Web server
pour les analyses d’enrichissement
Gene Ontology, Réseaux…

Céline Boby – Jérémy Tournayre – Muriel Bonnet

Groupe Biopuce – 19/10/2023

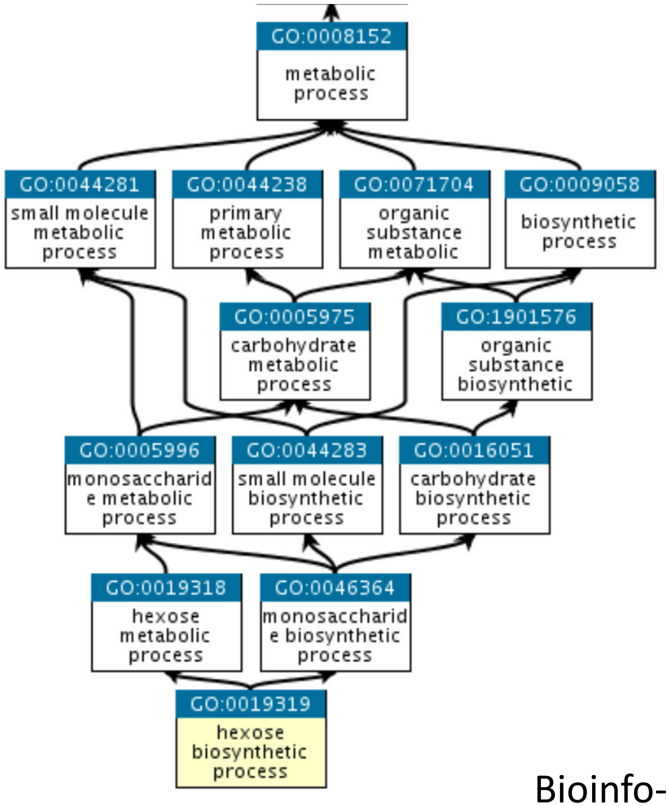
Quelques définitions : Ontologie

Ontologie :

- **ensemble structuré de termes et de concepts** d'un domaine particulier
- précise les **relations** entre ces termes et leurs **propriétés**.
- **structure hiérarchique** et l'ensemble des termes ancré par un terme de haut niveau : la **racine**.

Gene Ontology (GO) :

- Vocabulaire contrôlé et structuré qui **décrit les gènes et les protéines**
- Les protéines jouent un rôle dans **une ou plusieurs dizaines de fonctions biologiques.**
- 3 types d'information :
    - **processus biologiques (BP)**
    - **compartiments cellulaires (CC)**
    - fonctions moléculaires (MF).



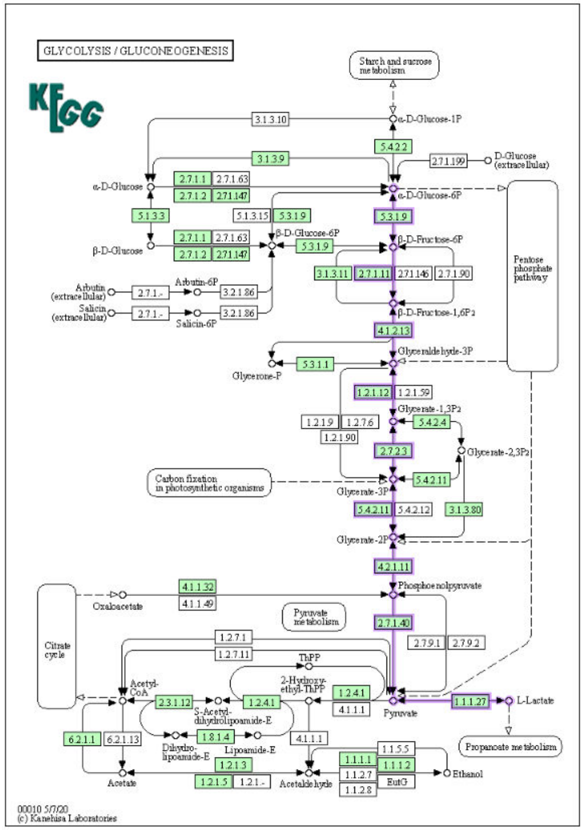
GENEONTOLOGY
Unifying Biology

http://geneontology.org/

Quelques définitions : Voie biologique/pathways

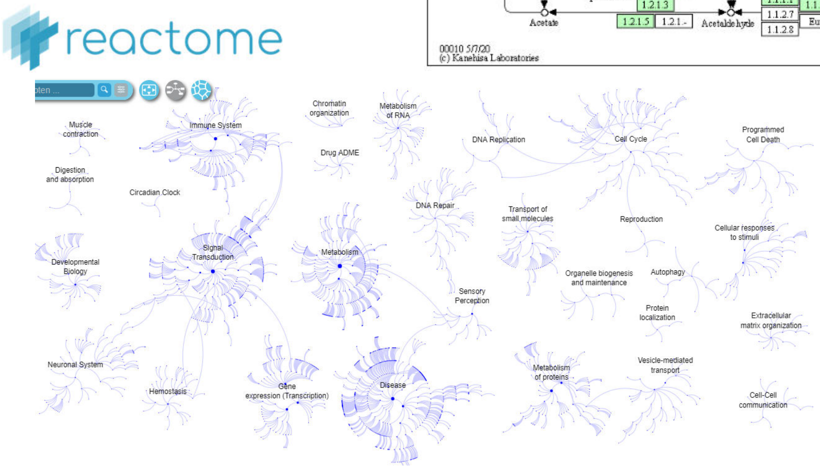
Processus biologiques :

- constitués d'un certain nombre de **réactions formant des voies** qui transforment des espèces chimiques en produits utiles.
- peuvent créer des **biomolécules complexes**, transformer **l'énergie** d'une forme à une autre ou diriger l'assemblage de **systèmes multicellulaires complexes**.
- sont souvent représentés sous la forme de **réseaux**



Différentes bases de données de pathways :

- **KEGG PATHWAY** : collection de « pathways maps » pour le **métabolisme**, les **processus cellulaires**, les maladies humaines…
- **Reactome** : Base de données de molécules de signalisation et de métabolisme
- **Panther** (Protein ANalysis THrough Evolutionary Relationships) : Base de données annotée d'arbres phylogénétiques de familles de gènes.
- **Biocarta**
- **WikiPathways :** ressource communautaire (non reviewé)…



Autres bases de données

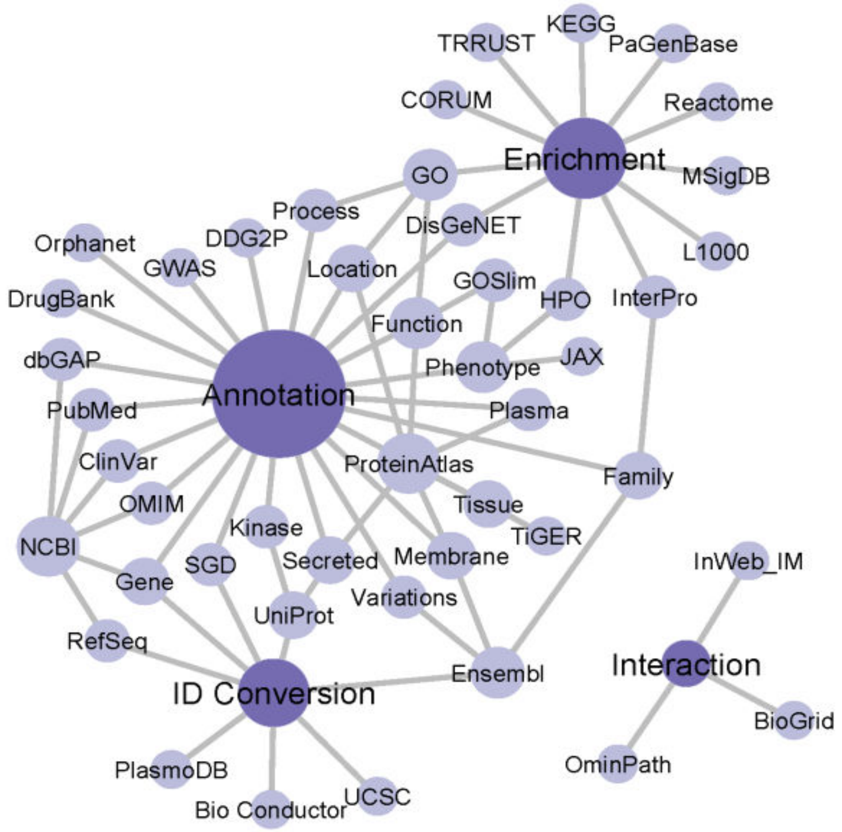
Interactions protéiques (PPI) :

- Psicquic
- String
- Uniprot
- Happi-2
- IntAct
- BioGrid…

Spécificité tissulaire : Human protein atlas (recense toutes les protéines humaines dans les cellules, les tissus et les organes)…
miRNA : miRTarBase…
Facteurs de transcriptions : Transfac…
Complexes protéiques : CORUM
Pathologies humaines : Human Phenotype Ontology
…

Bases de données Bovines :

- Bovine Genome DataBase
https://bovinegenome.elsiklab.missouri.edu/
- miRNA : **RumimiR** database
http://rumimir.sigenae.org/



