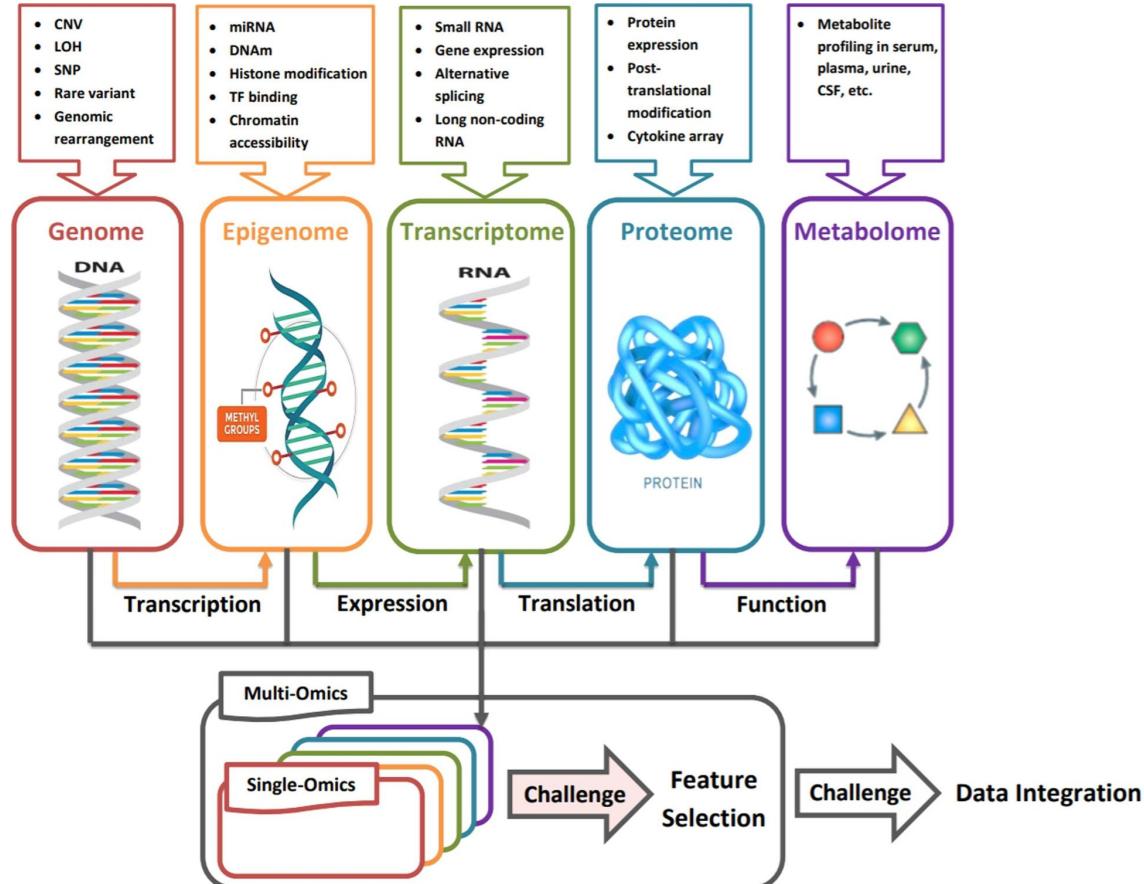
Denise Slenter

ORCID: 0000-0001-8449-1318

Tuesday July 2nd, 2024

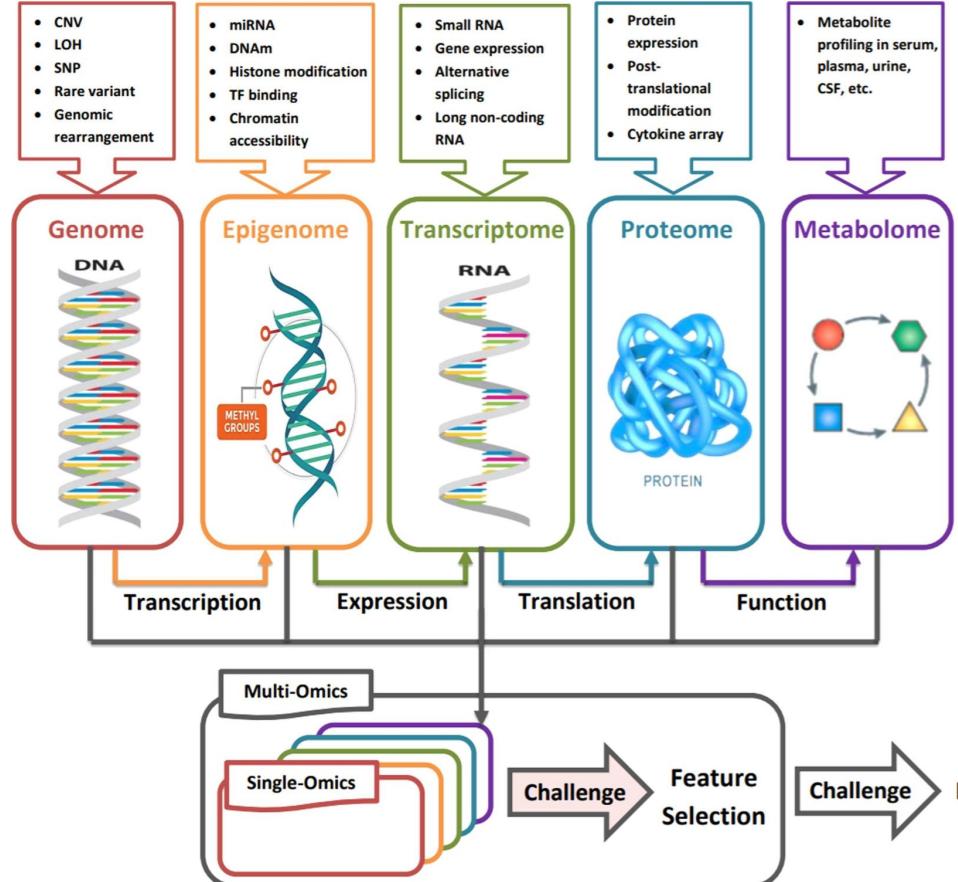


# Different data types and analysis techniques



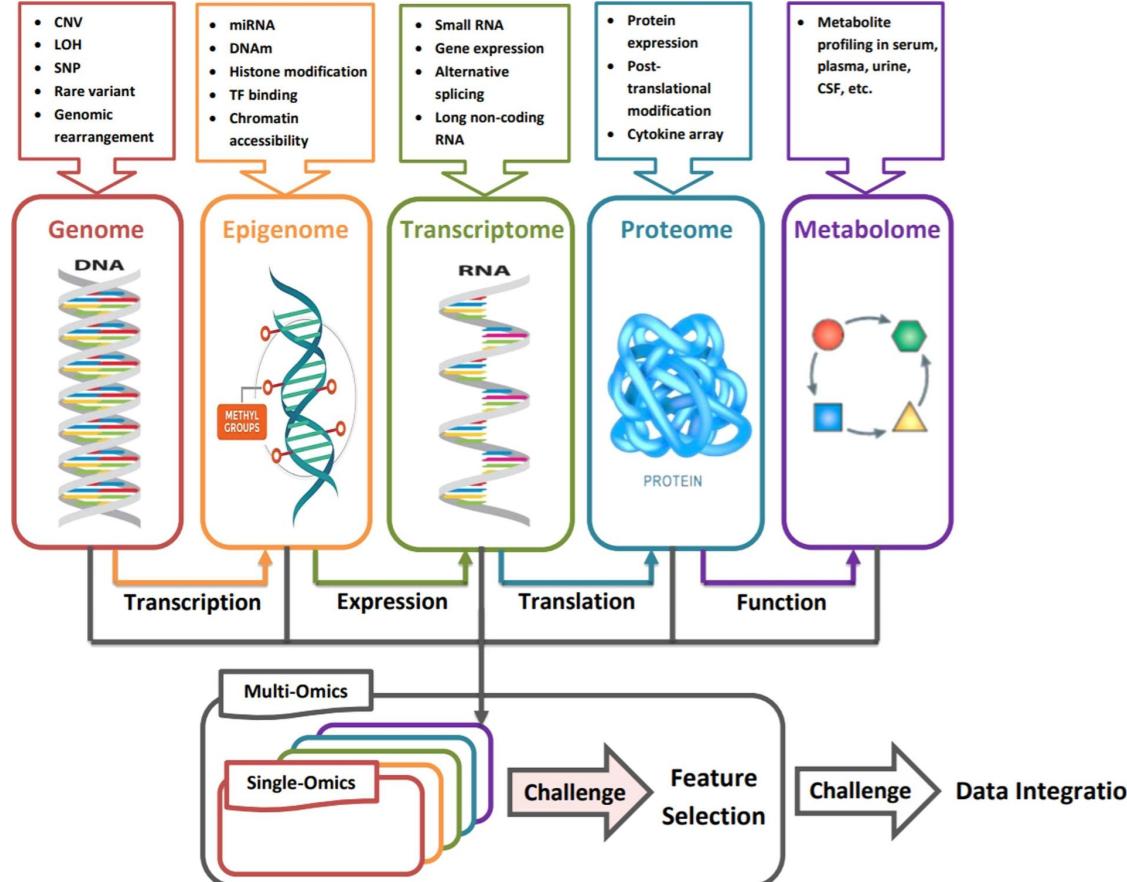
Adapted from: Momeni, Zahra, et al. "A survey on single and multi omics data mining methods in cancer data classification." Journal of Biomedical Informatics 107 (2020): 103466. DOI: [10.1016/j.jbi.2020.103466](https://doi.org/10.1016/j.jbi.2020.103466)

# Different data types and analysis techniques



What data is missing in this overview?

# Different data types and analysis techniques

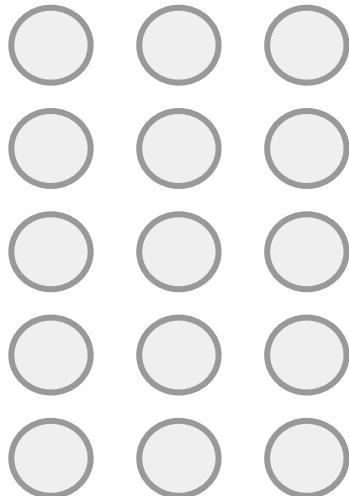


## Missing:

- Phenotype
- Imaging data
- Fluxomics
- ...

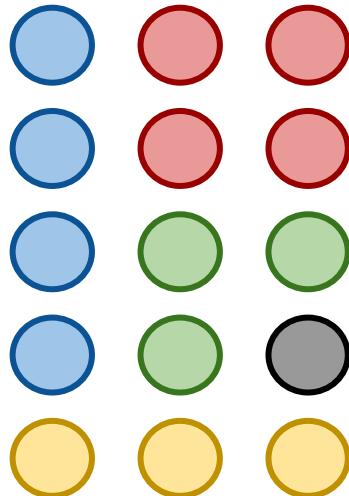
Adapted from: Momeni, Zahra, et al. "A survey on single and multi omics data mining methods in cancer data classification." Journal of Biomedical Informatics 107 (2020): 103466. DOI: [10.1016/j.jbi.2020.103466](https://doi.org/10.1016/j.jbi.2020.103466)

