AN INTERACTIVE PLATFORM TO EXPLORE AND INTEGRATE MULTI-OMICS DATA

Joe Wandy and Ronan Daly

http://graphomics.glasgowcompbio.org/

BACKGROUND

An increasing number of studies now produce **multiple omics measurements** that require sophisticated computational methods for analysis.

Each omics data can be examined separately but **integrating multiple omics** allows for a deeper understanding and insights to be gained from the study.

In particular, data integration can be performed **horizontally** by mapping different entities to shared reactions and pathways.

HORIZONTAL INTEGRATION

Web-based tools to perform horizontal integration using pathways have been gaining popularity.

MetaboAnalyst provides a functionality to map genes and metabolites to metabolic pathways and performs pathway enrichment analysis.

3Omics accepts human-only transcriptomics, proteomics and metabolomics datasets and performs pathways mapping and other analyses such as correlation and gene ontology (GO) analyses.
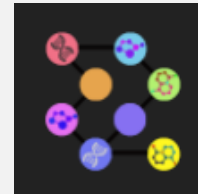
PaintOmics3 performs a complete integration of multiple data types to KEGG pathways, allowing for the enrichment and clustering analyses of pathways, and network visualisation.
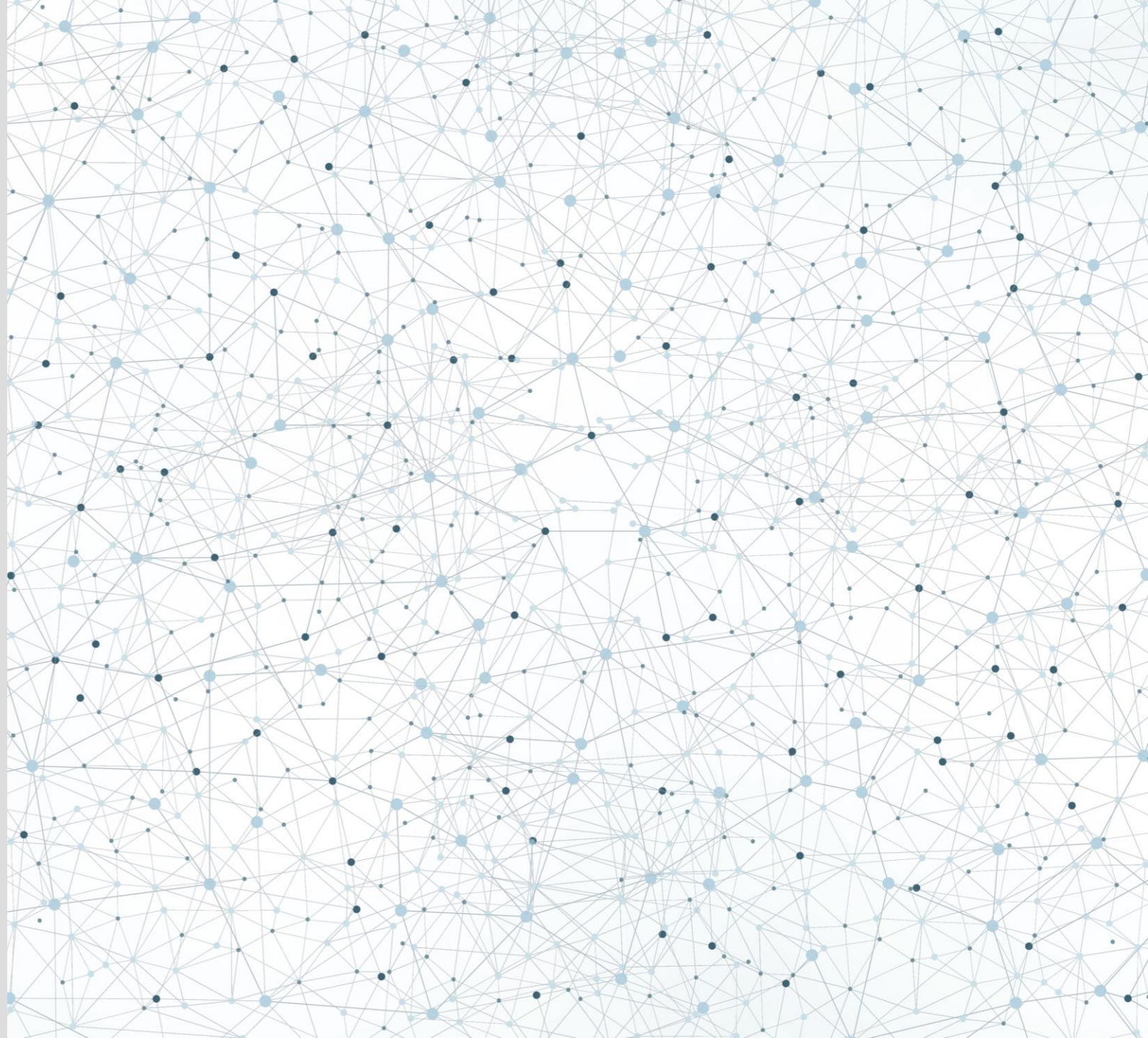
PROBLEM 1