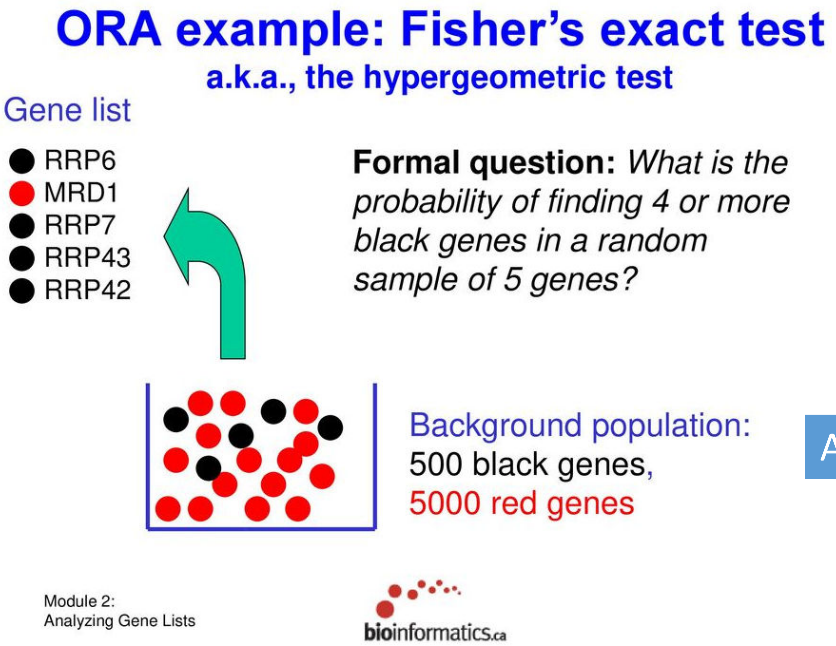
Supplementary Figure 8. Over 40 data sources are integrated in Metascape to support its CAME analysis workflow (also see Supplementary Data 2).

Zhou Y *et al.*, 2019

Analyses d'Enrichissement
Over-Representation Analysis (ORA)

Met en évidence les **fonctions biologiques (GO/pathways) les plus représentatives** d'un jeu de données (liste gènes/prot.)

La **p-value associée** à un terme représente la **probabilité d'obtenir ce terme** dans les résultats d'annotation si on effectuait un **tirage au hasard** → caractérise le caractère **significatif** ou non de la surreprésentation de ce terme.
Plus la p-value est **faible**, plus le terme GO est dit « **enrichi** » et plus la présence de ce terme dans la liste de gènes étudiés est significative.



ORA example: Fisher's exact test

a.k.a., the hypergeometric test

Gene list

- ● RRP6
- 🔴 MRD1
- ● RRP7
- ● RRP43
- ● RRP42

Formal question: *What is the probability of finding 4 or more black genes in a random sample of 5 genes?*

Background population:
500 black genes,
5000 red genes

Attendu vs observé ?

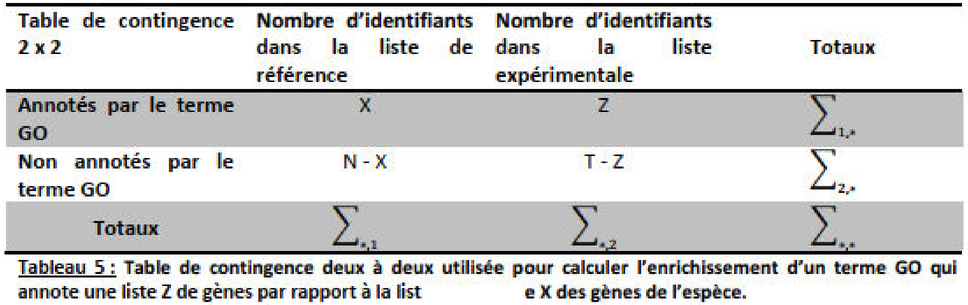


Table de contingence 2 x 2	Nombre d'identifiants dans la liste de référence	Nombre d'identifiants dans la liste expérimentale	Totaux
Annotés par le terme GO	X	Z	$\sum_{1,*}$
Non annotés par le terme GO	N - X	T - Z	$\sum_{2,*}$
Totaux	$\sum_{*,1}$	$\sum_{*,2}$	$\sum_{*,*}$

Tableau 5 : Table de contingence deux à deux utilisée pour calculer l'enrichissement d'un terme GO qui annote une liste Z de gènes par rapport à la list e X des gènes de l'espèce.

Kasspric, 2016

Etude bibliographique :

42 outils web d’enrichissement GO répertoriés et testés

Critères les plus importants :

- Web servers régulièrement maintenus
- Bases de données mises à jour
- Représentations graphiques informatives et fiables

→ **7 outils retenus** :

ProteINSIDE

g:Profiler

reactome

ShinyGO

WEADE Workflow for Enrichment Analysis and Data Exploration

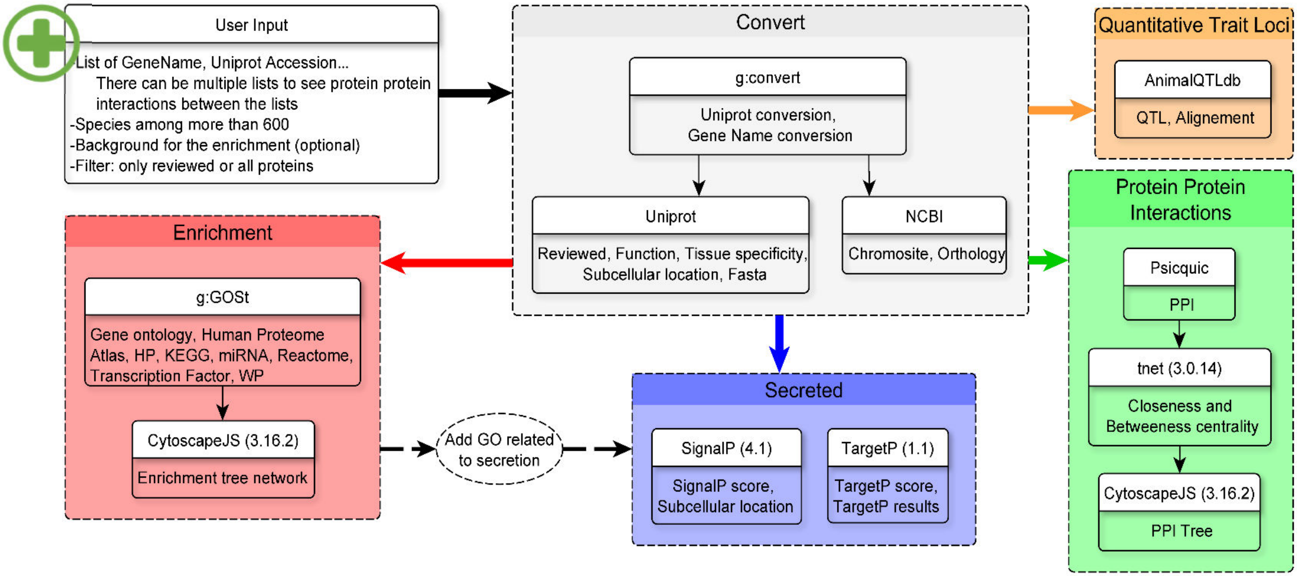
DAVID BIOINFORMATIC DATABASE

Metascape A Gene Annotation & Analysis Resource

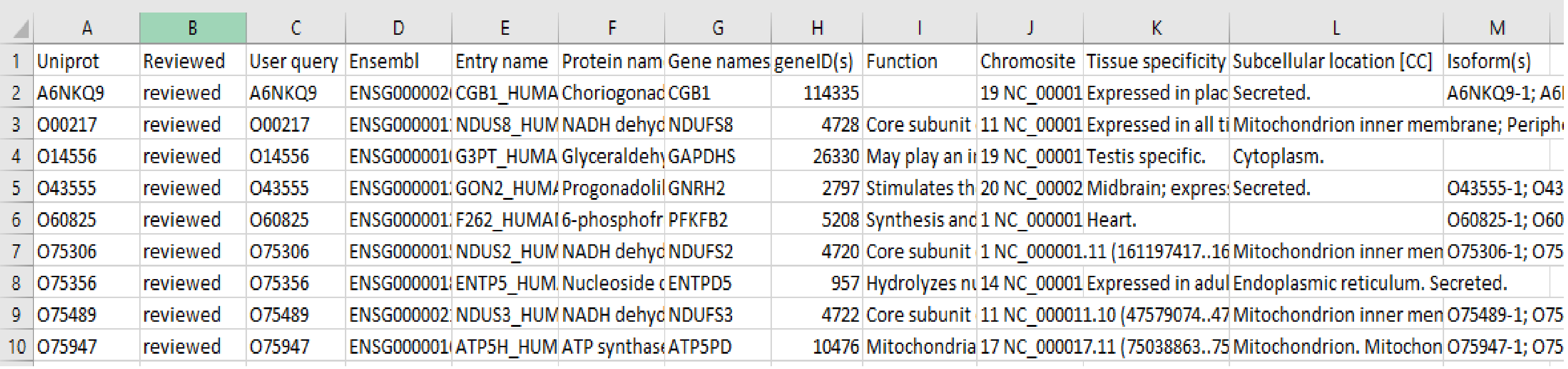
ProteINSIDE

https://umrh-bioinfo.clermont.inrae.fr/ProteINSIDE_2/

- Publié en 2015 – nouvelle version en 2021
- Outils interrogés en temps réels
- 600 espèces
- Grandes listes (> 3000 ID)



Convert :