# Pathway analysis



**Quantitative measurements**  
Isolated data points

# Pathway analysis



**Comparative statistics**

Isolated lists

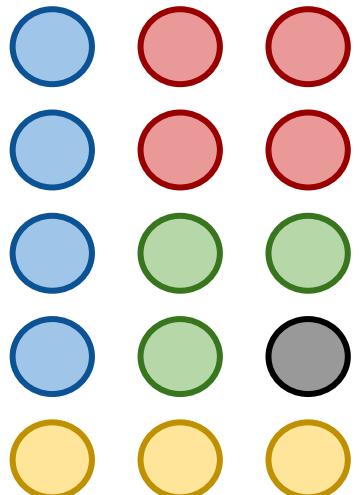
**Clustering**

Isolated groups

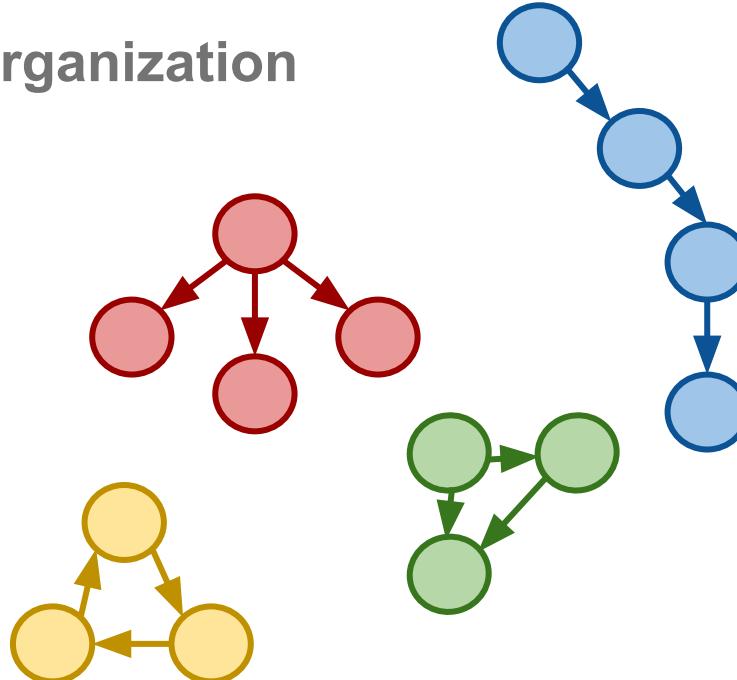
**Gene sets**

Functional groups

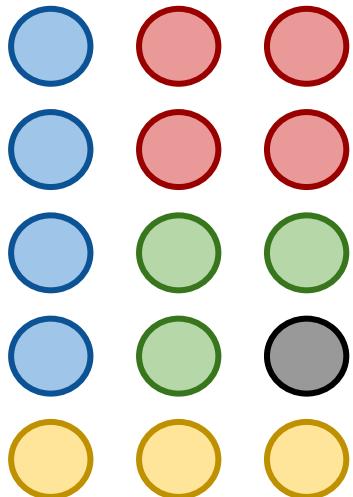
# Pathway analysis



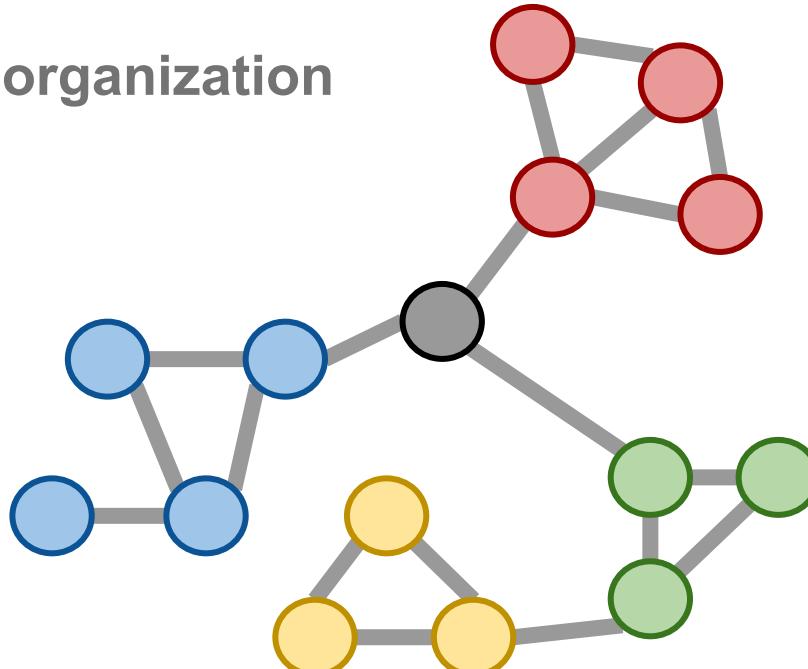
Functional organization  
Pathways

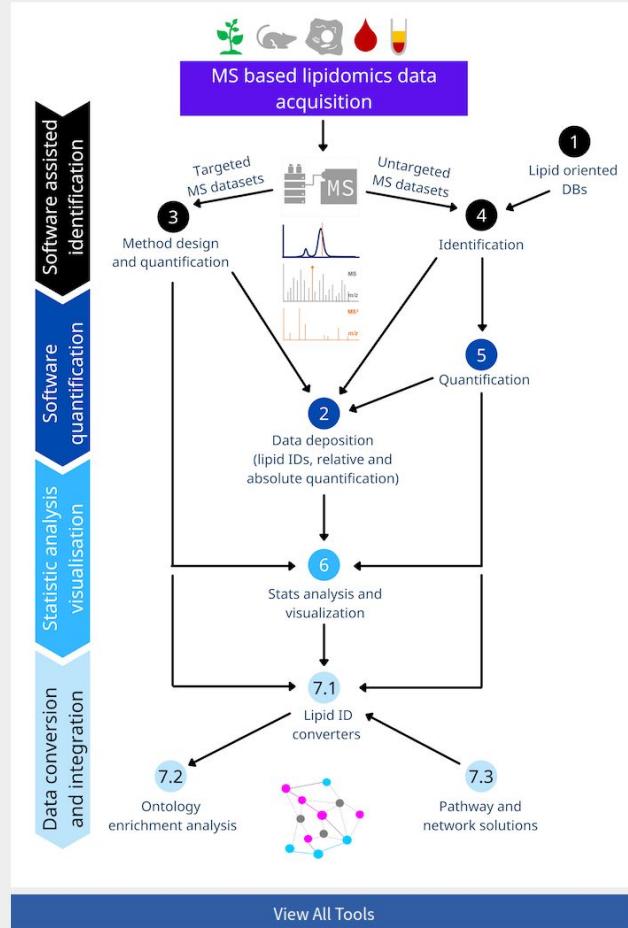


# Pathway analysis



Systems organization  
Networks

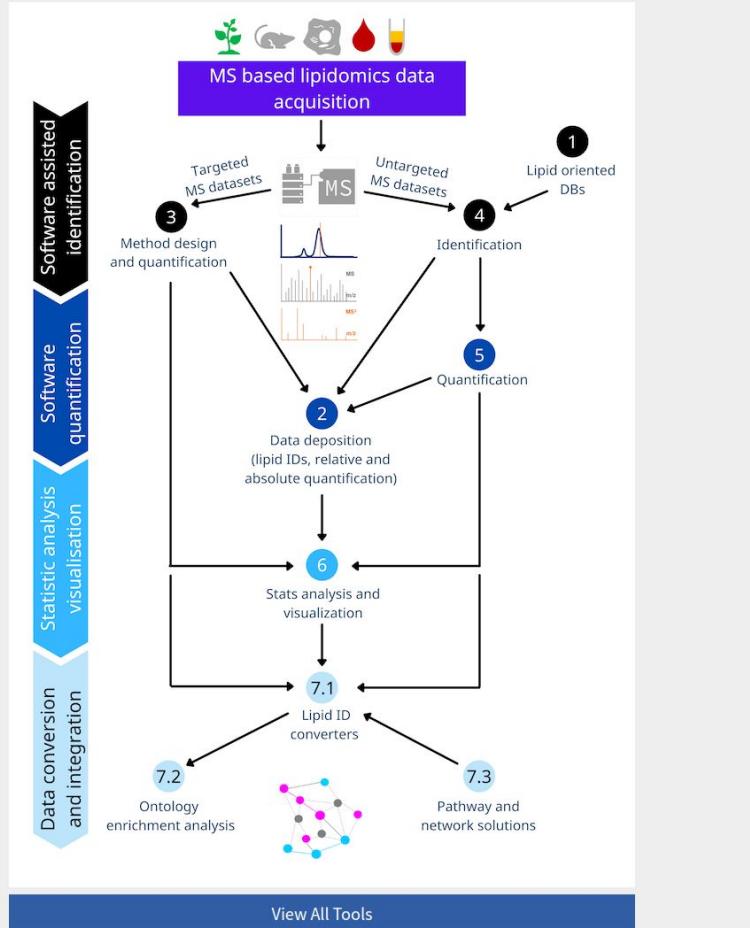




# Overview for Lipidomics analysis tools exists!

## For metabolomics not so much (unfortunately!)

Adapted from: Ni, Zhixu, et al. "Guiding the choice of informatics software and tools for lipidomics research applications." *Nature methods* 20.2 (2023): 193-204. DOI: [10.1038/s41592-022-01710-0](https://doi.org/10.1038/s41592-022-01710-0)



# Considerations when comparing tools:

- License & Source Code
- Graphical User Interface (GUI)
- Command Line Interface (CLI)
- Desktop client / web interface
- Input & output formats
- Operating Systems (Windows, Mac, Linux)
- Programming Language (R, Python, Java, Matlab, ...)
- Coverage and IDs used

Adapted from: Ni, Zhixu, et al. "Guiding the choice of informatics software and tools for lipidomics research applications." *Nature methods* 20.2 (2023): 193-204. DOI: [10.1038/s41592-022-01710-0](https://doi.org/10.1038/s41592-022-01710-0)

# Coverage of pathway data (according to RaMP, merging information from 4 pathway databases)

**Table 2.** Number of analytes and pathways (A) and chemical properties (B) available through RaMP-DB 2.0

A	Total <sup>a</sup>	HMDB v5.0	KEGG (from HMDB 5.0)	Reactome v81	WikiPathways v20220710
# Distinct metabolites	256 086 (+142 361)	216 683	5898	2355	3695
# Distinct genes/enzymes	15 827 (+410)	7111	-	11 227	13 393
# Distinct pathways	53 831 (+2035)	49 613	363	2583	1272
#Metabolite-pathway mappings	412 775 (+343 120)	367 609	1714	30 804	12 648
# Gene-pathway mappings	401 303 (-695 287)	208 211	8479	125 171	59 442

B

Total distinct compounds <sup>b</sup>	HMDB v5.0	ChEBI release 212	LIPID MAPS release July 13, 2022
Chemical properties <sup>c</sup>	256 592	217 776	13 066

- a The number in parentheses represents the difference in numbers compared to the previous RaMP version (1.1.0).
- b Distinct InChIKeys.
- c Chemical properties are only captured for compounds referenced within RaMP.

# Coverage of pathway data (according to RaMP, merging information from 4 pathway databases)

**Table 2.** Number of analytes and pathways (A) and chemical properties (B) available through RaMP-DB 2.0

	Total <sup>a</sup>	HMDB v5.0	KEGG (from HMDB 5.0)	Reactome v81	WikiPathways v20220710
# Distinct metabolites	256 086 (+142 361)	216 683	5898	2355	3695
# Distinct genes/enzymes	15 827 (+410)	7111	-	11 227	13 393
# Distinct pathways	53 831 (+2035)	49 613	363	2583	1272
# Metabolite-pathway mappings	412 775 (+343 120)	367 609	1714	30 804	12 648
# Gene-pathway mappings	401 303 (-695 287)	208 211	8479	125 171	59 442

B

	Total distinct compounds <sup>b</sup>	HMDB v5.0	ChEBI release 212	LIPID MAPS release July 13, 2022
Chemical properties <sup>c</sup>	256 592	217 776	13 066	44 981

a The number in parentheses represents the difference in numbers compared to the previous RaMP version (1.1.0).

b Distinct InChIKeys.

c Chemical properties are only captured for compounds referenced within RaMP.

# Coverage of pathway data (according to RaMP, merging information from 4 pathway databases)

**Table 2.** Number of analytes and pathways (A) and chemical properties (B) available through RaMP-DB 2.0

A	Total <sup>a</sup>	HMDB v5.0	KEGG (from HMDB 5.0)	Reactome v81	WikiPathways v20220710
# Distinct metabolites	256 086 (+142 361)	216 683	5898	2355	3695
# Distinct genes/enzymes	15 827 (+410)	7111	-	11 227	13 393
# Distinct pathways	53 831 (+2035)	49 613	363	2583	1272
# Metabolite-pathway mappings	412 775 (+343 120)	367 609	1714	30 804	12 648
# Gene-pathway mappings	401 303 (-695 287)	208 211	8479	125 171	59 442
B					
Total distinct compounds <sup>b</sup>	256 592	HMDB v5.0	ChEBI release 212	LIPID MAPS release July 13, 2022	
Chemical properties <sup>c</sup>	217 776	13 066	44 981		

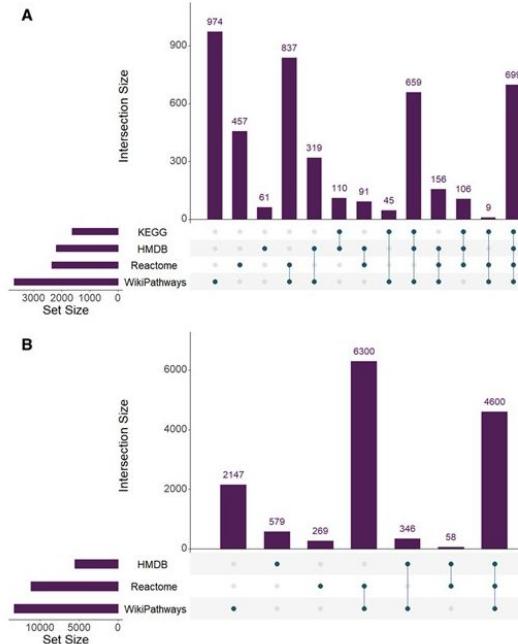
a The number in parentheses represents the difference in numbers compared to the previous RaMP version (1.1.0).

b Distinct InChIKeys.

c Chemical properties are only captured for compounds referenced within RaMP.