- Existing tools often present data integration results as either static tables, or complex network graphs
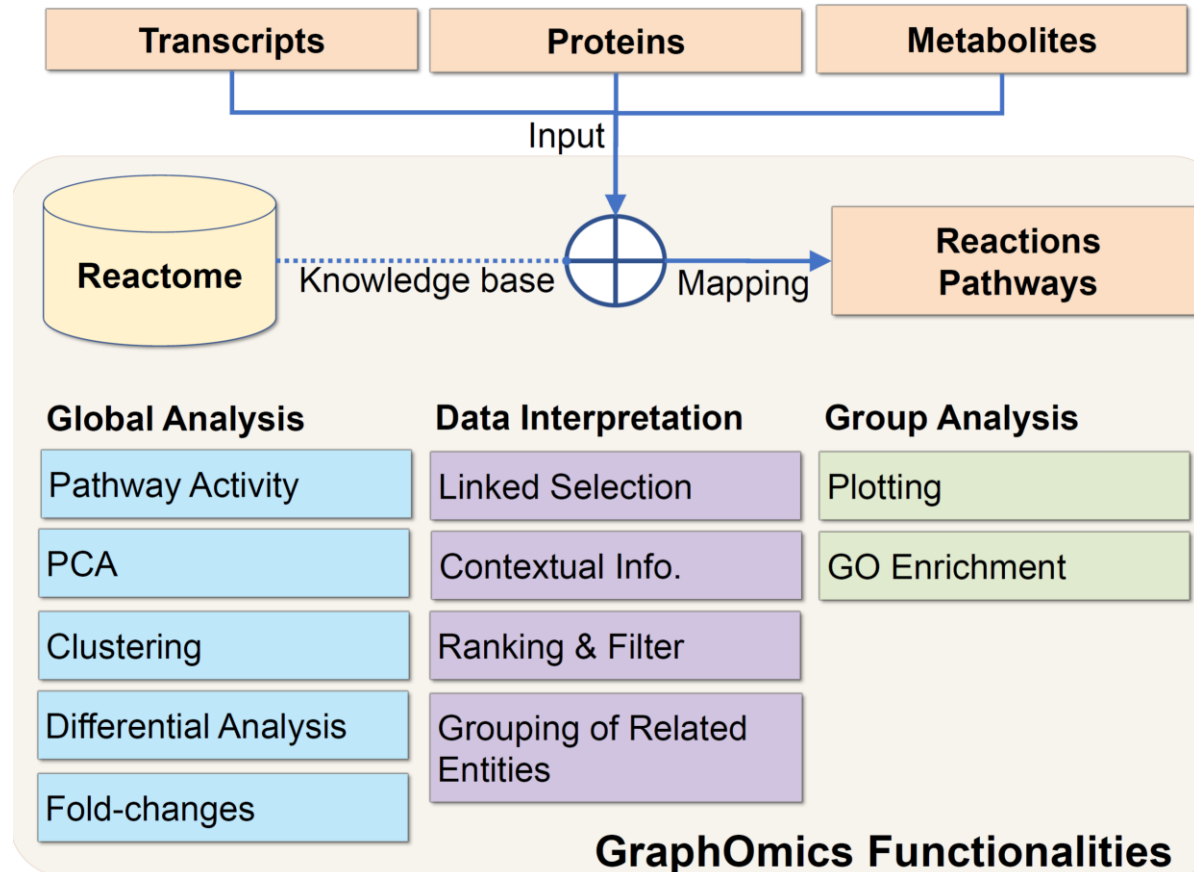
- More data = bigger tables

PROBLEM 2

- Existing tools often present data integration results as either static tables, or complex network graphs

- More data = bigger network ('hairball effect')

GraphOmics

- Accepts measurements of transcripts, proteins and metabolites.

- Performs data integration horizontally using Reactome as the graph knowledge base.

- Provides an interactive platform that Map entity onto Reactome reactions and pathways emphasising **interactivity** and **biological contexts.**

DATA INTEGRATION

- Horizontal integration through an automated mapping procedure written in Neo4j, the graph database used for Reactome.

- Retrieves the connections between transcripts, proteins, metabolites to reactions and pathways of the given species in Reactome.

- Entities in this graph are connected to one another. The connections are shown in interactive tables.

reactome

CONTEXTUAL INFORMATION PANEL

The info panel provides additional contextual information for selected entries in the Data Browser.