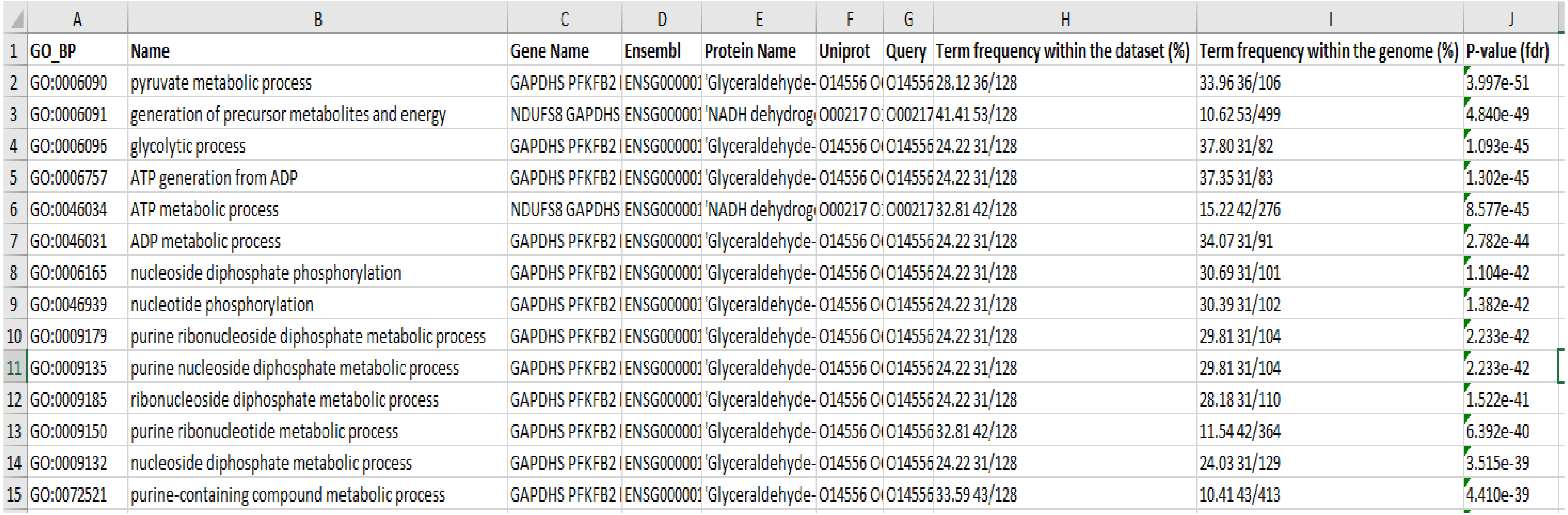


	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Uniprot	Reviewed	User query	Ensembl	Entry name	Protein nam	Gene names	geneID(s)	Function	Chromosite	Tissue specificity	Subcellular location [CC]	Isoform(s)
2	A6NKQ9	reviewed	A6NKQ9	ENSG000002(	CGB1_HUMA	Choriogonad	CGB1	114335		19 NC_00001	Expressed in plac	Secreted.	A6NKQ9-1; A6I
3	O00217	reviewed	O00217	ENSG000001:	NDUS8_HUM	NADH dehyd	NDUFS8	4728	Core subunit	11 NC_00001	Expressed in all ti	Mitochondrion inner membrane; Periph	
4	O14556	reviewed	O14556	ENSG000001(	G3PT_HUMA	Glyceraldehy	GAPDHS	26330	May play an ii	19 NC_00001	Testis specific.	Cytoplasm.	
5	O43555	reviewed	O43555	ENSG000001:	GON2_HUM/	Progonadolil	GNRH2	2797	Stimulates th	20 NC_00002	Midbrain; expres:	Secreted.	O43555-1; O43
6	O60825	reviewed	O60825	ENSG000001:	F262_HUMAI	6-phosphofr	PFKFB2	5208	Synthesis and	1 NC_000001	Heart.		O60825-1; O60
7	O75306	reviewed	O75306	ENSG000001:	NDUS2_HUM	NADH dehyd	NDUFS2	4720	Core subunit	1 NC_000001.11 (161197417..16		Mitochondrion inner men	O75306-1; O75
8	O75356	reviewed	O75356	ENSG000001:	ENTP5_HUM.	Nucleoside (	ENTPD5	957	Hydrolyzes nu	14 NC_00001	Expressed in adul	Endoplasmic reticulum. Secreted.	
9	O75489	reviewed	O75489	ENSG000002:	NDUS3_HUM	NADH dehyd	NDUFS3	4722	Core subunit	11 NC_000011.10 (47579074..47		Mitochondrion inner men	O75489-1; O75
10	O75947	reviewed	O75947	ENSG000001(	ATP5H_HUM	ATP synthase	ATP5PD	10476	Mitochondria	17 NC_000017.11 (75038863..75		Mitochondrion. Mitochon	O75947-1; O75

ProteInside v2 - suite

Enrichment :