Adapted from: Braisted, John, et al. "RaMP-DB 2.0: a renovated knowledgebase for deriving biological and chemical insight from metabolites, proteins, and genes." Bioinformatics 39.1 (2023). DOI: [10.1093/bioinformatics/btac726](https://doi.org/10.1093/bioinformatics/btac726)

**Fig 3.**



[Open in new tab](#)

[Download slide](#)

Overlap in content among source databases. Only analytes mapping to pathways are considered, as HMDB contains a large number of metabolites associated only with ontologies, which are not relevant to Reactome and WikiPathways as pathway-centric databases. (A) Overlap in metabolites associated with at least one pathway between source databases in RaMP. (B) Overlap of genes associated with at least one pathway. The filled circle(s) underneath each bar in the plots demonstrate the source databases that the analyte counts are drawn from

# Some existing tools in metabolomics analysis

# Based on pathway databases in RaMP

**Table 2.** Number of analytes and pathways (A) and chemical properties (B) available through RaMP-DB 2.0.

A	Total <sup>a</sup>	HMDB v5.0	KEGG (from HMDB 5.0)	Reactome v81	WikiPathways v20220710
# Distinct metabolites	256 086 (+142 361)	216 683	5898	2355	3695
# Distinct genes/enzymes	15 827 (+410)	7111	-	11 227	13 393
# Distinct pathways	53 831 (+2035)	49 613	363	2583	1272
# Metabolite-pathway mappings	412 775 (+343 120)	367 609	1714	30 804	12 648
# Gene-pathway mappings	401 303 (-695 287)	208 211	8479	125 171	59 442

B
Total distinct compounds <sup>b</sup>
HMDB v5.0
ChEBI release 212
LIPID MAPS release July 13, 2022

Chemical properties <sup>c</sup>	256 592	217 776	13 066	44 981
----------------------------------	---------	---------	--------	--------

a The number in parentheses represents the difference in numbers compared to the previous RaMP version (1.1.0).

b Distinct InChIKeys.

c Chemical properties are only captured for compounds referenced within RaMP.

DATABASE AND TOOLS	ANALYSIS OPTIONS	ID MAPPING																				
	<p><b>Input Data Type</b></p> <p>LC-MS Spectra (mzXML, mzML, or mzData)</p> <p>MS Peaks (peak list or intensity table)</p> <p>Generic Format (CSV or TSV table files)</p> <p>Annotated Features (metabolite list or table)</p> <p>Link to Genomics &amp; Phenotypes (metabolite list)</p>	<p><b>18 modules, Free to use</b></p> <table border="1"> <tr> <td></td> <td></td> <td>Spectra Processing (LC-MS with RQ)</td> <td></td> </tr> <tr> <td></td> <td>Peak Annotation (MS/MS DEconv)</td> <td>Functional Analysis (LC-MS)</td> <td>Functional Meta-analysis (LC-MS)</td> </tr> <tr> <td>Statistical Analysis (gene factors)</td> <td>Statistical Analysis (metabolite table)</td> <td>Biomarker Analysis</td> <td>Statistical Meta-analysis</td> </tr> <tr> <td></td> <td>Enrichment Analysis</td> <td>Pathway Analysis</td> <td>Network Analysis</td> </tr> <tr> <td></td> <td></td> <td>Causal Analysis (Metabolite concentration)</td> <td></td> </tr> </table>			Spectra Processing (LC-MS with RQ)			Peak Annotation (MS/MS DEconv)	Functional Analysis (LC-MS)	Functional Meta-analysis (LC-MS)	Statistical Analysis (gene factors)	Statistical Analysis (metabolite table)	Biomarker Analysis	Statistical Meta-analysis		Enrichment Analysis	Pathway Analysis	Network Analysis			Causal Analysis (Metabolite concentration)	
		Spectra Processing (LC-MS with RQ)																				
	Peak Annotation (MS/MS DEconv)	Functional Analysis (LC-MS)	Functional Meta-analysis (LC-MS)																			
Statistical Analysis (gene factors)	Statistical Analysis (metabolite table)	Biomarker Analysis	Statistical Meta-analysis																			
	Enrichment Analysis	Pathway Analysis	Network Analysis																			
		Causal Analysis (Metabolite concentration)																				

# Based on pathway databases in RaMP

**Table 2.** Number of analytes and pathways (A) and chemical properties (B) available through RaMP-DB 2.0.

**A**

	Total <sup>a</sup>	HMDB v5.0	KEGG (from HMDB 5.0)	Reactome v81	WikiPathways v20220710
# Distinct metabolites	256 086 (+142 361)	216 683	5898	2355	3695
# Distinct genes/enzymes	15 827 (+410)	7111	-	11 227	13 393
# Distinct pathways	53 831 (+2035)	49 613	363	2583	1272
# Metabolite-pathway mappings	412 775 (+343 120)	367 609	1714	30 804	12 648
# Gene-pathway mappings	401 303 (-695 287)	208 211	8479	125 171	59 442

**B**

Total distinct compounds <sup>b</sup>	HMDB v5.0	ChEBI release 212	LIPID MAPS release July 13, 2022
---------------------------------------	-----------	-------------------	----------------------------------

Chemical properties<sup>c</sup> 256 592 217 776 13 066 44 981

a The number in parentheses represents the difference in numbers compared to the previous RaMP version (1.1.0).

b Distinct InChIKeys.

c Chemical properties are only captured for compounds referenced within RaMP.

DATABASE AND TOOLS	ANALYSIS OPTIONS	ID MAPPING															
<p>The Human Metabolome Database (hmdb), PathBank, DRUGBANK, and Small Molecule Pathway Database logos.</p>	<p><b>Input Data Type</b></p> <ul style="list-style-type: none"> <li>LC-MS Spectra (mzXML, mzML, or mzData)</li> <li>MS Peaks (peak list or intensity table)</li> <li>Genomic Format (csv or .txt table files)</li> <li>Annotated Features (metabolite list or table)</li> <li>Link to Genomics &amp; Phenotypes (metabolite list)</li> </ul> <p><b>18 modules, Free to use</b></p> <table border="1"> <tr> <td></td> <td>Peak Annotation (MS/MS Deconv)</td> <td>Spectra Processing (LC-MS with MS2)</td> <td>Functional Analysis (LC-MS)</td> <td>Functional Meta-analysis (LC-MS)</td> </tr> <tr> <td>Statistical Analysis (gene factors)</td> <td>Statistical Analysis (metabolite table)</td> <td>Biomarker Analysis</td> <td>Statistical Meta-analysis</td> <td>Dose Response Analysis</td> </tr> <tr> <td></td> <td>Enrichment Analysis</td> <td>Pathway Analysis</td> <td>Causal Analysis (Metabolite concentration)</td> <td>Network Analysis</td> </tr> </table>		Peak Annotation (MS/MS Deconv)	Spectra Processing (LC-MS with MS2)	Functional Analysis (LC-MS)	Functional Meta-analysis (LC-MS)	Statistical Analysis (gene factors)	Statistical Analysis (metabolite table)	Biomarker Analysis	Statistical Meta-analysis	Dose Response Analysis		Enrichment Analysis	Pathway Analysis	Causal Analysis (Metabolite concentration)	Network Analysis	<p>Metabolite ID conversion</p> <p><a href="https://www.metaboanalyst.ca/MetaboAnalyst/upload/ConvertView.xhtml">https://www.metaboanalyst.ca/MetaboAnalyst/upload/ConvertView.xhtml</a></p>
	Peak Annotation (MS/MS Deconv)	Spectra Processing (LC-MS with MS2)	Functional Analysis (LC-MS)	Functional Meta-analysis (LC-MS)													
Statistical Analysis (gene factors)	Statistical Analysis (metabolite table)	Biomarker Analysis	Statistical Meta-analysis	Dose Response Analysis													
	Enrichment Analysis	Pathway Analysis	Causal Analysis (Metabolite concentration)	Network Analysis													
<p>Kyoto Encyclopedia of Genes and Genomes (KEGG) logo.</p> <p>Reaction scheme showing L-threonine conversion to N-acetyl-L-threonine.</p> <p>Legend: Reaction center (blue circle), Difference arrow (yellow circle), Matched arrow (orange circle), HO - CH<sub>2</sub> (green circle), C(=O) - CH<sub>2</sub> (red circle).</p>	<p><b>Licence required; Academic usage Of website for free</b></p>	<p>KEGG Mapper</p> <p><a href="https://www.genome.jp/kegg/mapper/">https://www.genome.jp/kegg/mapper/</a></p>															

# Based on pathway databases in RaMP

**Table 2.** Number of analytes and pathways (A) and chemical properties (B) available through RaMP-DB 2.0.

## A

	Total <sup>a</sup>	HMDB v5.0	KEGG (from HMDB 5.0)	Reactome v81	WikiPathways v20220710
# Distinct metabolites	256 086 (+142 361)	216 683	5898	2355	3695
# Distinct genes/enzymes	15 827 (+410)	7111	-	11 227	13 393
# Distinct pathways	53 831 (+2035)	49 613	363	2583	1272
# Metabolite-pathway mappings	412 775 (+343 120)	367 609	1714	30 804	12 648
# Gene-pathway mappings	401 303 (-695 287)	208 211	8479	125 171	59 442

## B

Total distinct compounds <sup>b</sup>	HMDB v5.0	ChEBI release 212	LIPID MAPS release July 13, 2022
Chemical properties <sup>c</sup>	256 592	217 776	13 066

a The number in parentheses represents the difference in numbers compared to the previous RaMP version (1.1.0).

b Distinct InChIKeys.

c Chemical properties are only captured for compounds referenced within RaMP.

DATABASE AND TOOLS	ANALYSIS OPTIONS	ID MAPPING															
	<p><b>Input Data Type</b></p> <p>LC-MS Spectra (mzXML, mzML, or mzData)</p> <p>MS Peaks (peak list or intensity table)</p> <p>Genomic Format (csv or .txt table files)</p> <p>Annotated Features (metabolite list or table)</p> <p>Link to Genomics &amp; Phenotypes (metabolite list)</p> 	<p><b>18 modules, Free to use</b></p> <table border="1"> <tr> <td></td> <td>Peak Annotation (MS/MS Deconv)</td> <td>Spectra Processing (LC-MS with MS2)</td> <td>Functional Analysis (LC-MS)</td> <td>Metabolite ID conversion</td> </tr> <tr> <td>Statistical Analysis (gene factors)</td> <td>Statistical Analysis (metabolite table)</td> <td>Biomarker Analysis</td> <td>Statistical Meta-analysis</td> <td>Dose Response Analysis</td> </tr> <tr> <td></td> <td>Enrichment Analysis</td> <td>Pathway Analysis</td> <td>Causal Analysis (Metabolite concentration)</td> <td>Network Analysis</td> </tr> </table>		Peak Annotation (MS/MS Deconv)	Spectra Processing (LC-MS with MS2)	Functional Analysis (LC-MS)	Metabolite ID conversion	Statistical Analysis (gene factors)	Statistical Analysis (metabolite table)	Biomarker Analysis	Statistical Meta-analysis	Dose Response Analysis		Enrichment Analysis	Pathway Analysis	Causal Analysis (Metabolite concentration)	Network Analysis
	Peak Annotation (MS/MS Deconv)	Spectra Processing (LC-MS with MS2)	Functional Analysis (LC-MS)	Metabolite ID conversion													
Statistical Analysis (gene factors)	Statistical Analysis (metabolite table)	Biomarker Analysis	Statistical Meta-analysis	Dose Response Analysis													
	Enrichment Analysis	Pathway Analysis	Causal Analysis (Metabolite concentration)	Network Analysis													
	<p>L-threonine OXIDATION</p> <p><chem>CC(=O)N[C@@H](C)C</chem> → <chem>CC(=O)N([O-])[C@@H](C)C</chem></p> <p>Reaction center: <chem>Ni+ Ni2+</chem> Difference atom: <chem>H+ - Cl-</chem> Matched atom: <chem>Cx - Cx</chem></p>	<p><b>Licence required; Academic usage Of website for free</b></p>															
		<p><b>KEGG Mapper</b></p> <p><a href="https://www.genome.jp/kegg/mapper/">https://www.genome.jp/kegg/mapper/</a></p> <p><b>Integrated in analysis tool, downloadable files available</b></p> <p><a href="https://reactome.org/download-data">https://reactome.org/download-data</a></p>															

# Based on pathway databases in RaMP

**Table 2.** Number of analytes and pathways (A) and chemical properties (B) available through RaMP-DB 2.0.

## A

	Total <sup>a</sup>	HMDB v5.0	KEGG (from HMDB 5.0)	Reactome v81	WikiPathways v20220710
# Distinct metabolites	256 086 (+142 361)	216 683	5898	2355	3695
# Distinct genes/enzymes	15 827 (+410)	7111	-	11 227	13 393
# Distinct pathways	53 831 (+2035)	49 613	363	2583	1272
# Metabolite-pathway mappings	412 775 (+343 120)	367 609	1714	30 804	12 648
# Gene-pathway mappings	401 303 (-695 287)	208 211	8479	125 171	59 442

## B

	Total distinct compounds <sup>b</sup>	HMDB v5.0	ChEBI release 212	LIPID MAPS release July 13, 2022
Chemical properties <sup>c</sup>	256 592	217 776	13 066	44 981

a The number in parentheses represents the difference in numbers compared to the previous RaMP version (1.1.0).

b Distinct InChIKeys.

c Chemical properties are only captured for compounds referenced within RaMP.