(A) An example info panel entry for the transcript identified by the gene *Aldh1a2*
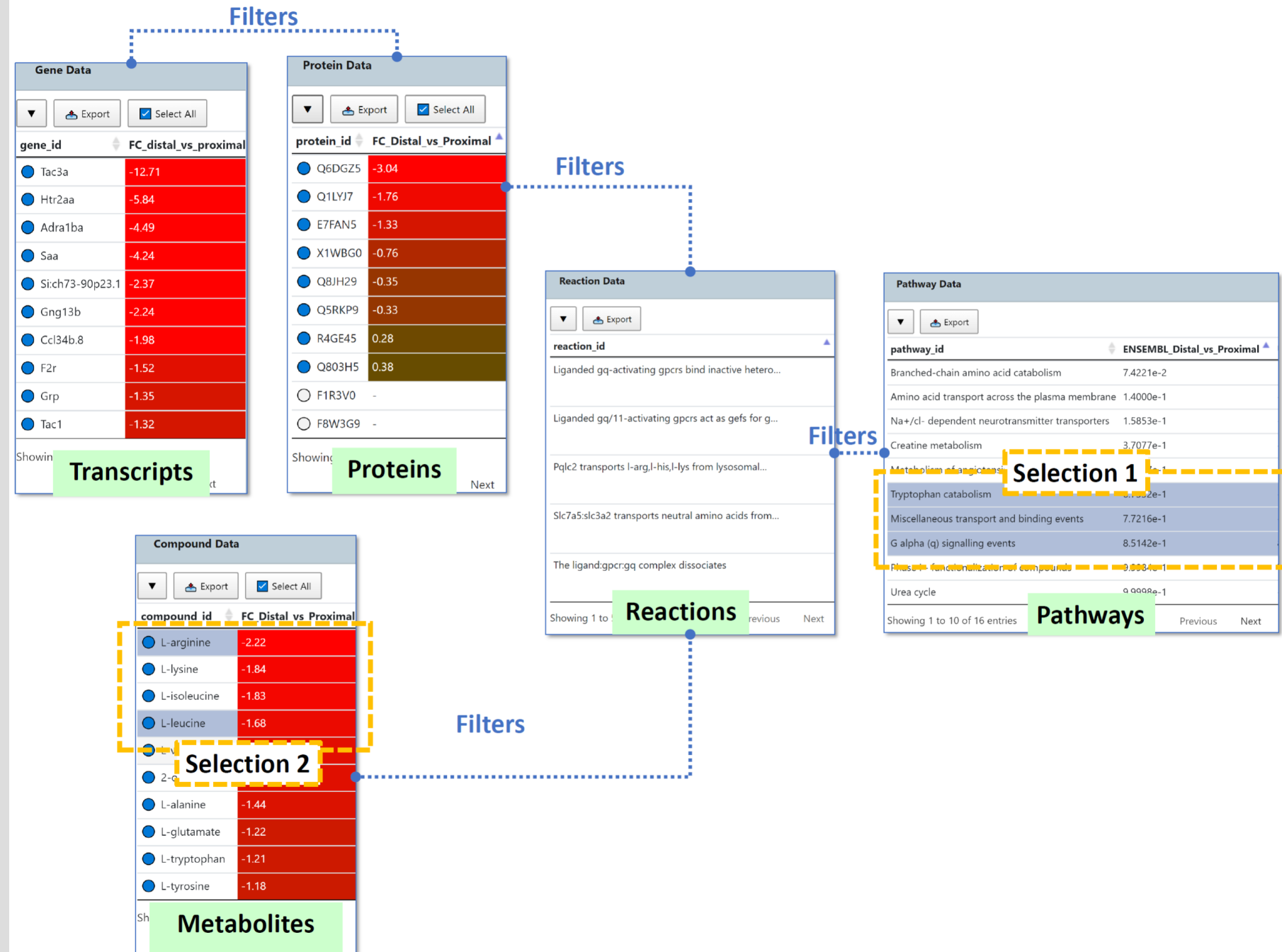
(B) The measurements of *Aldh1a2* in the data.