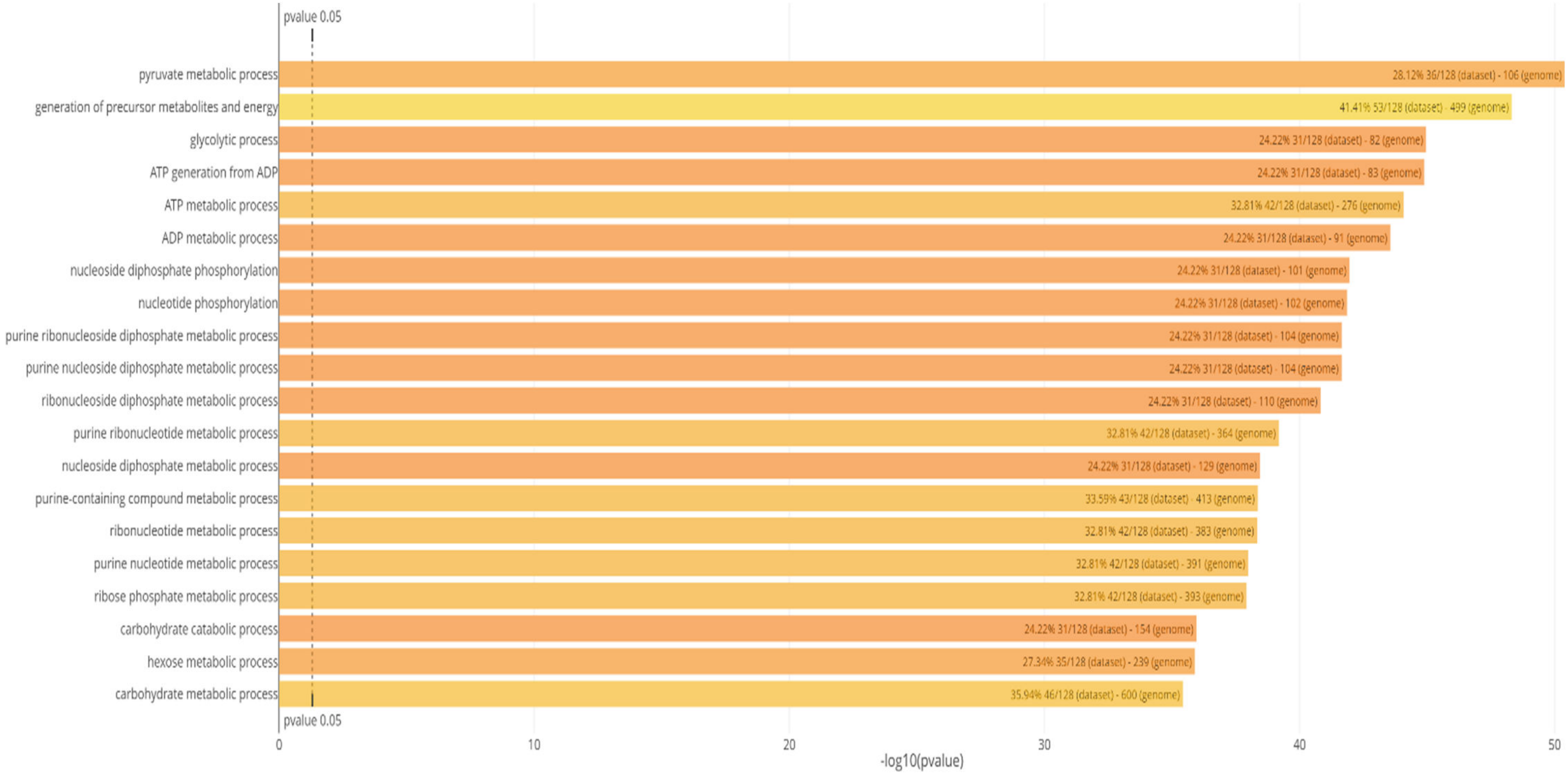
A- table



GO_BP	Name	Gene Name	Ensembl	Protein Name	Uniprot	Query	Term frequency within the dataset (%)	Term frequency within the genome (%)	P-value (fdr)
GO:0006090	pyruvate metabolic process	GAPDHS PFKFB2	ENSG000001	'Glyceraldehyde-	O14556	O14556	28.12 36/128	33.96 36/106	3.997e-51
GO:0006091	generation of precursor metabolites and energy	NDUFS8 GAPDHS	ENSG000001	'NADH dehydrog	O00217	O00217	41.41 53/128	10.62 53/499	4.840e-49
GO:0006096	glycolytic process	GAPDHS PFKFB2	ENSG000001	'Glyceraldehyde-	O14556	O14556	24.22 31/128	37.80 31/82	1.093e-45
GO:0006757	ATP generation from ADP	GAPDHS PFKFB2	ENSG000001	'Glyceraldehyde-	O14556	O14556	24.22 31/128	37.35 31/83	1.302e-45
GO:0046034	ATP metabolic process	NDUFS8 GAPDHS	ENSG000001	'NADH dehydrog	O00217	O00217	32.81 42/128	15.22 42/276	8.577e-45
GO:0046031	ADP metabolic process	GAPDHS PFKFB2	ENSG000001	'Glyceraldehyde-	O14556	O14556	24.22 31/128	34.07 31/91	2.782e-44
GO:0006165	nucleoside diphosphate phosphorylation	GAPDHS PFKFB2	ENSG000001	'Glyceraldehyde-	O14556	O14556	24.22 31/128	30.69 31/101	1.104e-42
GO:0046939	nucleotide phosphorylation	GAPDHS PFKFB2	ENSG000001	'Glyceraldehyde-	O14556	O14556	24.22 31/128	30.39 31/102	1.382e-42
GO:0009179	purine ribonucleoside diphosphate metabolic process	GAPDHS PFKFB2	ENSG000001	'Glyceraldehyde-	O14556	O14556	24.22 31/128	29.81 31/104	2.233e-42
GO:0009135	purine nucleoside diphosphate metabolic process	GAPDHS PFKFB2	ENSG000001	'Glyceraldehyde-	O14556	O14556	24.22 31/128	29.81 31/104	2.233e-42
GO:0009185	ribonucleoside diphosphate metabolic process	GAPDHS PFKFB2	ENSG000001	'Glyceraldehyde-	O14556	O14556	24.22 31/128	28.18 31/110	1.522e-41
GO:0009150	purine ribonucleotide metabolic process	GAPDHS PFKFB2	ENSG000001	'Glyceraldehyde-	O14556	O14556	32.81 42/128	11.54 42/364	6.392e-40
GO:0009132	nucleoside diphosphate metabolic process	GAPDHS PFKFB2	ENSG000001	'Glyceraldehyde-	O14556	O14556	24.22 31/128	24.03 31/129	3.515e-39
GO:0072521	purine-containing compound metabolic process	GAPDHS PFKFB2	ENSG000001	'Glyceraldehyde-	O14556	O14556	33.59 43/128	10.41 43/413	4.410e-39

ProteInside v2 - suite

Enrichment :
B- barplot

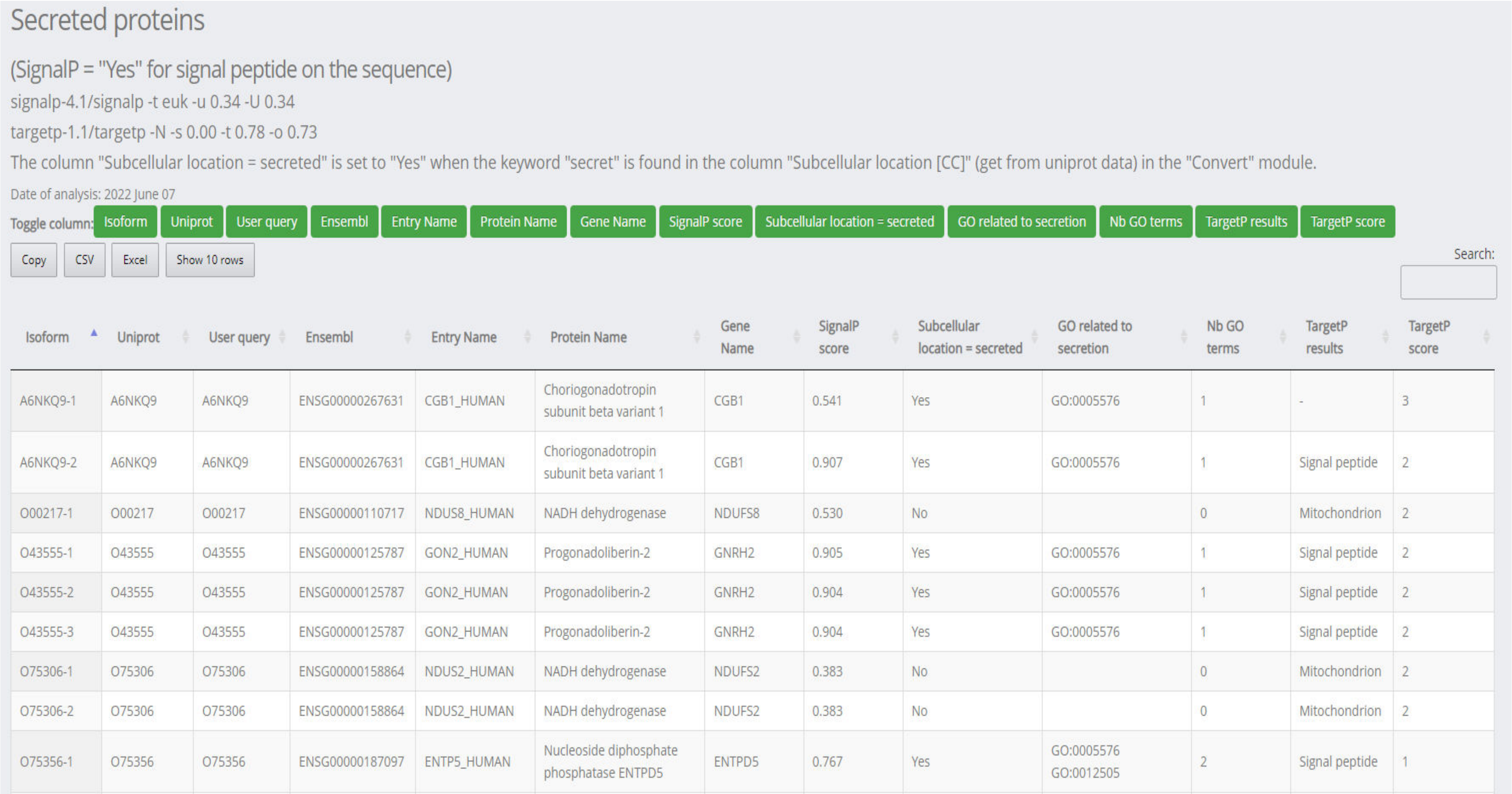


ProteInside v2 - suite

Enrichment :
C- Enrichment Tree Network

ProteInside v2 - suite

Prédiction des protéines secrétées : secretome



Secreted proteins

(SignalP = "Yes" for signal peptide on the sequence)

```
signalp-4.1/signalp -t euk -u 0.34 -U 0.34
targetp-1.1/targetp -N -s 0.00 -t 0.78 -o 0.73
```

The column "Subcellular location = secreted" is set to "Yes" when the keyword "secret" is found in the column "Subcellular location [CC]" (get from uniprot data) in the "Convert" module.