DATABASE AND TOOLS	ANALYSIS OPTIONS	ID MAPPING
	<p><b>Input Data Type</b></p> <ul style="list-style-type: none"> <li>LC-MS Spectra (mzXML, mzML, or mzData)</li> <li>MS Peaks (peak list or intensity table)</li> <li>Generic Format (csv or .txt table files)</li> <li>Annotated Features (metabolite list or table)</li> <li>Link to Genomics &amp; Phenotypes (metabolite list)</li> </ul> <p><b>18 modules, Free to use</b></p>	Metabolite ID conversion <a href="https://www.metaboanalyst.ca/MetaboAnalyst/upload/ConvertView.xhtml">https://www.metaboanalyst.ca/MetaboAnalyst/upload/ConvertView.xhtml</a>
	<p>L- glutamic acid → N-acetyl-L-glutamate</p> <p>Reaction center: NH<sub>2</sub>—NH<sub>2</sub> Difference atom: HO—CH<sub>2</sub> Matched atom: C=O—C=O</p> <p><b>Licence required; Academic usage Of website for free</b></p>	KEGG Mapper <a href="https://www.genome.jp/kegg/mapper/">https://www.genome.jp/kegg/mapper/</a>
	<p><b>Analysis tools</b></p> <ul style="list-style-type: none"> <li>Analyze gene list</li> <li>GO</li> <li>Protein Interaction</li> <li>Species Comparison</li> <li>Transcriptome</li> </ul>	Integrated in analysis tool, downloadable files available <a href="https://reactome.org/download-data">https://reactome.org/download-data</a>
		<a href="https://www.bridgedb.org/">https://www.bridgedb.org/</a>

# Four main types of pathway analysis

Over-Representation Analysis (ORA)

Functional Class Scoring (FCS)

Pathway Topology Analysis (TPA)

Network Enrichment Analysis (NEA)

# Four main types of pathway analysis

## Over-Representation Analysis (ORA)

Compares the overlap between metabolites of interest, metabolites present in a pathway, and the total number of metabolites measured and identified in a sample. This method does not include the arrangement of the elements in a PWM (topology), nor the ranking of the metabolites of interest.

## Functional Class Scoring (FCS)

## Pathway Topology Analysis (TPA)

## Network Enrichment Analysis (NEA)

# Four main types of pathway analysis

## Over-Representation Analysis (ORA)

Compares the overlap between metabolites of interest, metabolites present in a pathway, and the total number of metabolites measured and identified in a sample. This method does not include the arrangement of the elements in a PWM (topology), nor the ranking of the metabolites of interest.

## Functional Class Scoring (FCS)

Ranks the metabolites according to a statistical variable (e.g. p-value) and then compares the overlap between metabolites of interest to the ones present in a PWM. Topology is not considered in this approach.

## Pathway Topology Analysis (TPA)

## Network Enrichment Analysis (NEA)

# Four main types of pathway analysis

## Over-Representation Analysis (ORA)

Compares the overlap between metabolites of interest, metabolites present in a pathway, and the total number of metabolites measured and identified in a sample. This method does not include the arrangement of the elements in a PWM (topology), nor the ranking of the metabolites of interest.

## Functional Class Scoring (FCS)

Ranks the metabolites according to a statistical variable (e.g. p-value) and then compares the overlap between metabolites of interest to the ones present in a PWM. Topology is not considered in this approach.

## Pathway Topology Analysis (TPA)

The connections between individual metabolites are considered to estimate how a change in one particular metabolite might alter the complete pathway. This type of analysis depends on the relationships within a pathway

## Network Enrichment Analysis (NEA)

# Four main types of pathway analysis

## Over-Representation Analysis (ORA)

Compares the overlap between metabolites of interest, metabolites present in a pathway, and the total number of metabolites measured and identified in a sample. This method does not include the arrangement of the elements in a PWM (topology), nor the ranking of the metabolites of interest.

## Functional Class Scoring (FCS)

Ranks the metabolites according to a statistical variable (e.g. p-value) and then compares the overlap between metabolites of interest to the ones present in a PWM. Topology is not considered in this approach.

## Pathway Topology Analysis (TPA)

The connections between individual metabolites are considered to estimate how a change in one particular metabolite might alter the complete pathway. This type of analysis depends on the relationships within a pathway

## Network Enrichment Analysis (NEA)

This method surpasses the boundaries of a PWM, by comparing all relationships present in a chemical reaction network for overlap between metabolites of interest and metabolites present in the network.

# Four main types of pathway analysis

## Over-Representation Analysis (ORA)

Compares the overlap between metabolites of interest, metabolites present in a pathway, and the total number of metabolites measured and identified in a sample. This method does not include the arrangement of the elements in a PWM (topology), nor the ranking of the metabolites of interest.

## Functional Class Scoring (FCS)

Ranks the metabolites according to a statistical variable (e.g. p-value) and then compares the overlap between metabolites of interest to the ones present in a PWM. Topology is not considered in this approach.

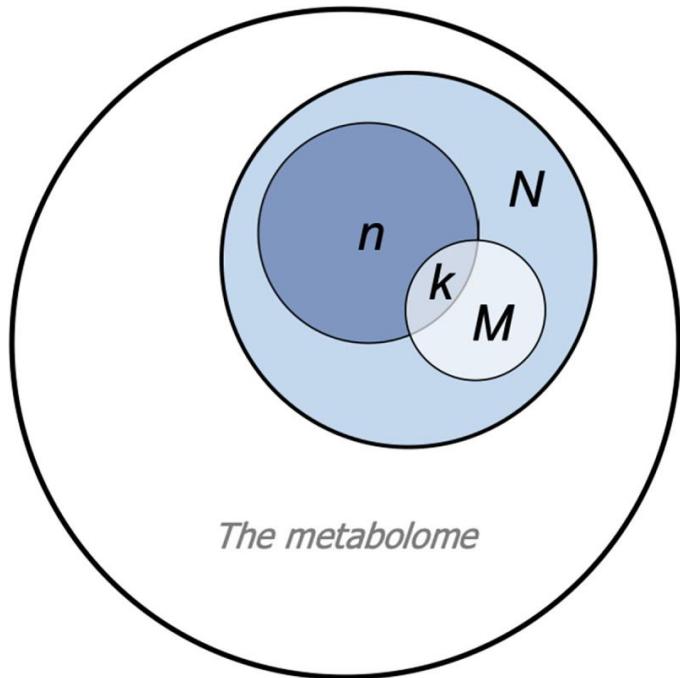
## Pathway Topology Analysis (TPA)

The connections between individual metabolites are considered to estimate how a change in one particular metabolite might alter the complete pathway. This type of analysis depends on the relationships within a pathway

## Network Enrichment Analysis (NEA)

This method surpasses the boundaries of a PWM, by comparing all relationships present in a chemical reaction network for overlap between metabolites of interest and metabolites present in the network.

# Why ORA?

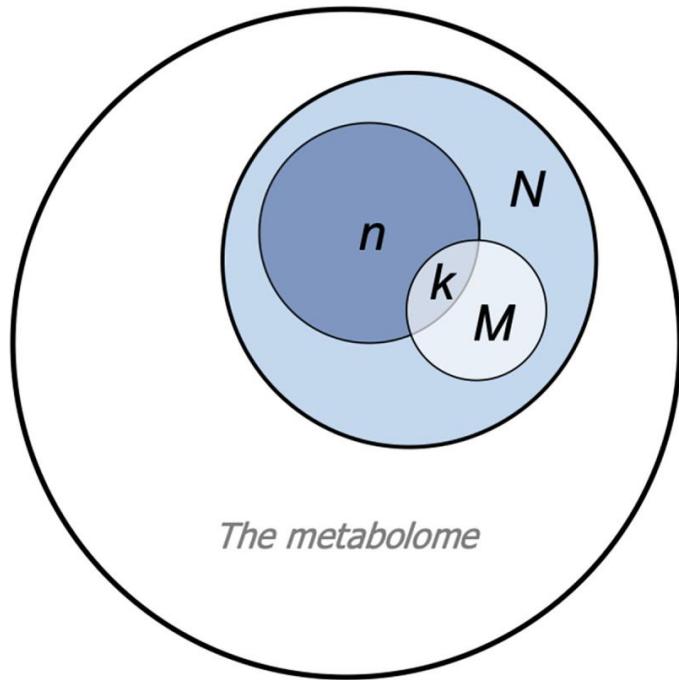


Over Representation Analysis (ORA).

Venn diagram representing ORA parameters corresponding to Eq 1. N represents compounds forming the background set, which covers part of the full metabolome. M represents compounds in the pathway of interest. n represents compounds of interest (i.e., differentially abundant metabolites), and k represents the overlap between the list of compounds of interest and compounds in the pathway.



# Why ORA?



Commonly performed for transcriptomics data

Not so easy for metabolomics!

# Why ORA through SPARQL?

The screenshot shows a SPARQL endpoint interface for WikiPathways. At the top, there's a logo for "WIKIPATHWAYS Pathways for the People" and a "SPARQL Endpoint" button with the URL "https://sparql.wikipathways.org/sparql/".