(C) An interactive pathway diagram with expression data mapped onto it.
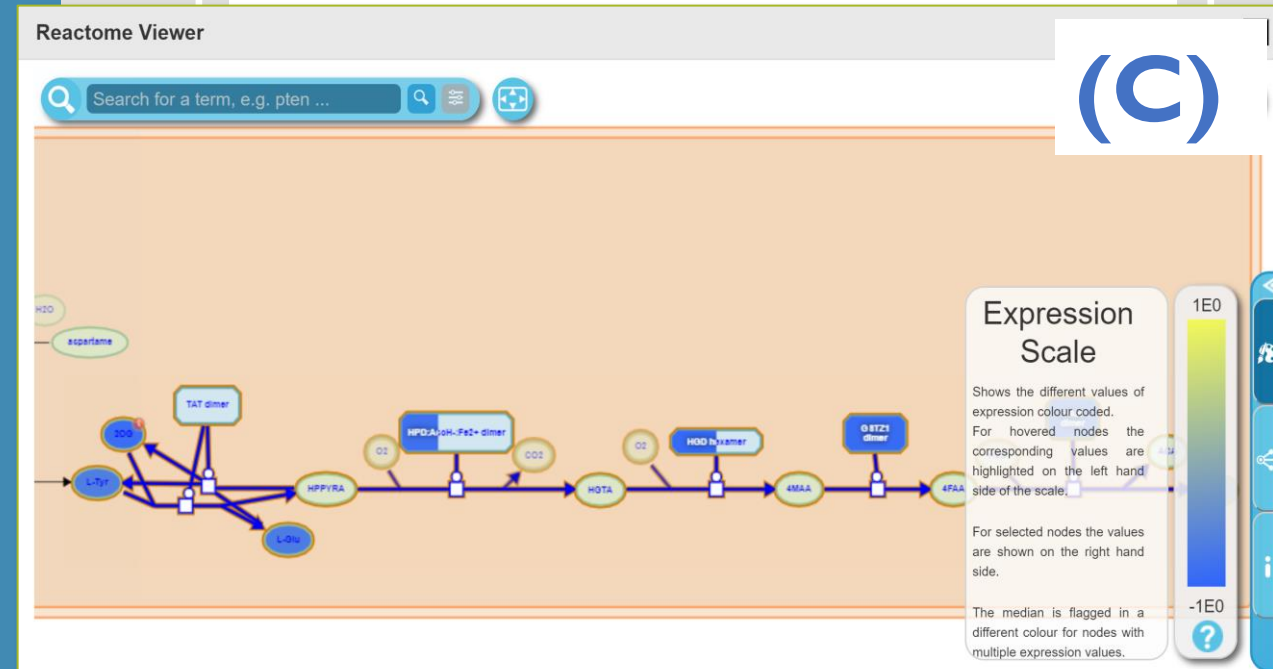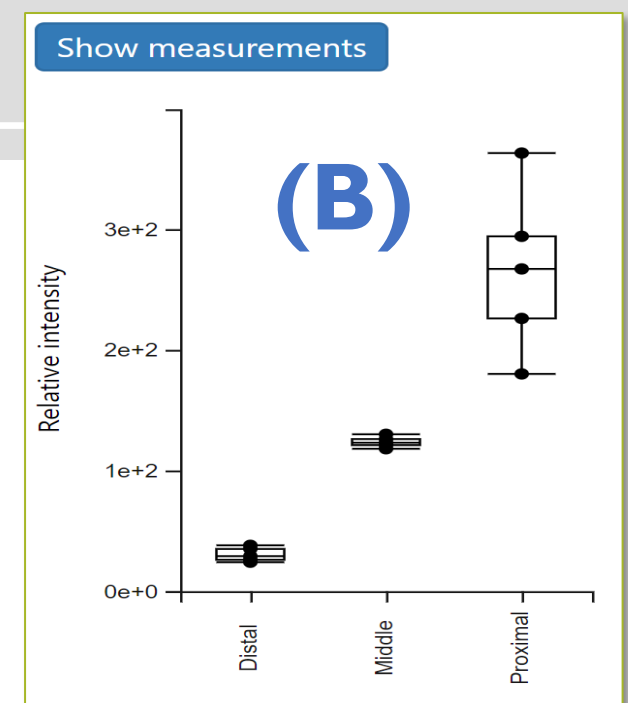
Gene (1/1)

Aldh1a2

📝 Annotate **(A)**

Description: aldehyde dehydrogenase 1 family, member A2

Species: danio_rerio

Description: This protein belongs to the aldehyde dehydrogenase family of proteins. The product of this gene is an enzyme that catalyzes the synthesis of retinoic acid (RA) from retinaldehyde. Retinoic acid, the active derivative of vitamin A (retinol), is a hormonal signaling molecule that functions in developing and adult tissues. The studies of a similar mouse gene suggest that this enzyme and the cytochrome CYP26A1, concurrently establish local embryonic retinoic acid levels which facilitate posterior organ development and prevent spina bifida. Four transcript variants encoding distinct isoforms have been identified for this gene. [provided by RefSeq, May 2011]

Link to Ensembl

Link to GeneCard

Show measurements

(B)

Reactome Viewer

(C)

Search for a term, e.g. pten ...

Expression Scale

Shows the different values of expression colour coded. For hovered nodes the corresponding values are highlighted on the left hand side of the scale.

For selected nodes the values are shown on the right hand side.

The median is flagged in a different colour for nodes with multiple expression values.

FLEXIBLE QUERY BUILDER

- The Query Builder is used to filter entities of data tables by specifying rules that will be concatenated using a logical AND operator.

- Example: Filter for transcripts and proteins that are both statistically significant (p-values less than 0.05) and having transcript fold changes at least 0.5 both ways

Define rules for filtering of data tables here. Rules will be concatenated using an AND logical operator.