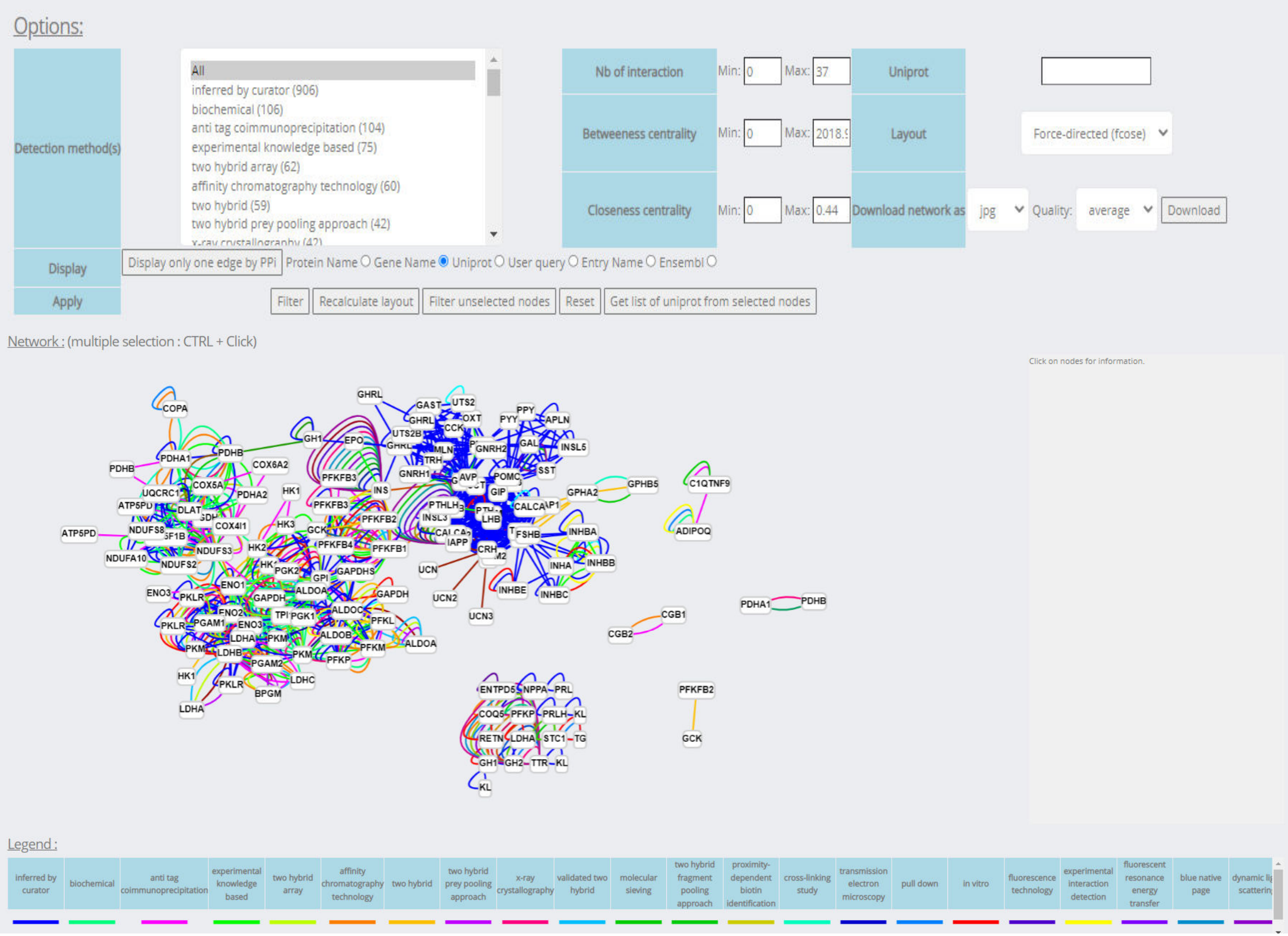
Date of analysis: 2022 June 07

Isoform	Uniprot	User query	Ensembl	Entry Name	Protein Name	Gene Name	SignalP score	Subcellular location = secreted	GO related to secretion	Nb GO terms	TargetP results	TargetP score
A6NKQ9-1	A6NKQ9	A6NKQ9	ENSG00000267631	CGB1_HUMAN	Choriogonadotropin subunit beta variant 1	CGB1	0.541	Yes	GO:0005576	1	-	3
A6NKQ9-2	A6NKQ9	A6NKQ9	ENSG00000267631	CGB1_HUMAN	Choriogonadotropin subunit beta variant 1	CGB1	0.907	Yes	GO:0005576	1	Signal peptide	2
O00217-1	O00217	O00217	ENSG00000110717	NDUS8_HUMAN	NADH dehydrogenase	NDUFS8	0.530	No		0	Mitochondrion	2
O43555-1	O43555	O43555	ENSG00000125787	GON2_HUMAN	Progonadoliberin-2	GNRH2	0.905	Yes	GO:0005576	1	Signal peptide	2
O43555-2	O43555	O43555	ENSG00000125787	GON2_HUMAN	Progonadoliberin-2	GNRH2	0.904	Yes	GO:0005576	1	Signal peptide	2
O43555-3	O43555	O43555	ENSG00000125787	GON2_HUMAN	Progonadoliberin-2	GNRH2	0.904	Yes	GO:0005576	1	Signal peptide	2
O75306-1	O75306	O75306	ENSG00000158864	NDUS2_HUMAN	NADH dehydrogenase	NDUFS2	0.383	No		0	Mitochondrion	2
O75306-2	O75306	O75306	ENSG00000158864	NDUS2_HUMAN	NADH dehydrogenase	NDUFS2	0.383	No		0	Mitochondrion	2
O75356-1	O75356	O75356	ENSG00000187097	ENTP5_HUMAN	Nucleoside diphosphate phosphatase ENTPD5	ENTPD5	0.767	Yes	GO:0005576 GO:0012505	2	Signal peptide	1

ProteInside v2 - suite

Proteins-proteins interactions :
PPi Tree

Interactions protéiques
entre 2 listes



Options:

Detection method(s)

All
inferred by curator (906)
biochemical (106)
anti tag coimmunoprecipitation (104)
experimental knowledge based (75)
two hybrid array (62)
affinity chromatography technology (60)
two hybrid (59)
two hybrid prey pooling approach (42)
x-ray crystallography (42)

Nb of interaction Min: 0 Max: 37
Betweeness centrality Min: 0 Max: 2018.9
Closeness centrality Min: 0 Max: 0.44
Uniprot
Layout: Force-directed (fcose)
Download network as: jpg Quality: average

Display: Protein Name ○ Gene Name ● Uniprot ○ User query ○ Entry Name ○ Ensembl ○

Network : (multiple selection : CTRL + Click)

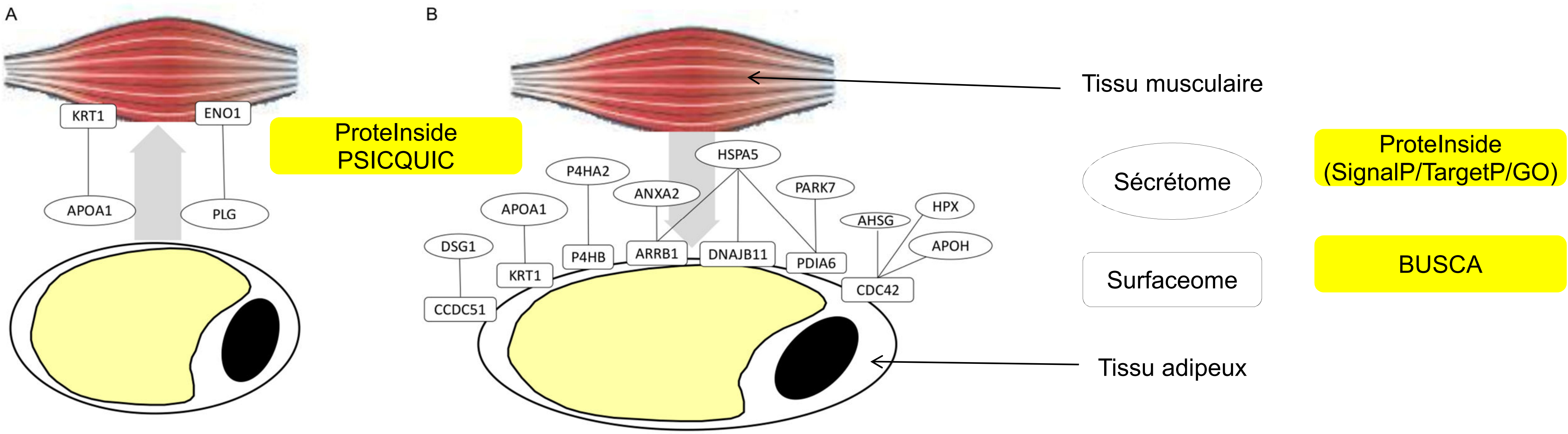
Click on nodes for information.

Legend :

ProteInside v2 - suite

Prochainement : **prédiction des protéines membranaires** (**surfaceome**)

➔ **Interaction entre le surfaceome** d'un tissu et le **secretome** d'un autre tissu pour étudier le **dialogue inter-organe**.



Réseau de protéines en interaction entre (A) les protéines potentiellement sécrétées par les cellules adipeuses et les protéines de la surface cellulaire musculaire ou (B) entre les protéines musculaires et de la surface cellulaire adipeuse (Bonnet et al. 2020).

g:Profiler

https://biit.cs.ut.ee/gprofiler/gost

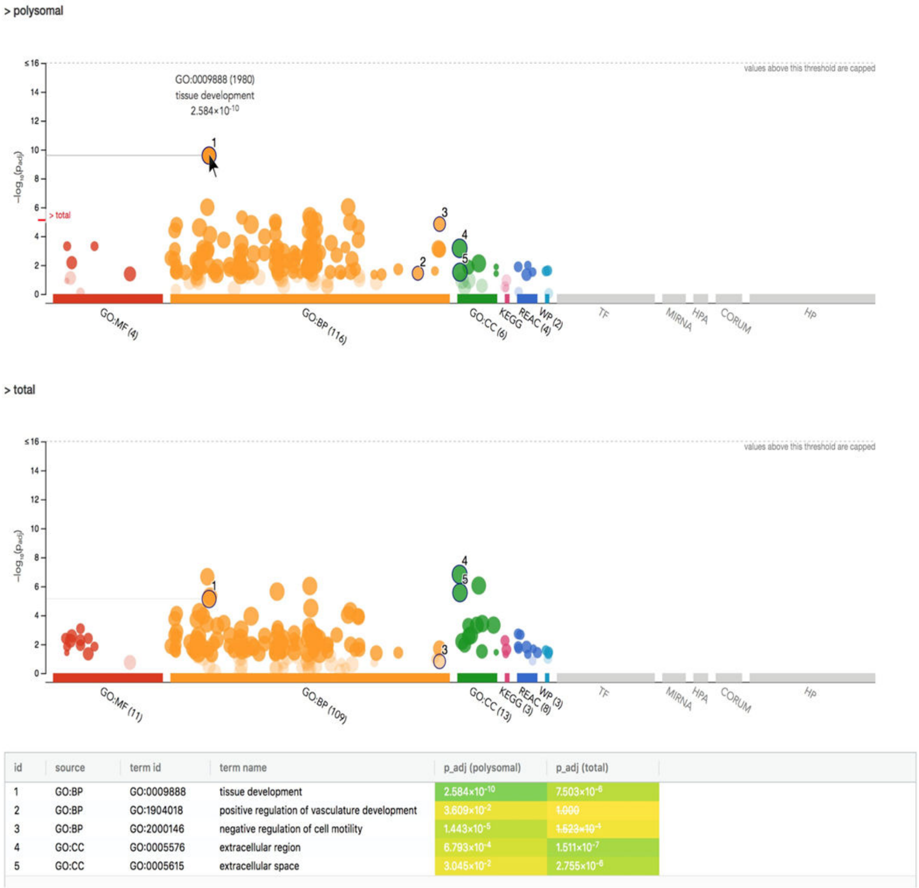
g:GOSt : **enrichissement** et analyse de **sous-représentation**