**SPARQL Query:**

```
1 SELECT count(DISTINCT ?lipidID) as ?IndividualLipidsPerClass
2 WHERE {
3   ?metabolite a wp:Metabolite ;
4   dcterms:identifier ?id ;
5   dcterms:isPartOf ?pathwayRes ;
6   wp:bdbLipidMaps ?lipidID . #Metabolite DataNodes need to have a LIPID MAPS ID, for this query to count correctly (some lipid
7   ?pathwayRes a wp:Pathway ;
8   wp:organismName "Homo sapiens"; #Filter for a species (omit when querying all pathways available for all species)
9   dcterms:identifier ?wpid ;
10  dcterms:title .
11 } FILTER regex(str(?lipidID), "FA" ). #Filter for a LIPID MAPS ID subclass: 'FA' Fatty Acids ; 'GL' Glycerolipid ; 'GP' Glycerop
12 }
```

**SPARQL Examples:**

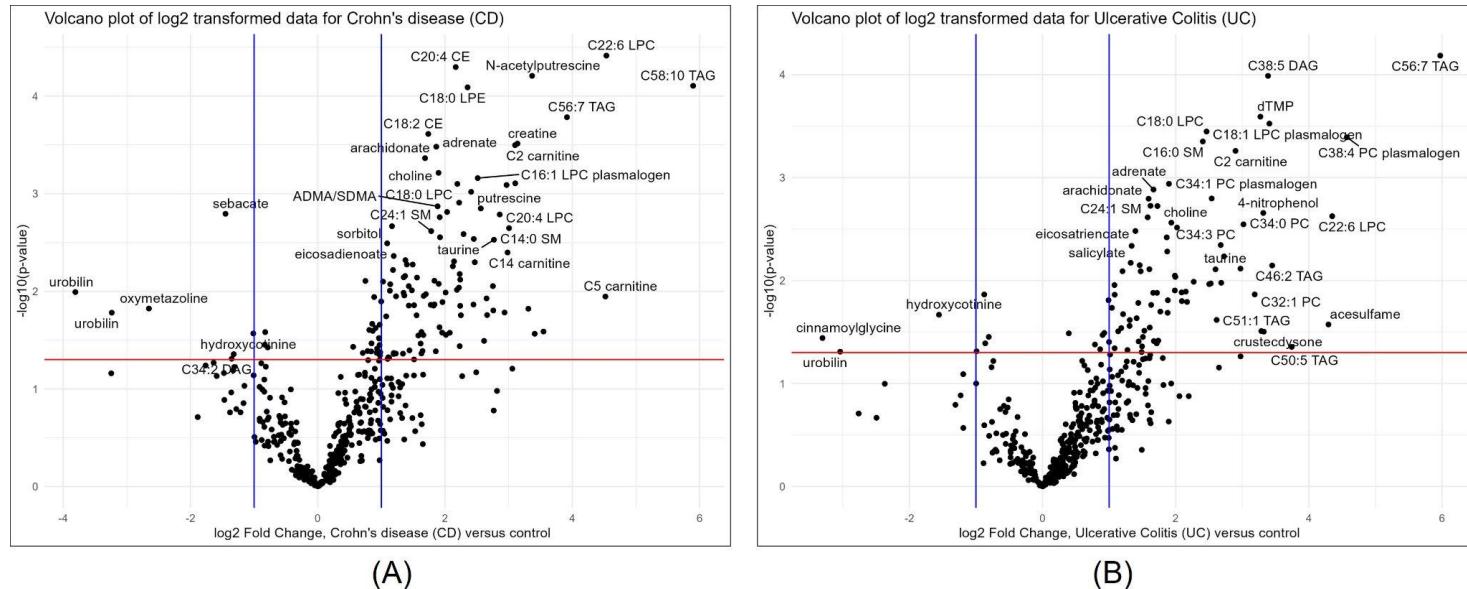
- A. Metadata
- B. Communities
- AOP
- CIRM Stem Cell Pathways
- COVID19
- Inborn Errors of Metabolism
- Lipids
  - LIPIDMAPS\_Federated...
  - LipidClassesTotal.rq
  - LipidsClassesCountPe..

**SPARQL results** (1 results in 0.08890000000037253 seconds)

IndividualLipidsPerClass
609

DEMO: <https://sparql.wikipathways.org/>

# Metabolomics results IBD study



**Figure 7.** Volcano plots representing altered metabolites in the stool of IBD patients, with the vertical axis depicting the log2FC (blue line showing the threshold absolute log2Fold change  $\geq 1$ ), versus the y-axis showing the  $-\log_{10}(p\text{-value})$  (red line depicting the p-value  $\leq 0.05$ ). (A) control versus Crohn's disease (CD) patients, (B) control group versus Ulcerative colitis (UC) patients.

# Metabolomics results IBD study

**Table 3.** Pathway Analysis results for pathways containing 4 or more significantly changed metabolites in Crohn's disease (CD) and ulcerative colitis (UC). Overlapping pathways between both disorders are highlighted with a #, significant p-values ( $\leq 0.05$ ) with a \*.

Disorder	Pathway ID	Pathway Title	# Sign. Metabolites in PW	P-value	# Proteins in PW
CD	WP3925	Amino acid metabolism	8	0.0000*	91
	WP2525	Trans-sulfuration, one-carbon metabolism, and related pathways*	8	0.0468*	67
	WP4723	Omega-3 / omega-6 fatty acid synthesis*	7	0.1263	15
	WP15	Selenium micronutrient network*	6	0.0000*	86
	WP661	Glucose homeostasis	6	0.1695	1
	WP4726	Sphingolipid metabolism: integrated pathway*	5	0.0008*	26
	WP3953	mRNA, protein, and metabolite induction pathway by cyclosporin A	5	0.1353	7
	WP550	Biogenic amine synthesis	5	0.1844	15
	WP5176	Disorders of bile acid synthesis and biliary transport	4	0.0109	20
	WP706	Sudden infant death syndrome (SIDS) susceptibility pathways	4	0.2163	159
UC	WP4723	Omega-3 / omega-6 fatty acid synthesis*	6	0.1777	15
	WP15	Selenium micronutrient network*	5	0.0002*	86
	WP4726	Sphingolipid metabolism: integrated pathway*	4	0.0075*	26
	WP2525	Trans-sulfuration, one-carbon metabolism, and related pathways*	4	0.0238*	67

# Metabolomics results IBD study

**Table 3.** Pathway Analysis results for pathways containing 4 or more significantly changed metabolites in Crohn's disease (CD) and ulcerative colitis (UC). Overlapping pathways between both disorders are highlighted with a #, significant p-values ( $\leq 0.05$ ) with \*

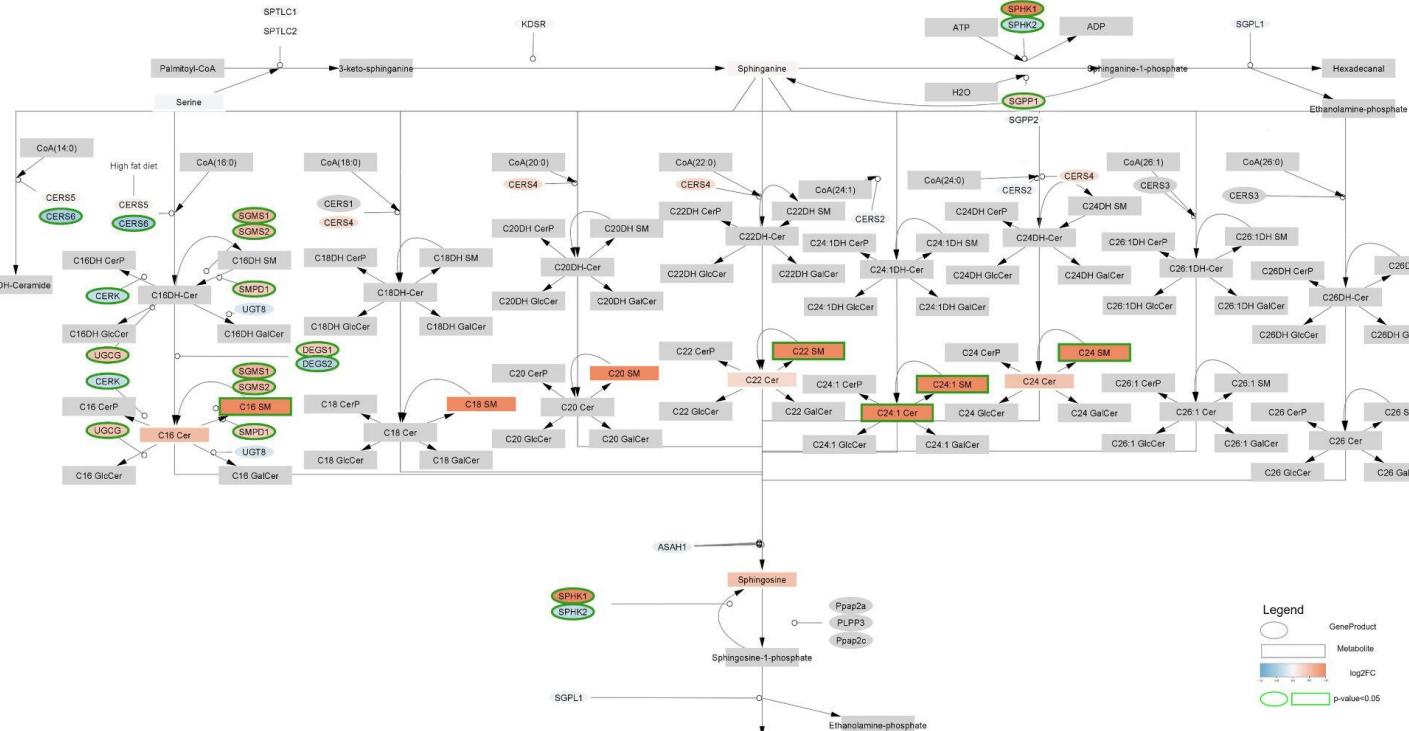
Disorder	Pathway ID	Pathway Title	# Sign. Metabolites in PW	P-value	# Proteins in PW
CD	WP3925	Amino acid metabolism	8	0.0000*	91
	WP2525	Trans-sulfuration, one-carbon metabolism, and related pathways*	8	0.0468*	67
	WP4723	Omega-3 / omega-6 fatty acid synthesis*	7	0.1263	15
	WP15	Selenium micronutrient network*	6	0.0000*	86
	WP661	Glucose homeostasis	6	0.1695	1
	WP4726	Sphingolipid metabolism: integrated pathway*	5	0.0008*	26
	WP3953	mRNA, protein, and metabolite induction pathway by cyclosporin A	5	0.1353	7
	WP550	Biogenic amine synthesis	5	0.1844	15
	WP5176	Disorders of bile acid synthesis and biliary transport	4	0.0109	20
	WP706	Sudden infant death syndrome (SIDS) susceptibility pathways	4	0.2163	159
UC	WP4723	Omega-3 / omega-6 fatty acid synthesis*	6	0.1777	15
	WP15	Selenium micronutrient network*	5	0.0002*	86
	WP4726	Sphingolipid metabolism: integrated pathway*	4	0.0075*	26
	WP2525	Trans-sulfuration, one-carbon metabolism, and related pathways*	4	0.0238*	67

# Metabolomics results IBD study

**Table 3.** Pathway Analysis results for pathways containing 4 or more significantly changed metabolites in Crohn's disease (CD) and ulcerative colitis (UC). Overlapping pathways between both disorders are highlighted with a #, significant p-values ( $\leq 0.05$ ) with a \*.

Disorder	Pathway ID	Pathway Title	# Sign. Metabolites in PW	P-value	# Proteins in PW
CD	WP3925	Amino acid metabolism	8	0.0000*	91
	WP2525	Trans-sulfuration, one-carbon metabolism, and related pathways*	8	0.0468*	67
	WP4723	Omega-3 / omega-6 fatty acid synthesis*	7	0.1263	15
	WP15	Selenium micronutrient network*	6	0.0000*	86
	WP661	Glucose homeostasis	6	0.1695	1
	WP4726	Sphingolipid metabolism: integrated pathway*	5	0.0008*	26
	WP3953	mRNA, protein, and metabolite induction pathway by cyclosporin A	5	0.1353	7
	WP550	Biogenic amine synthesis	5	0.1844	15
	WP5176	Disorders of bile acid synthesis and biliary transport	4	0.0109	20
	WP706	Sudden infant death syndrome (SIDS) susceptibility pathways	4	0.2163	159
UC	WP4723	Omega-3 / omega-6 fatty acid synthesis*	6	0.1777	15
	WP15	Selenium micronutrient network*	5	0.0002*	86
	WP4726	Sphingolipid metabolism: integrated pathway*	4	0.0075*	26
	WP2525	Trans-sulfuration, one-carbon metabolism, and related pathways*	4	0.0238*	67