AND   **Add rule**

| genes: padj_distal_vs_proximal | ⌄ | less or equal | 0.05 | **Delete** |

| proteins: padj_Distal_vs_Proximal | ⌄ | less or equal | 0.05 | **Delete** |

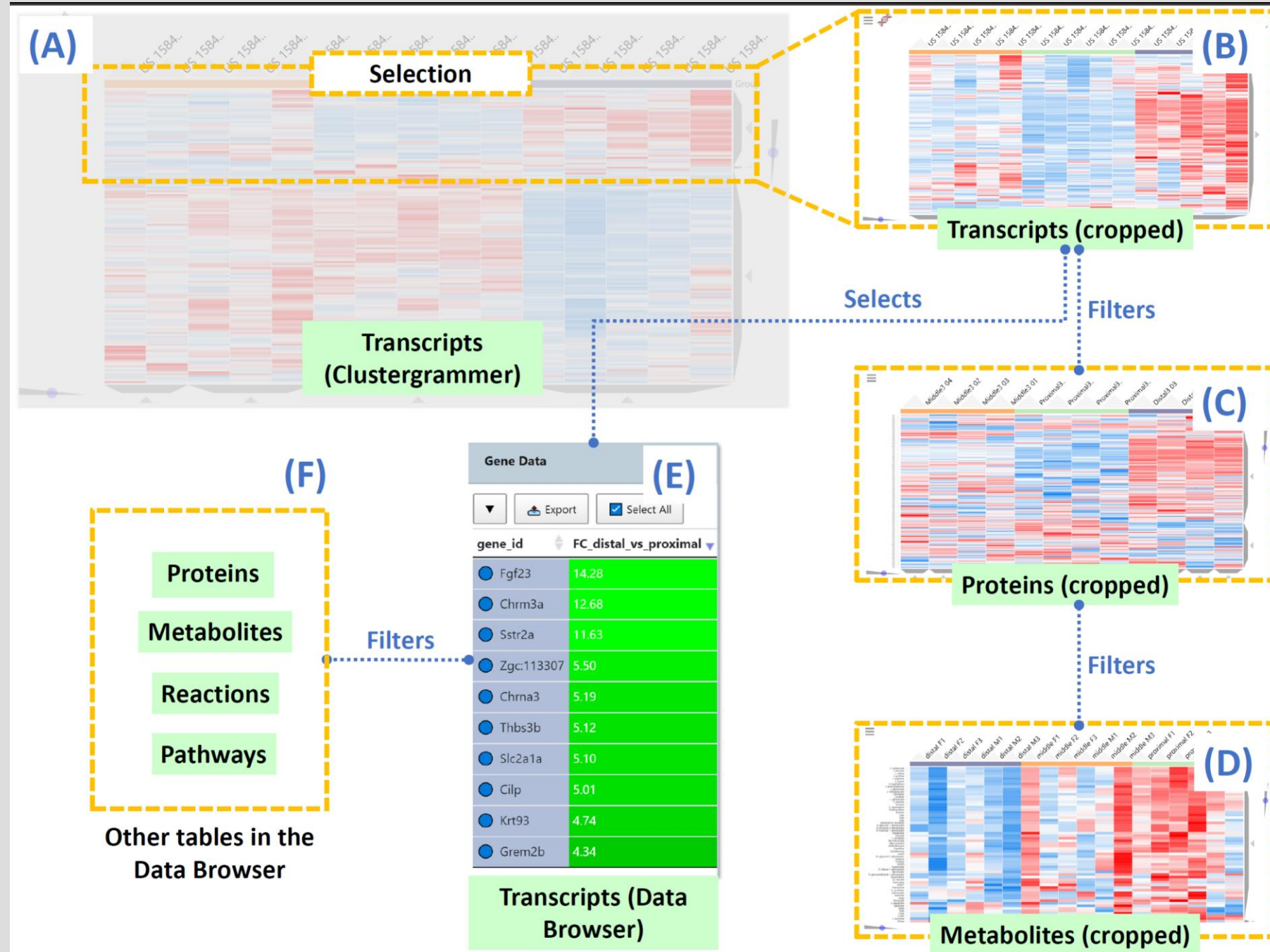| genes: FC_distal_vs_proximal | ⌄ | not between ⌄ | -0.5 | 0.5 | **Delete** |

Apply   **Reset**

INTERACTIVE CLUSTERING AND HEATMAPS

Interactive clustering and heatmap is available for each omics type.

Selecting a cluster in the heatmap will also select related entities in views and othe heatmaps.

OTHER FEATURES

- Differential expression analysis (t-test, limma, DeSEQ2), or load the results of any DE analysis from outside.
- Gene ontology (GO) analysis.
- Pathway activity level analysis (single/multi-omics).
- Principal Component Analysis.

Table 1 A comparison of GraphOmics to other Web-based multi-omics systems.