Nombreuses bases de données interrogées : GO, pathways…

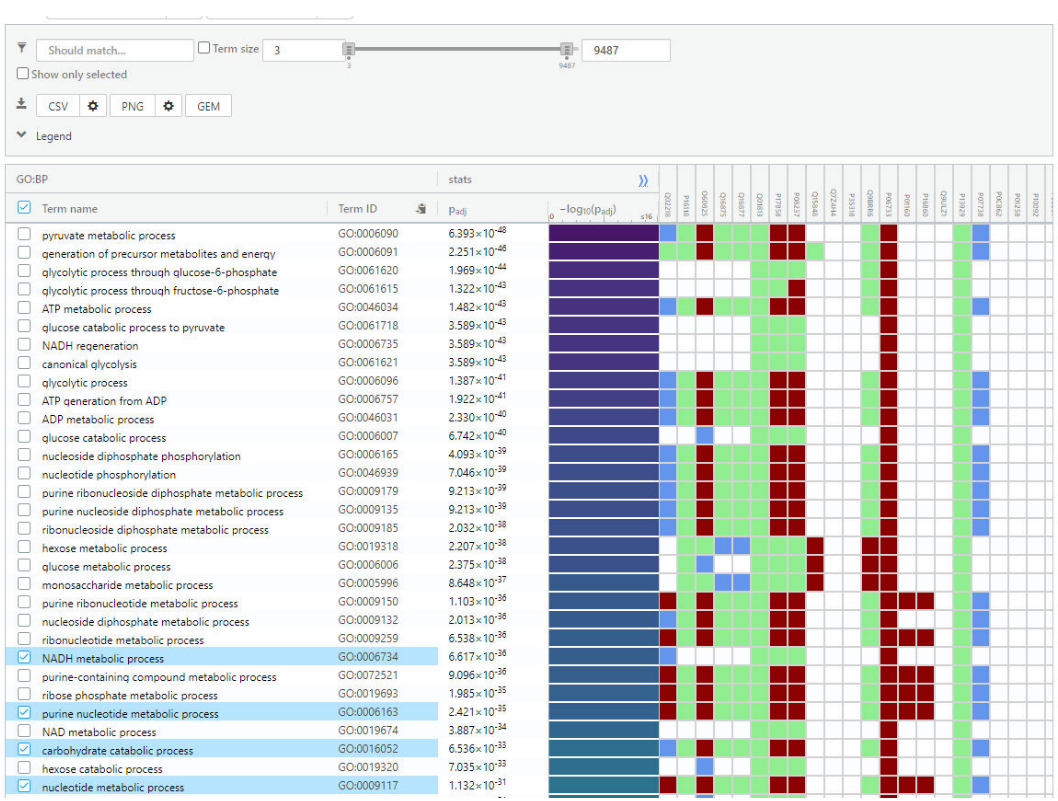
Comparaison **multi-liste**

Analyse de **données classées** (rankées)

- Publié en 2007 et évolue régulièrement (highly cited)
- Grandes listes (> 3000 ID)
- Nombreuses espèces



id	source	term id	term name	p_adj (polysomal)	p_adj (total)
1	GO:BP	GO:0009888	tissue development	2.584×10^{-10}	7.503×10^{-6}
2	GO:BP	GO:1904018	positive regulation of vasculature development	3.609×10^{-2}	~~1.000~~
3	GO:BP	GO:2000146	negative regulation of cell motility	1.443×10^{-5}	~~1.523×10^{-4}~~
4	GO:CC	GO:0005576	extracellular region	6.793×10^{-4}	1.511×10^{-7}
5	GO:CC	GO:0005615	extracellular space	3.045×10^{-2}	2.755×10^{-6}



GO:BP

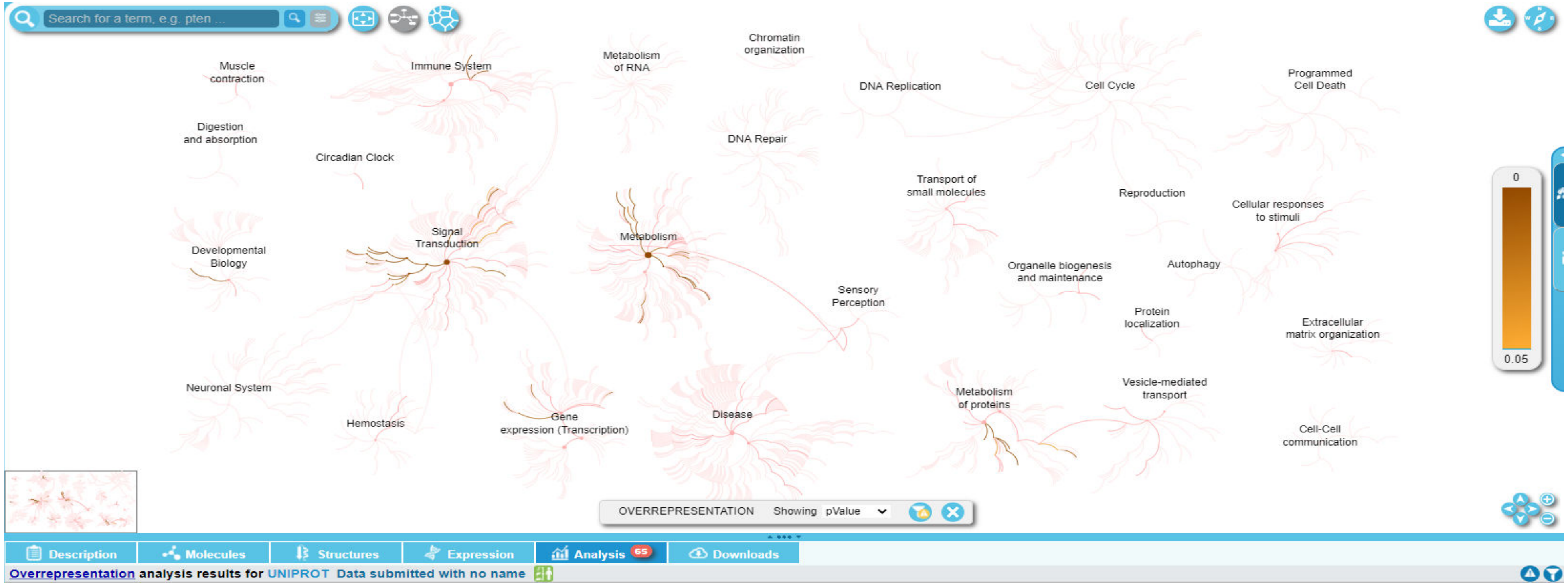
☑	Term name	Term ID	Padj
☐	pyruvate metabolic process	GO:0006090	6.393×10^{-48}
☐	generation of precursor metabolites and energy	GO:0006091	2.251×10^{-46}
☐	glycolytic process through glucose-6-phosphate	GO:0061620	1.969×10^{-44}
☐	glycolytic process through fructose-6-phosphate	GO:0061615	1.322×10^{-43}
☐	ATP metabolic process	GO:0046034	1.482×10^{-43}
☐	glucose catabolic process to pyruvate	GO:0061718	3.589×10^{-43}
☐	NADH regeneration	GO:0006735	3.589×10^{-43}
☐	canonical glycolysis	GO:0061621	3.589×10^{-43}
☐	glycolytic process	GO:0006096	1.387×10^{-41}
☐	ATP generation from ADP	GO:0006757	1.922×10^{-41}
☐	ADP metabolic process	GO:0046031	2.330×10^{-40}
☐	glucose catabolic process	GO:0006007	6.742×10^{-40}
☐	nucleoside diphosphate phosphorylation	GO:0006165	4.093×10^{-39}
☐	nucleotide phosphorylation	GO:0046939	7.046×10^{-39}
☐	purine ribonucleoside diphosphate metabolic process	GO:0009179	9.213×10^{-39}
☐	purine nucleoside diphosphate metabolic process	GO:0009135	9.213×10^{-39}
☐	ribonucleoside diphosphate metabolic process	GO:0009185	2.032×10^{-38}
☐	hexose metabolic process	GO:0019318	2.207×10^{-38}
☐	glucose metabolic process	GO:0006006	2.375×10^{-38}
☐	monosaccharide metabolic process	GO:0005996	8.648×10^{-37}
☐	purine ribonucleotide metabolic process	GO:0009150	1.103×10^{-36}
☐	nucleoside diphosphate metabolic process	GO:0009132	2.013×10^{-36}
☐	ribonucleotide metabolic process	GO:0009259	6.538×10^{-36}
☑	NADH metabolic process	GO:0006734	6.617×10^{-36}
☐	purine-containing compound metabolic process	GO:0072521	9.096×10^{-36}
☐	ribose phosphate metabolic process	GO:0019693	1.985×10^{-35}
☑	purine nucleotide metabolic process	GO:0006163	2.421×10^{-35}
☐	NAD metabolic process	GO:0019674	3.887×10^{-34}
☑	carbohydrate catabolic process	GO:0016052	6.536×10^{-33}
☐	hexose catabolic process	GO:0019320	7.035×10^{-33}
☑	nucleotide metabolic process	GO:0009117	1.132×10^{-31}