# Multi-Omics integration

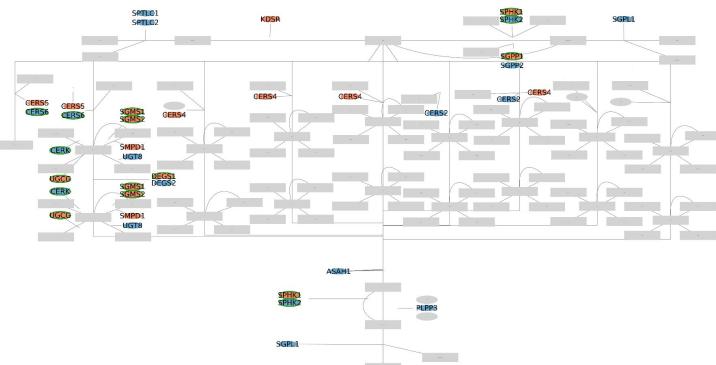
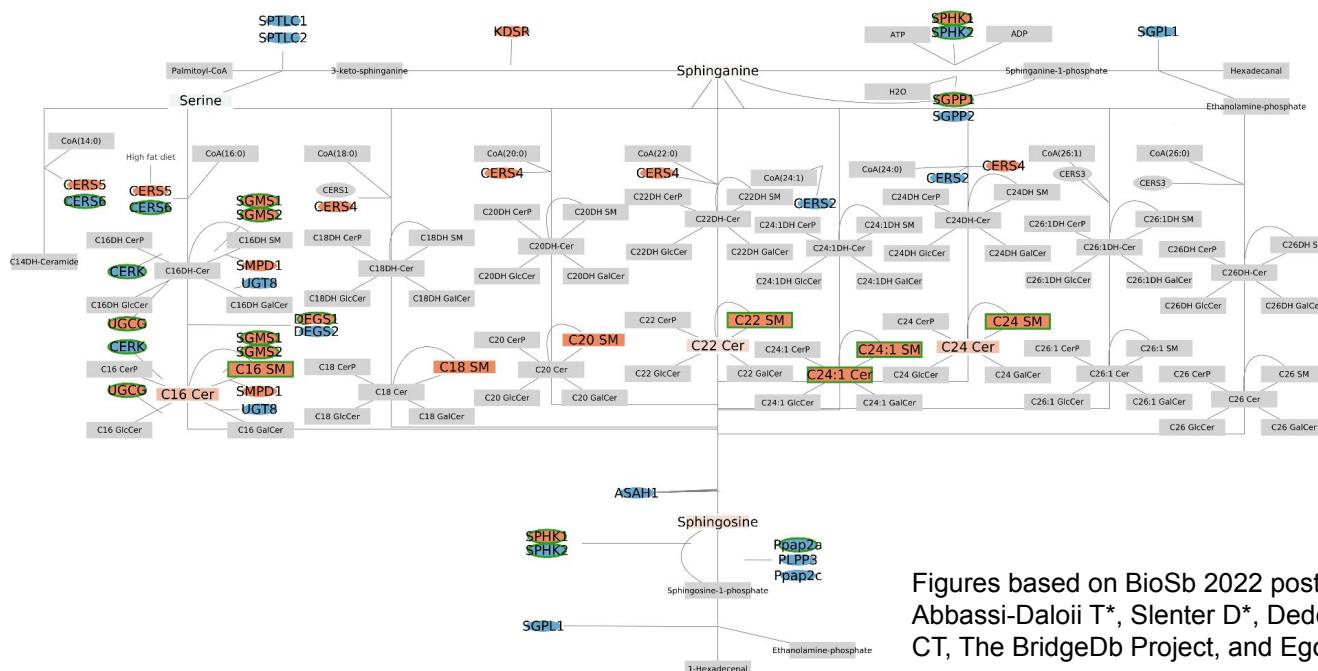


**Figure 8.** Sphingolipid metabolism: integrated pathway, <https://www.wikipathways.org/instance/WP4726>. The log2FC is indicated by a color gradient from blue (down-regulated) over white to red (up-regulated) on the nodes, significance ( $p\text{-value} < 0.05$ ) is represented by a light-green border and omics data type is shown with a rectangle (metabolomics) and an ellipse (transcriptomics) shape. Nodes without measurement in the dataset are colored gray.

# Secondary IDs problem

Growing issue in (meta) analysis of biological data

- (1) entries withdrawn/deleted from a database,
- (2) entries split/merged in a database,
- (3) entries referring to the same entity ('unknown' duplicates)



Enhanced multi-omics visualization with BridgeDb identifier mapping adding data to:

2 proteins and 13 metabolites

Figures based on BioSb 2022 poster, authors:  
Abbassi-Daloii T\*, Slenter D\*, Dede Sener D, Basaric H, Kutmon M, Evelo CT, The BridgeDb Project, and Egon Willighagen

# Cytoscape Demo

Required:

- a. Install Cytoscape
- b. Get the WikiPathways App for Cytoscape
- c. Download the 'extended ID' dataset

# Cytoscape Demo

The screenshot shows the Cytoscape application interface. The menu bar includes File, Edit, View, Select, Layout, Apps, Tools, and Help. The toolbar contains icons for network operations like DSGNN, network visualization, file operations (Save, Import), and search functions. On the left, there are tabs for Network (selected) and Style, with a Filter button below. The main workspace displays a recent session titled "WP4726". A "WikiPathways Search" dialog is open in the foreground, containing a search bar with "WP4726", a checkbox for "Only: Unspecified", and a table with columns for Pathway, Species, and ID. The first row in the table is highlighted with a blue background and shows "Sphingolipid metabolism: integrated pathway", "Homo sapiens", and "WP4726". At the bottom of the dialog are buttons for "Import as Pathway" and "Import".

Pathway	Species	ID
Sphingolipid metabolism: integrated pathway	Homo sapiens	WP4726

Recent Sessions

WP4726

WikiPathways Search

WP4726

Only: Unspecified

Pathway Species ID

Sphingolipid metabolism: integrated pathway Homo sapiens WP4726

Import as Pathway Import

# Cytoscape Demo

Network ▾

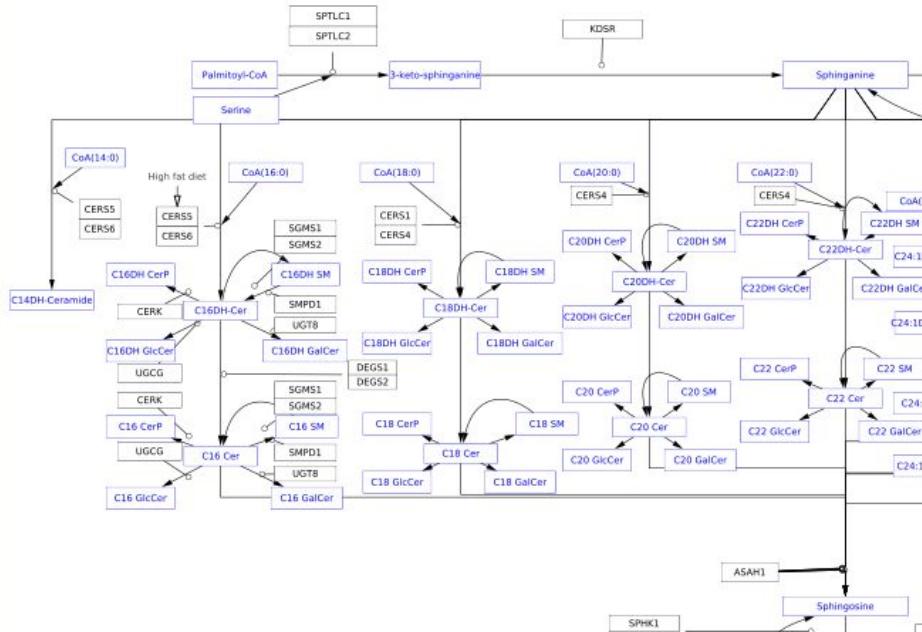
WP4726

1 of 1 Network selected

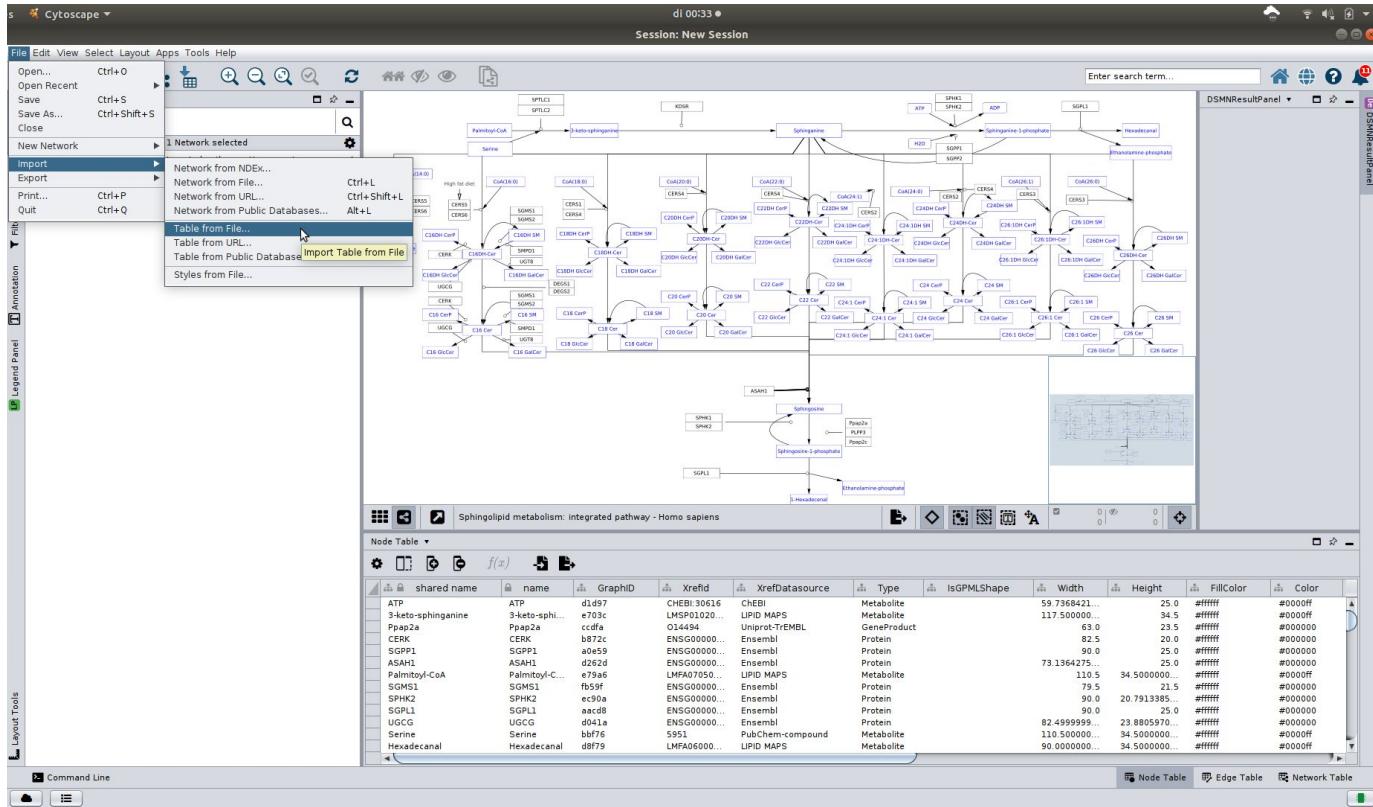
Network Sphingolipid metabolism: integrated pathway - Homo sapiens 1

Sphingolipid metabolism: integrated pathway - Homo sa... 204 166

The screenshot shows the Cytoscape interface with a network titled "Sphingolipid metabolism: integrated pathway - Homo sapiens". The network contains 204 nodes and 166 edges. Nodes represent various metabolites and enzymes involved in sphingolipid metabolism, such as Palmitoyl-CoA, Serine, 3-keto-sphinganine, SPTLC1, SPTLC2, KDSR, Sphinganine, CoA(14:0), C14DH-Ceramide, C16DH CerP, C16DH-Cer, C16DH SM, C16DH GicCer, C16 CerP, C16 Cer, C16 SM, C16 GalCer, C16 GlcCer, SGMS1, SGMS2, SMPD1, UGT8, DEGS1, DEGS2, C18DH CerP, C18DH Cer, C18DH SM, C18DH GicCer, C18DH GalCer, C18 DH G, C18 CerP, C18 Cer, C18 SM, C18 GalCer, C18 GlcCer, C20DH CerP, C20DH Cer, C20DH SM, C20DH GicCer, C20DH GalCer, C20 DH G, C20 CerP, C20 Cer, C20 SM, C20 GalCer, C20 GlcCer, C22DH CerP, C22DH Cer, C22DH SM, C22DH GicCer, C22DH GalCer, C22 DH G, C22 CerP, C22 Cer, C22 SM, C22 GalCer, C22 GlcCer, C24:1DH, C24:1G, ASAH1, SPHK1, and Sphingosine. The network is organized into several main pathways, including the synthesis of sphinganine from palmitoyl-CoA and serine, the generation of various ceramides and glycerophospholipids (GicCer, GalCer, GlcCer) from sphinganine, and the degradation of sphingolipids via enzymes like CERK, SGMS1, SGMS2, SMPD1, UGT8, DEGS1, DEGS2, and ASAH1.



# Cytoscape Demo



- File
- Import
- Table from file

Select the file with extended IDs (including ChEBI)

# Cytoscape Demo

Import Columns From Table

**Target Table Data**

Where to Import Table Data: To a Network Collection

**Select a Network Collection**

Network Collection: Sphingolipid metabolism: integrated pathway - Homo sapiens

Import Data as: Node Table Columns

Key Column for Network: ChEBI

Case Sensitive Key Values:

**Preview**

Click on a column to edit it.