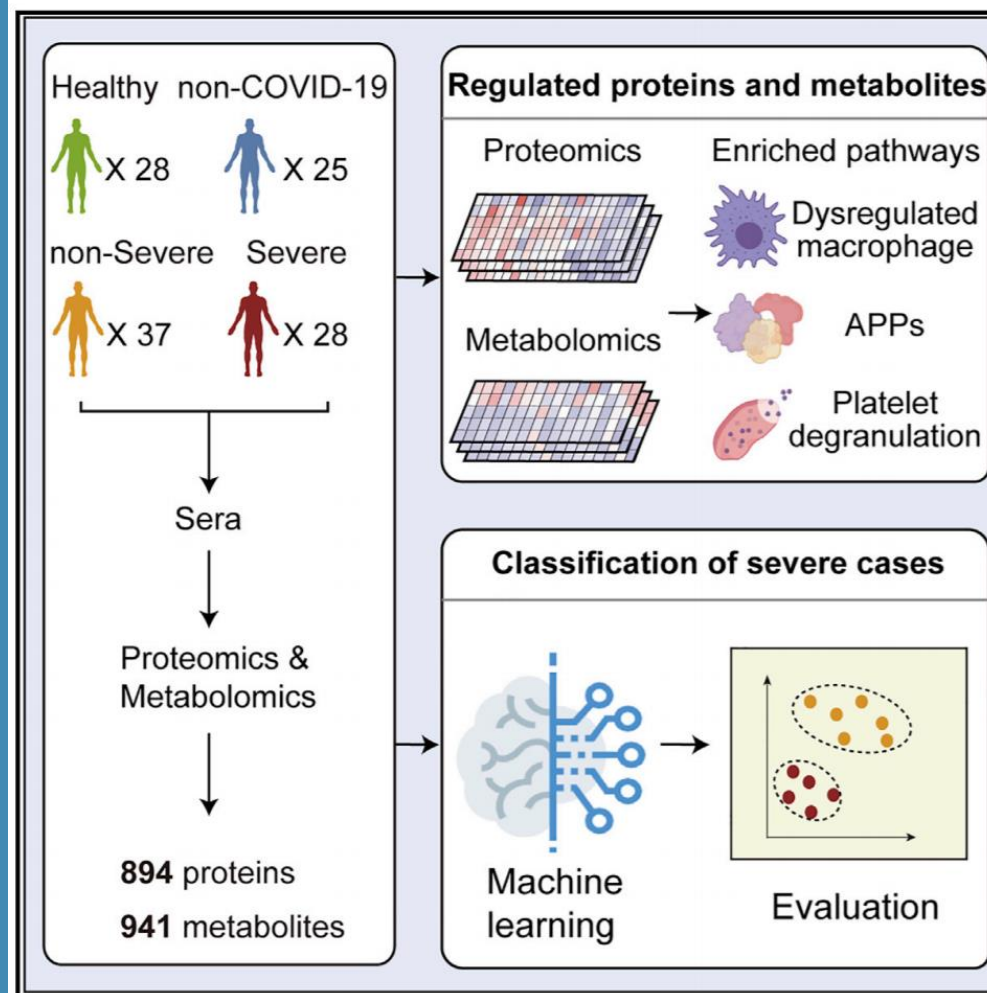| Tool	Database	Omics Types	Analysis Types	Results Presentation
GraphOmics	Reactome	- Transcripts - Proteins - Metabolites	- Pathway enrichment: ORA, GSEA, PLAGE, Reactome Analysis Service - GO enrichment	- Interactive tables - Interactive pathway diagrams - Interactive heatmaps - Interactive clustering
MetaboAnalyst	KEGG	- Genes - Metabolites	- Pathway enrichment: ORA, Topology	- Static tables - Static pathway diagrams
3Omics	KEGG (human only)	- Transcripts - Proteins - Metabolites	- Correlation analysis - Coexpression profiles - Phenotype analysis - Pathway enrichment (ORA) - GO enrichment	- Static tables - Static pathway diagrams - Static heatmaps
PaintOmics3	KEGG	- Transcripts - Proteins - Metabolites - DNase-seq - miRNA-seq	- Pathway enrichment (ORA) - Clustering of pathways	- Interactive tables - Interactive pathway diagrams - Interactive heatmaps

CASE STUDY

Dual-omics Covid19 Data

https://graphomics.glasgowcompbio.org/app/linker/explore_data/6

Cell

Proteomic and Metabolomic Characterization of COVID-19 Patient Sera

Graphical Abstract

Healthy non-COVID-19

X 28   X 25

non-Severe   Severe

X 37   X 28

Sera

Proteomics & Metabolomics

894 proteins
941 metabolites

Regulated proteins and metabolites

Proteomics   Enriched pathways

Dysregulated macrophage

Metabolomics   APPs

Platelet degranulation

Classification of severe cases

Machine learning   Evaluation

Authors

Bo Shen, Xiao Yi, Yaoting Sun, ...,
Huafen Liu, Haixiao Chen, Tiannan

Correspondence

zhuyi@westlake.edu.cn (Y.Z.),
liuhf1@dazd.cn (H.L.),
chenhx@enzemed.com (H.C.),
guotiannan@westlake.edu.cn (T.G.

In Brief

Proteomic and metabolomic analys
COVID-19 sera identifies differentia
expressed factors that correlate w
disease severity and highlights
dysregulation of multiple immune a
metabolic components in clinically
patients.

ACTIVATION OF ACUTE PHASE PROTEINS AND COMPLEMENT PATHWAYS.

- The complement system is a part of the immune system that enhances (complements) the ability of antibodies and phagocytic cells to clear microbes and damaged cells from an organism.


Pathway Data

	pathway_id	UniProt_severe_vs_healthy	CON
va...	Neutrophil degranulation	6.5757e-1	-
	Platelet degranulation	3.6178e-2	7.41
iva...	Post-translational protein phosphorylation	9.9989e-1	-
	Regulation of complement cascade	3.0948e-2	-
	Regulation of insulin-like growth factor (igf) tr...	1.0000e+0	-
	Cytosolic trna aminoacylation	-	3.34
	Activation of c3 and c5	2.5138e-6	
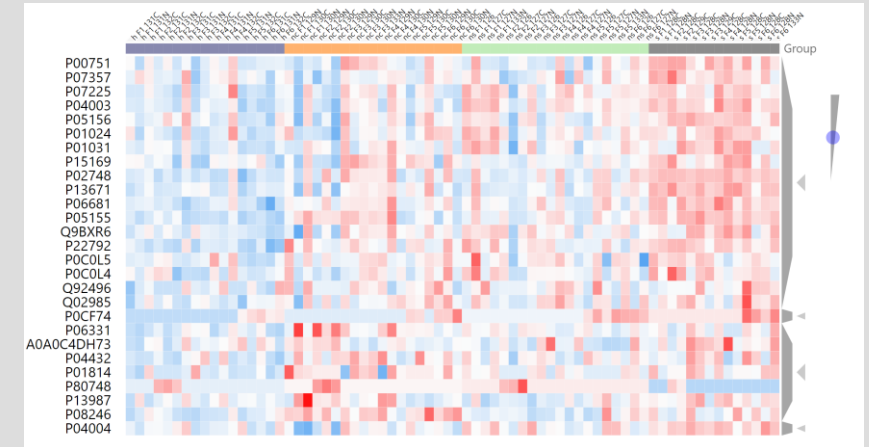
# SUPPRESSED PLATELET DEGRANULATION IN SEVERE PATIENTS

- Downregulation of many proteins involved in platelet degranulation.