reactome

https://reactome.org/

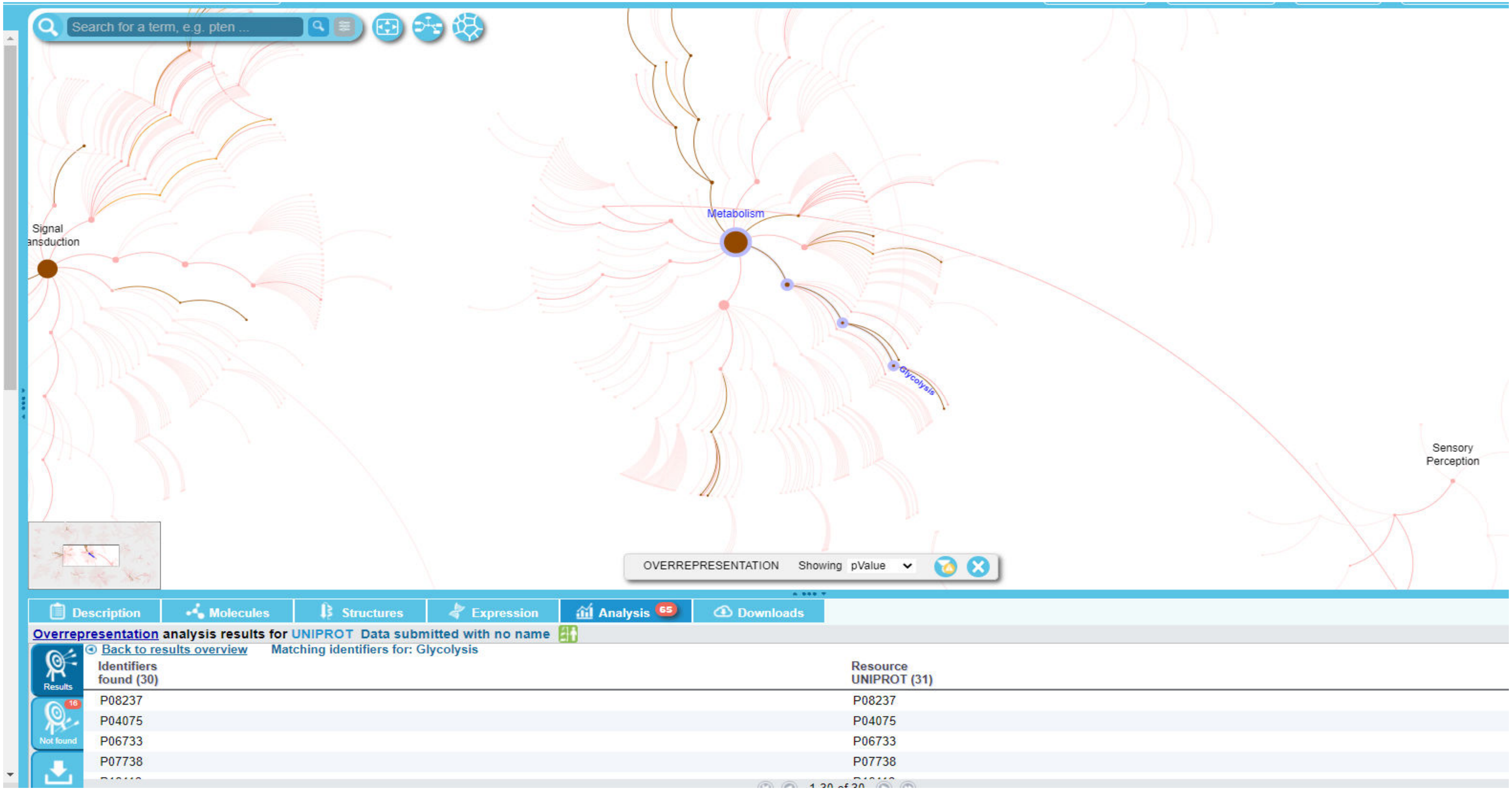
- Publié en 2004 et évolue régulièrement (14 publi highly cited)
- Grandes listes (> 3000 ID)

Base de données de réseaux et de réactions en **biologie humaine**
(tout événement qui modifie l'état d'une molécule biologique : liaison, activation, translocation, dégradation...)
+ stratégie d'**inférence par orthologie** pour 15 espèces eucaryotes dont **bovin**
Interroge d’autres bases de données dont **KEGG** et **GO**
Possibilité d'analyses **time-lapse**
Guide pour **analyses multi-omiques**



Overrepresentation analysis results for UNIPROT Data submitted with no name

Reactome - suite

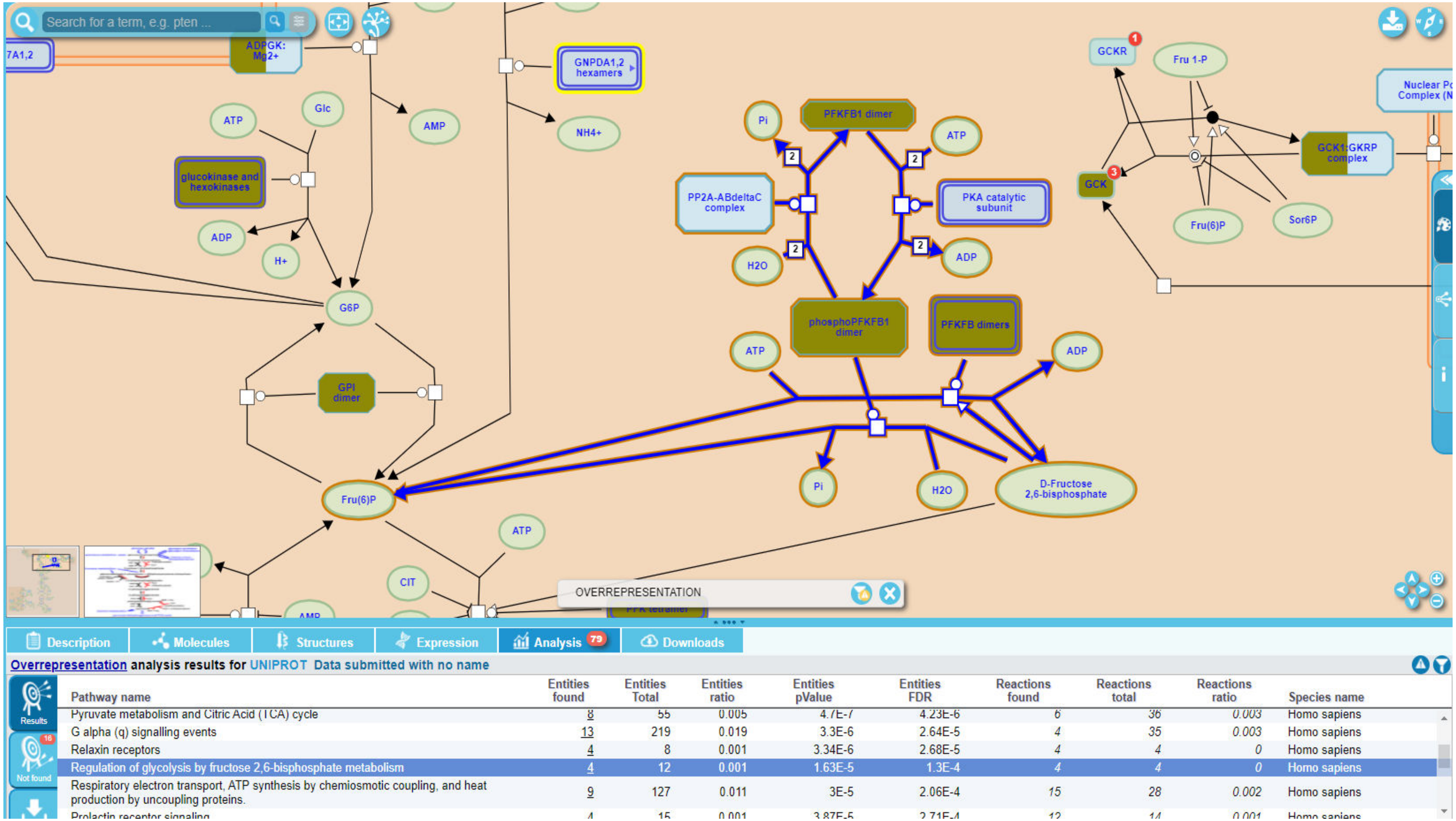


Overrepresentation analysis results for UNIPROT Data submitted with no name

Back to results overview Matching identifiers for: Glycolysis

Identifiers found (30)	Resource UNIPROT (31)
P08237	P08237
P04075	P04075
P06733	P06733
P07738	P07738