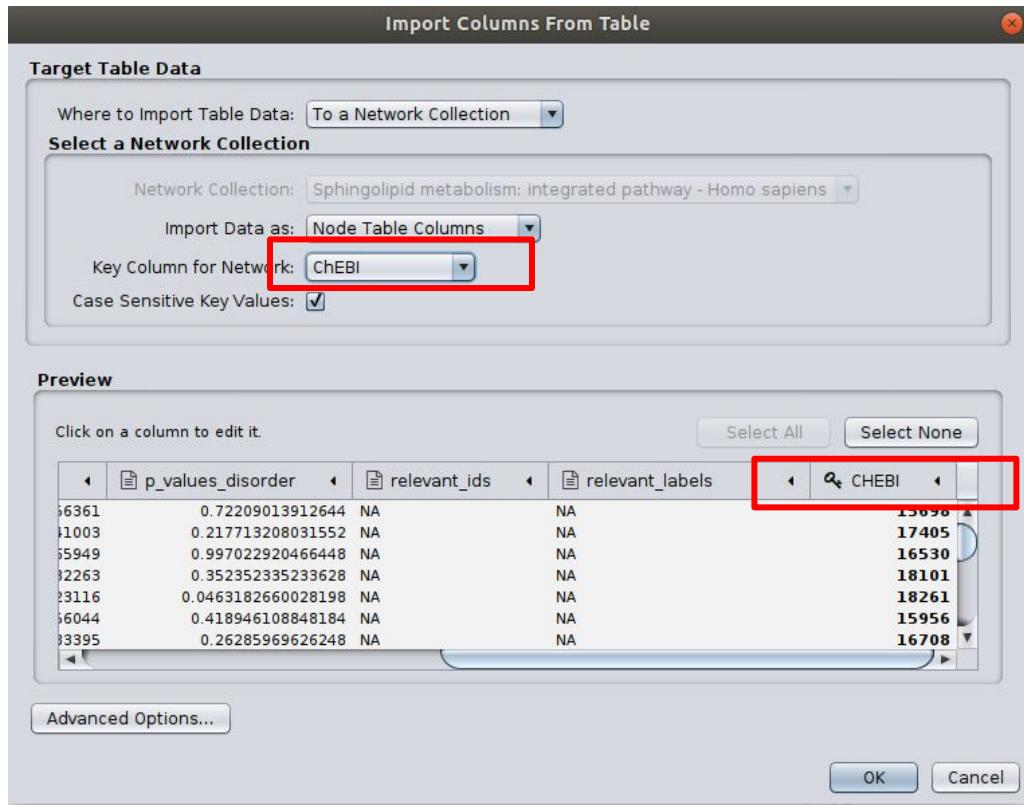
Select All    Select None

	p_values_disorder	relevant_ids	relevant_labels	ChEBI
6361	0.72209013912644	NA	NA	15698
11003	0.217713208031552	NA	NA	17405
5949	0.997022920466448	NA	NA	16530
32263	0.352352335233628	NA	NA	18101
31116	0.0463182660028198	NA	NA	18261
6044	0.418946108848184	NA	NA	15956
33395	0.26285969626248	NA	NA	16708

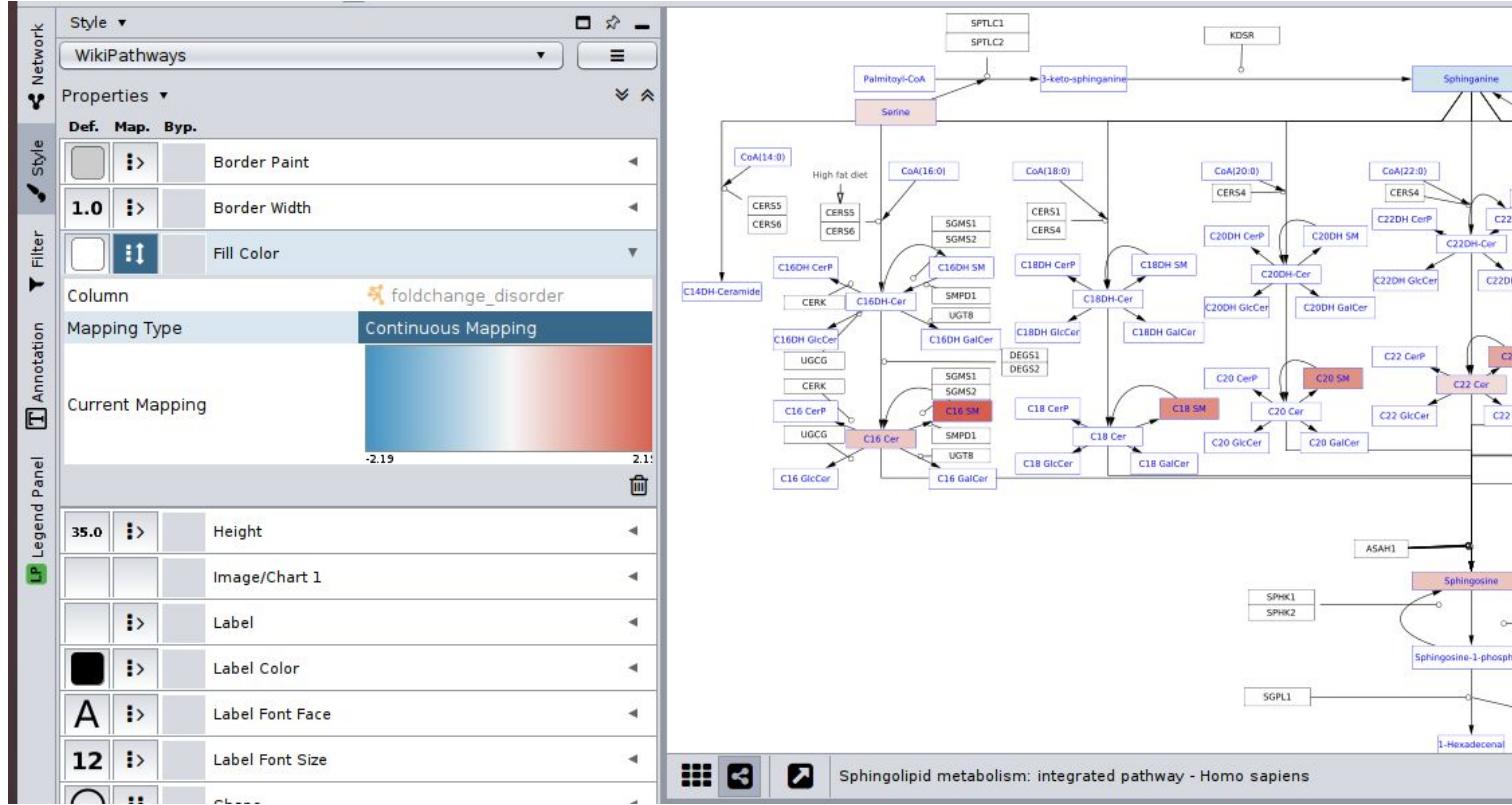
Advanced Options...