- Platelets are produced in the lung, and are degranulated as macrophages are activated in response to lung injury.

Pathway Data

▼   📤 Export

pathway_id	UniProt_severe_vs_healthy
Neutrophil degranulation	6.5757e-1
Platelet degranulation	3.6178e-2
Post-translational protein phosphorylation	9.9989e-1
Regulation of complement cascade	3.0948e-2
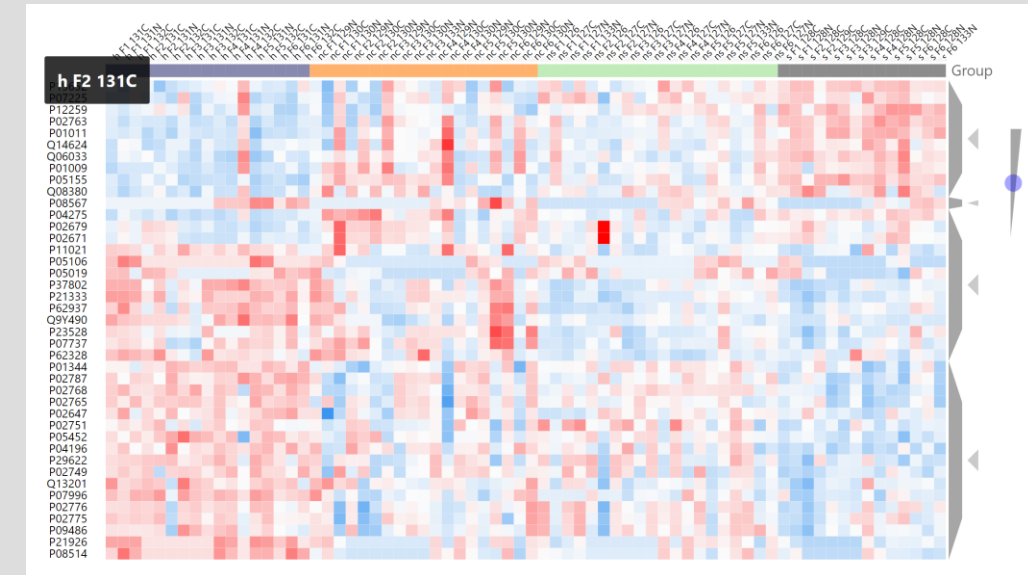
MASSIVE SUPPRESSION OF AMINO ACID METABOLISM IN THE SERA OF COVID-19 PATIENTS

- Two clusters, one upregulated in the severe cohort, and one downregulated can be seen

- The first cluster contains kynurerine and NAD+, and its upregulation is explained by the activation of kyunerine pathways in severe patients due to macrophage responses.