OK    Cancel

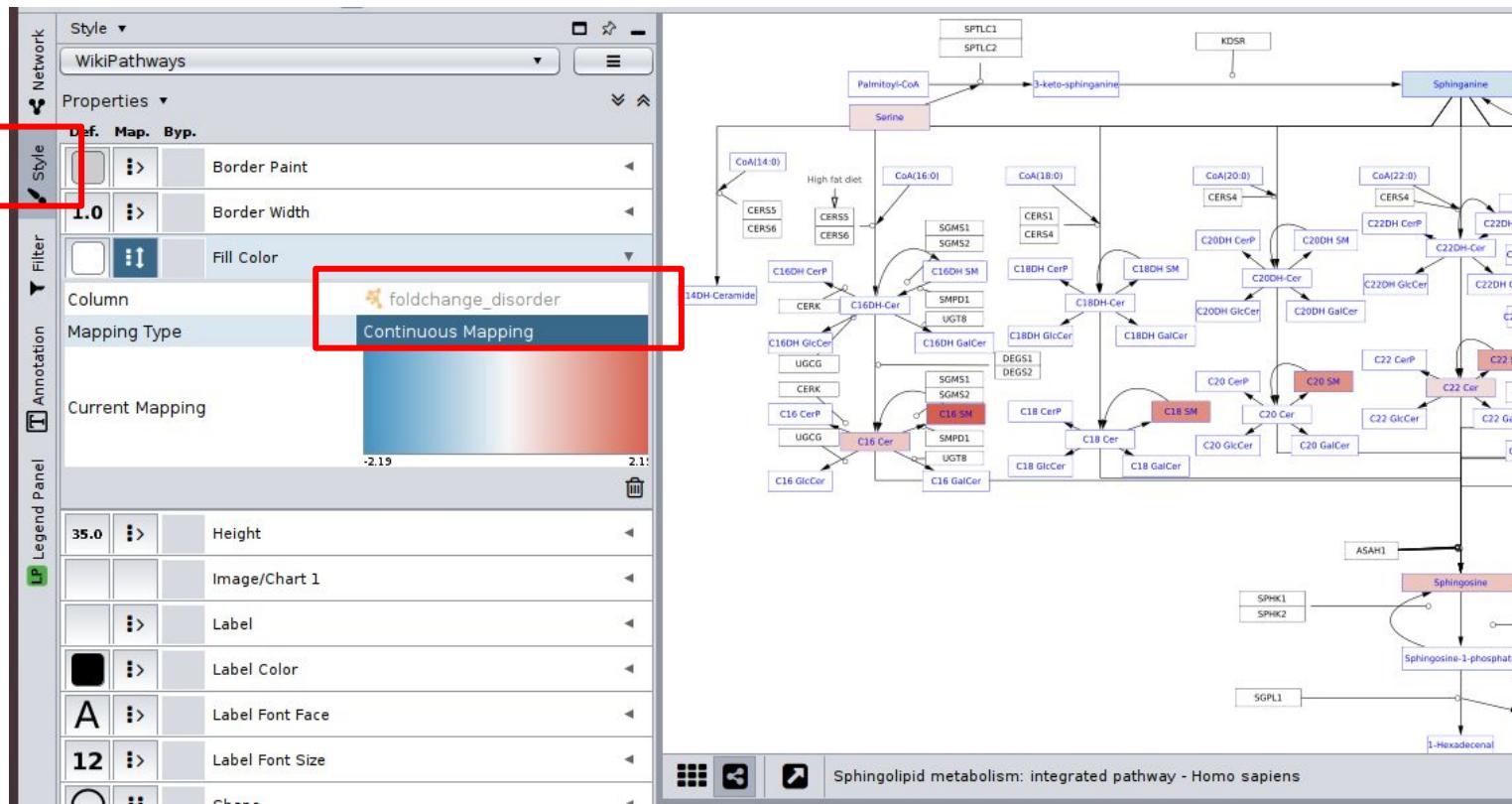
# Cytoscape Demo



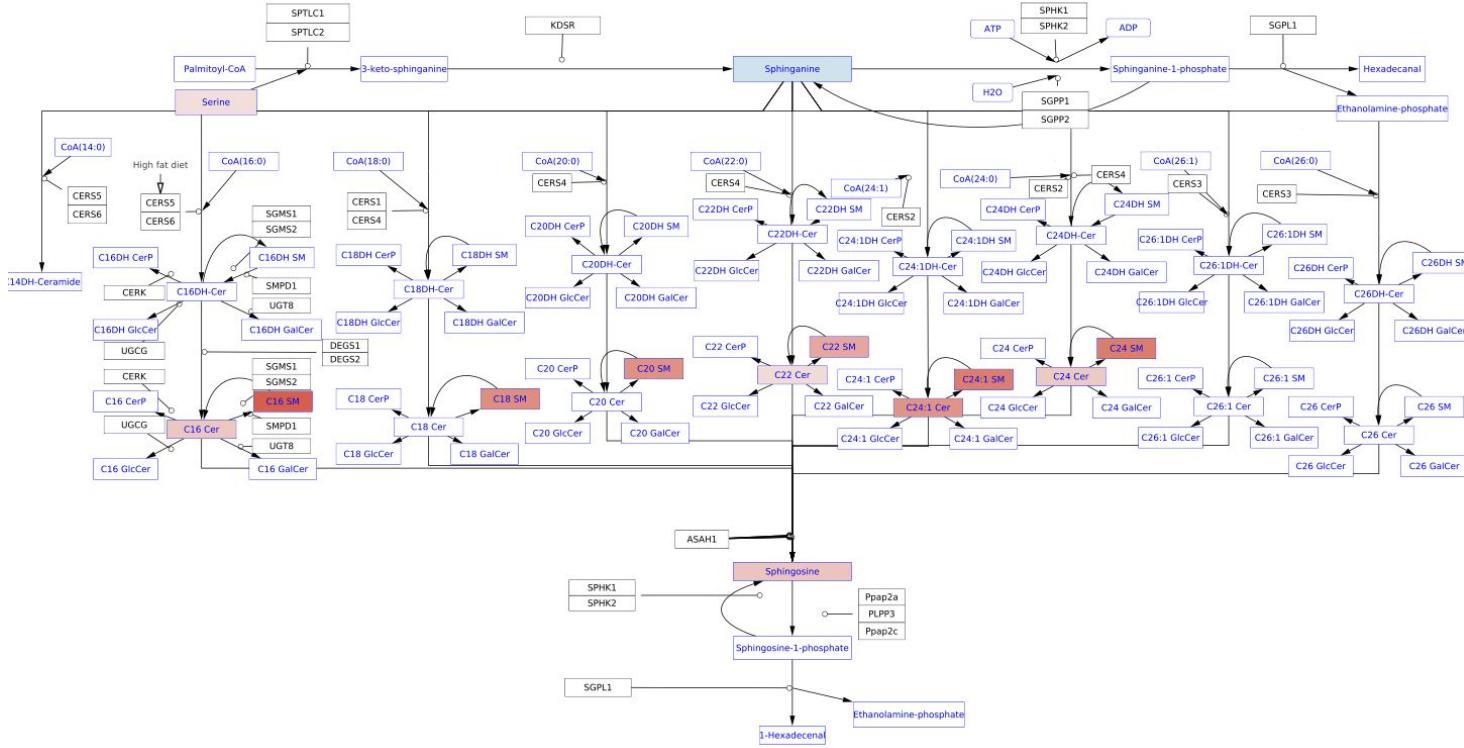
# Cytoscape Demo



# Cytoscape Demo



# Cytoscape Demo



# Multi-Omics Integration & Wrap-up

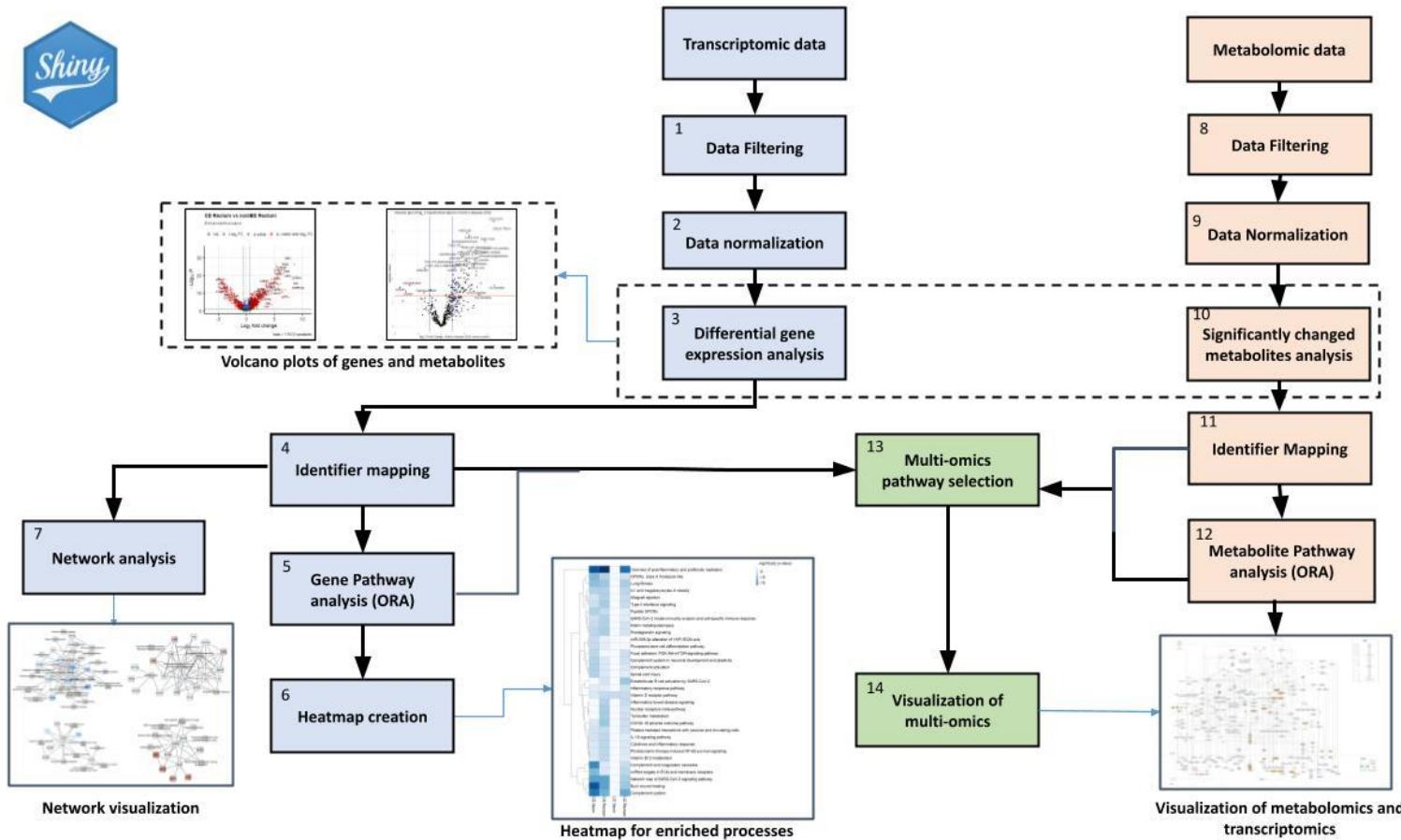
**Denise Slenter**

ORCID: 0000-0001-8449-1318

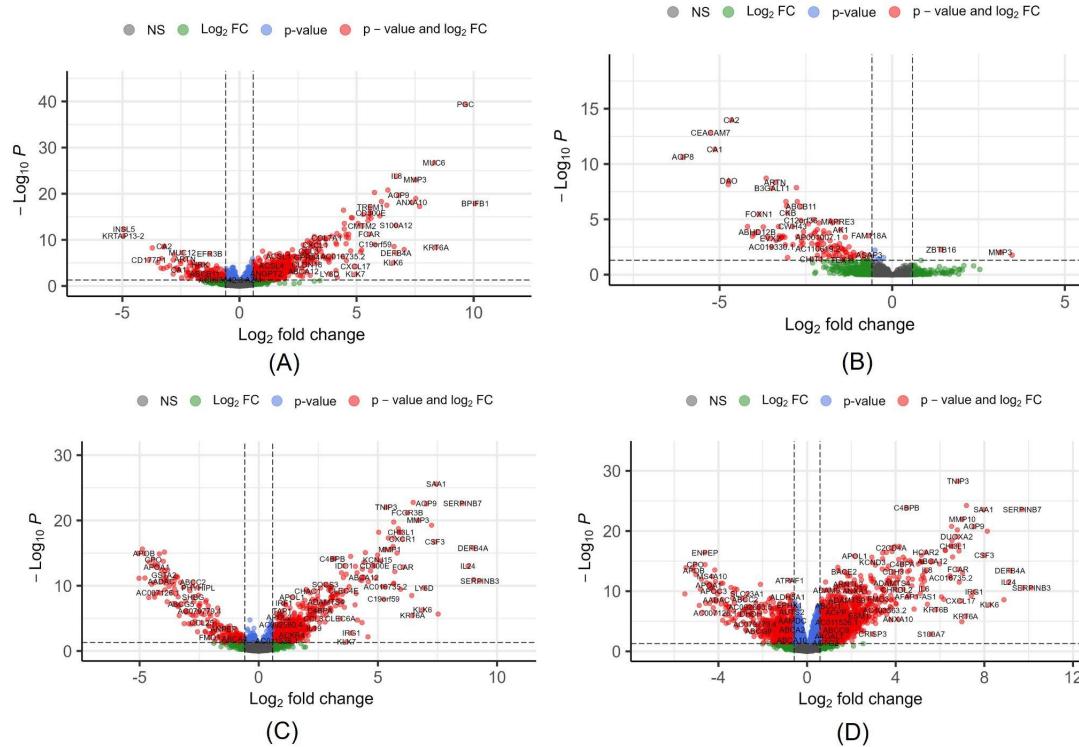
Tuesday July 2nd, 2024



# Workflow transcriptomics and metabolomics integration



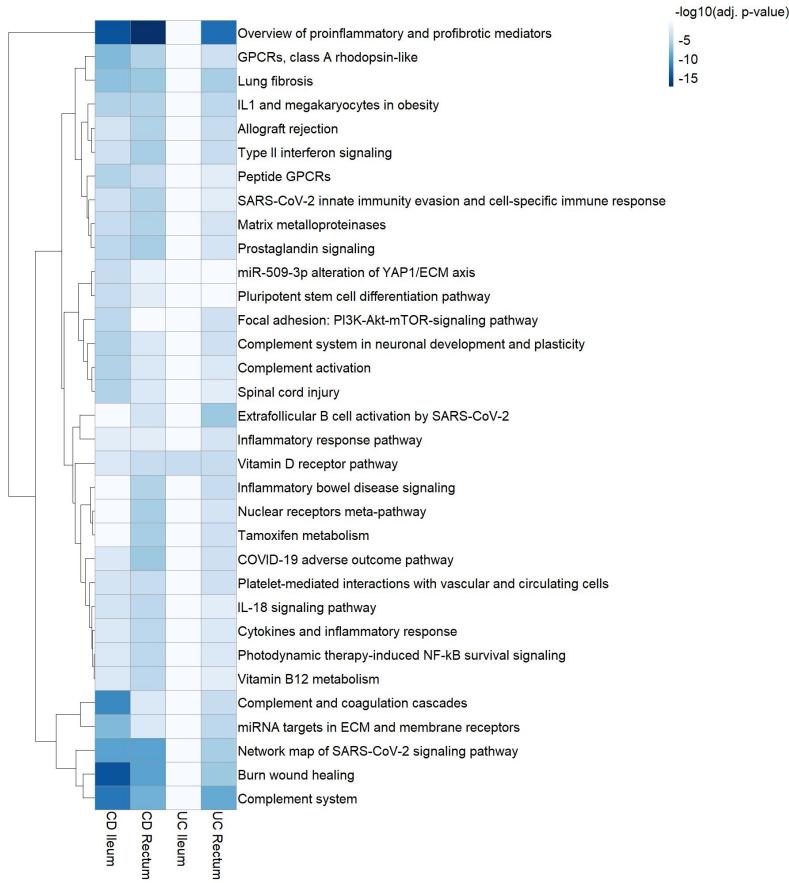
# Transcriptomics results IBD study



**Figure 3.** Volcano plots for differentially expressed genes in the ileum and rectum. (A) Crohn's disease patients versus controls on the ileum (B) Ulcerative colitis patients versus controls on the ileum (C) Crohn's disease patients versus controls on the rectum (D) Ulcerative colitis patients versus controls on the rectum. CD = Crohn's disease, UC = Ulcerative Colitis, non-IBD = healthy controls. The x-axis represents log<sub>2</sub>FC and the y-axis represents the corresponding significance given by  $-\log_{10}(p\text{-value})$ . 17,670 genes were analyzed.



# Transcriptomics results IBD study



**Figure 4.** Comparison of altered pathways in Crohn's disease and ulcerative colitis on both the rectum and ileum samples. Rows represent enriched pathways while columns represent comparison pairs including disease and the biopsy location. Dark blue indicates a more significant pathway while light blue represents a less significant pathway.

# Transcriptomics and Metabolomics differences

## Size:

- Genome: 3.055 billion nucleic acid pairs within DNA
- Epigenome: high-throughput techniques covers less than 2% of the genomic sites where methylation occurs
  - From each gene several transcripts may arise, which can be categorized as (protein)-coding and (long and short) non-coding RNA
- Proteomics: minimum of 70.000 proteoforms to 1.5 million
- (endogenous) metabolome: into the thousands for vertebrates and tens or hundreds of thousands for plants
- Fluxomics: ??? combination of all metabolic reactions possible, as well as their involved enzyme kinetics.

## Dynamics

## Tissue (distribution)

## Model (organism)

# Transcriptomics and Metabolomics differences