- The second downregulated cluster contains many amino acids such as histidine, arginine, proline, and many others. Its downregulation has been hypothesised to be due to damage to the liver from the disease.
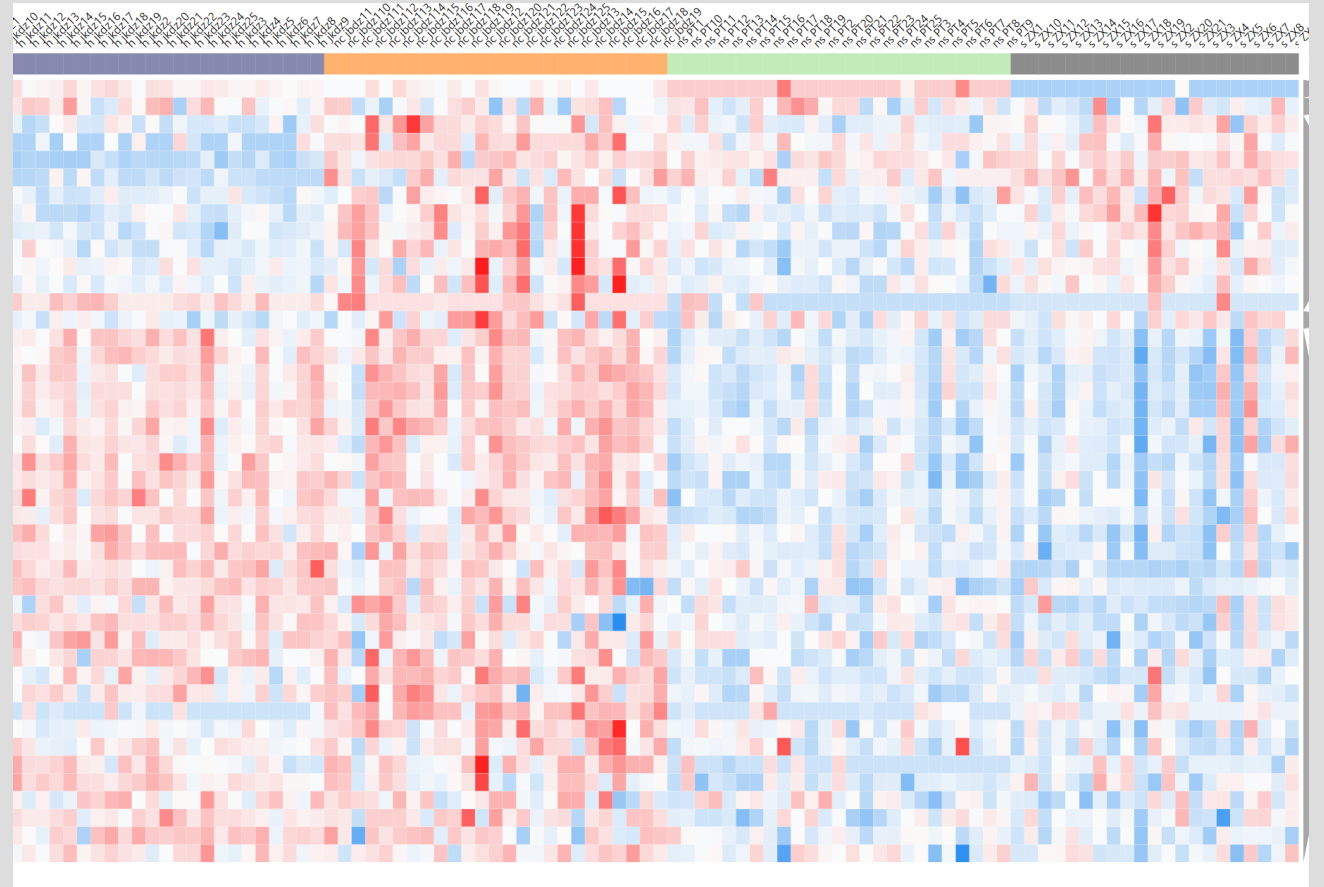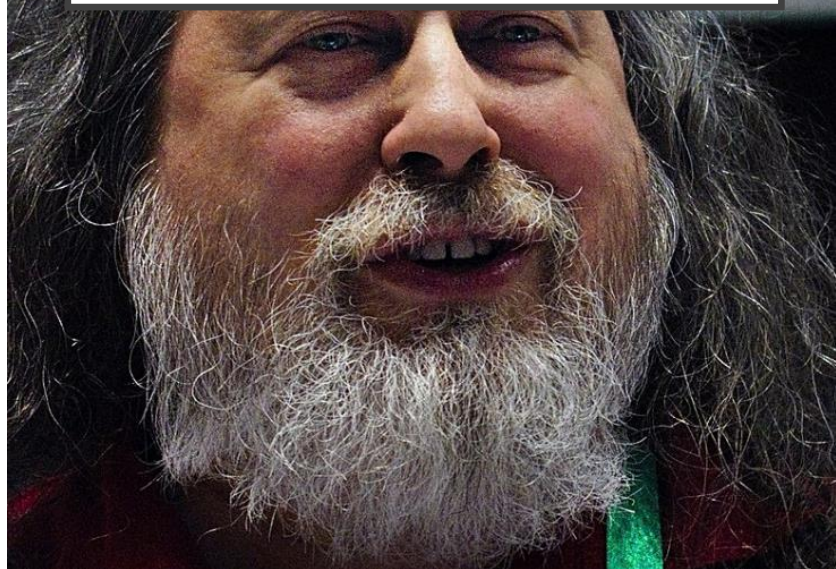
- Visit https://graphomics.glasgowcompbio.org
- GraphOmics can be used from any Web interface.
- Accompanying Python package: *pip install pyMultiOmics*

IF YOU LIKE PROGRAMMING ..

Create a mapping object

The mapping object uses Reactome to map the different biological entities in the data:

- Transcripts (or genes) are connected to the proteins they encode
- Proteins and compounds are connected to reactions they're involved in
- Reactions are connected to pathways

```
m = Mapper(DANIO_RERIO, metabolic_pathway_only=True) \
        .set_gene(gene_data, gene_design) \
        .set_protein(protein_data, protein_design) \
        .set_compound(compound_data, compound_design) \
        .build()
```

Retrieve top-10 significantly changing genes

Find the compounds that are connected to the DE genes above

```
res = QueryBuilder(ap) \
        .add(Select(GENES)) \
        .add(SignificantDE(case, control, pval, fc_lte=fc_lte, fc_gte=fc_gte, N=N)) \
        .add(Connected(data_type=COMPOUNDS)) \
        .run()
res
```

```
from pyMultiOmics.info import get_info
```

```
entity_ids = ['ENSDARG00000091254', 'F1QAA7', '15378', 'R-DRE-469659', 'R-DRE-174403']
data_types = ['genes', 'proteins', 'compounds', 'reactions', 'pathways']
for entity_id, data_type in zip(entity_ids, data_types):
    print(entity_id, data_type)
    print(get_info(entity_id, data_type))
    print()
```

DE analysis

```
ap = AnalysisPipeline(m)
```

```
method = INFERENCE_T_TEST
ap.run_de(method, GENES, 'Distal', 'Proximal')
ap.run_de(method, GENES, 'Distal', 'Middle')
ap.run_de(method, GENES, 'Proximal', 'Middle')
```