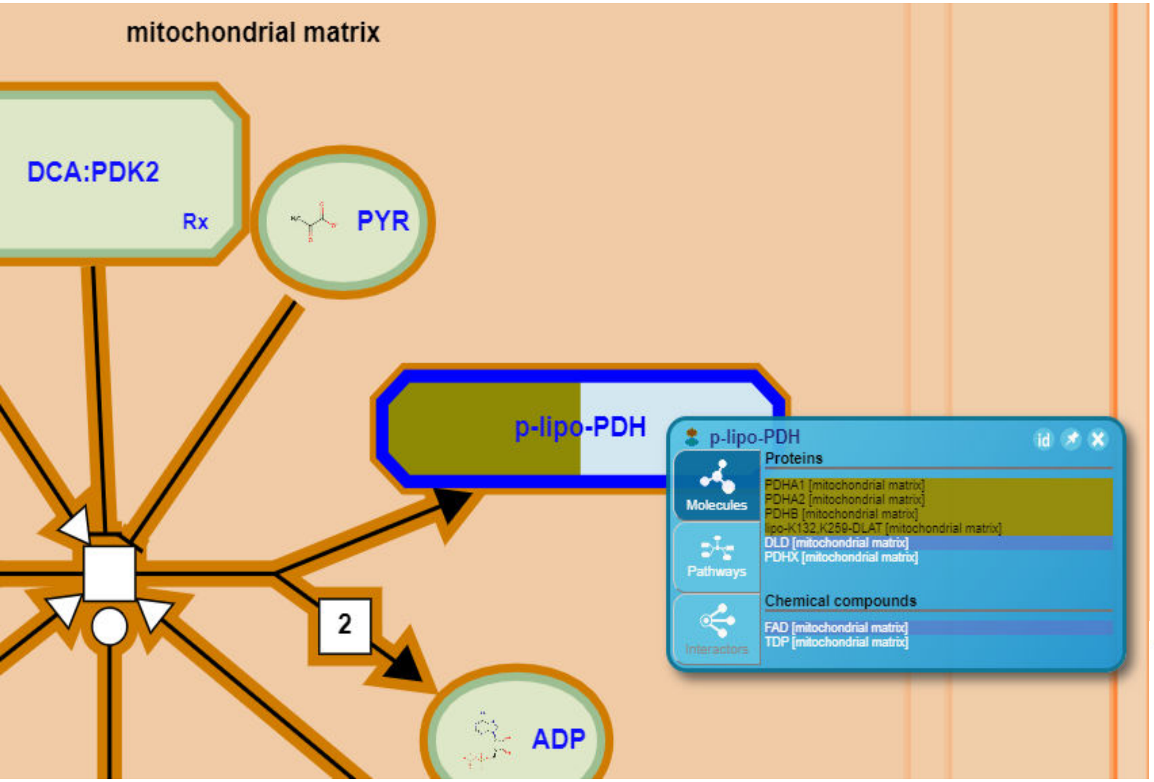
Reactome - suite



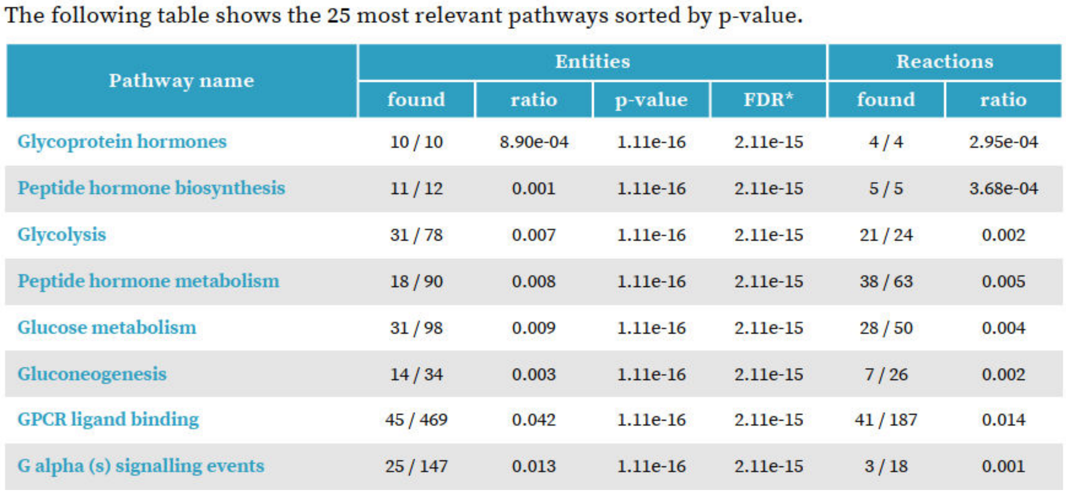
Overrepresentation analysis results for UNIPROT Data submitted with no name

Pathway name	Entities found	Entities Total	Entities ratio	Entities pValue	Entities FDR	Reactions found	Reactions total	Reactions ratio	Species name
Pyruvate metabolism and Citric Acid (TCA) cycle	8	55	0.005	4.7E-7	4.23E-6	6	36	0.003	Homo sapiens
G alpha (q) signalling events	13	219	0.019	3.3E-6	2.64E-5	4	35	0.003	Homo sapiens
Relaxin receptors	4	8	0.001	3.34E-6	2.68E-5	4	4	0	Homo sapiens
Regulation of glycolysis by fructose 2,6-bisphosphate metabolism	4	12	0.001	1.63E-5	1.3E-4	4	4	0	Homo sapiens
Respiratory electron transport, ATP synthesis by chemiosmotic coupling, and heat production by uncoupling proteins.	9	127	0.011	3E-5	2.06E-4	15	28	0.002	Homo sapiens
Prolactin receptor signaling	4	15	0.001	3.87E-5	2.71E-4	12	14	0.001	Homo sapiens

Reactome - suite



PDF report :

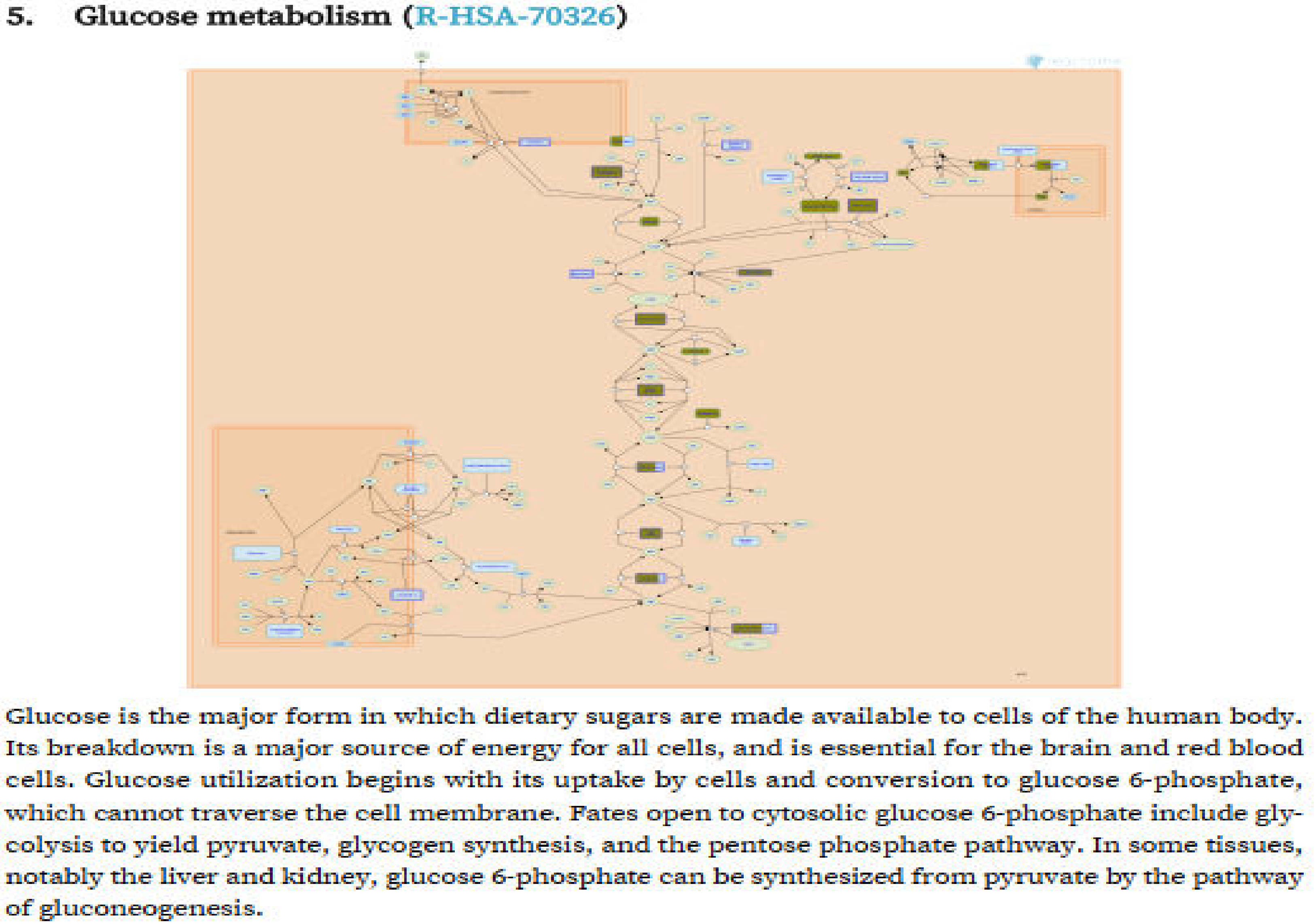


The following table shows the 25 most relevant pathways sorted by p-value.

Pathway name	Entities				Reactions	
	found	ratio	p-value	FDR*	found	ratio
Glycoprotein hormones	10 / 10	8.90e-04	1.11e-16	2.11e-15	4 / 4	2.95e-04
Peptide hormone biosynthesis	11 / 12	0.001	1.11e-16	2.11e-15	5 / 5	3.68e-04
Glycolysis	31 / 78	0.007	1.11e-16	2.11e-15	21 / 24	0.002
Peptide hormone metabolism	18 / 90	0.008	1.11e-16	2.11e-15	38 / 63	0.005
Glucose metabolism	31 / 98	0.009	1.11e-16	2.11e-15	28 / 50	0.004
Gluconeogenesis	14 / 34	0.003	1.11e-16	2.11e-15	7 / 26	0.002
GPCR ligand binding	45 / 469	0.042	1.11e-16	2.11e-15	41 / 187	0.014
G alpha (s) signalling events	25 / 147	0.013	1.11e-16	2.11e-15	3 / 18	0.001

Reactome - suite

PDF report :



5. **Glucose metabolism** (R-HSA-70326)

Glucose is the major form in which dietary sugars are made available to cells of the human body. Its breakdown is a major source of energy for all cells, and is essential for the brain and red blood cells. Glucose utilization begins with its uptake by cells and conversion to glucose 6-phosphate, which cannot traverse the cell membrane. Fates open to cytosolic glucose 6-phosphate include glycolysis to yield pyruvate, glycogen synthesis, and the pentose phosphate pathway. In some tissues, notably the liver and kidney, glucose 6-phosphate can be synthesized from pyruvate by the pathway of gluconeogenesis.

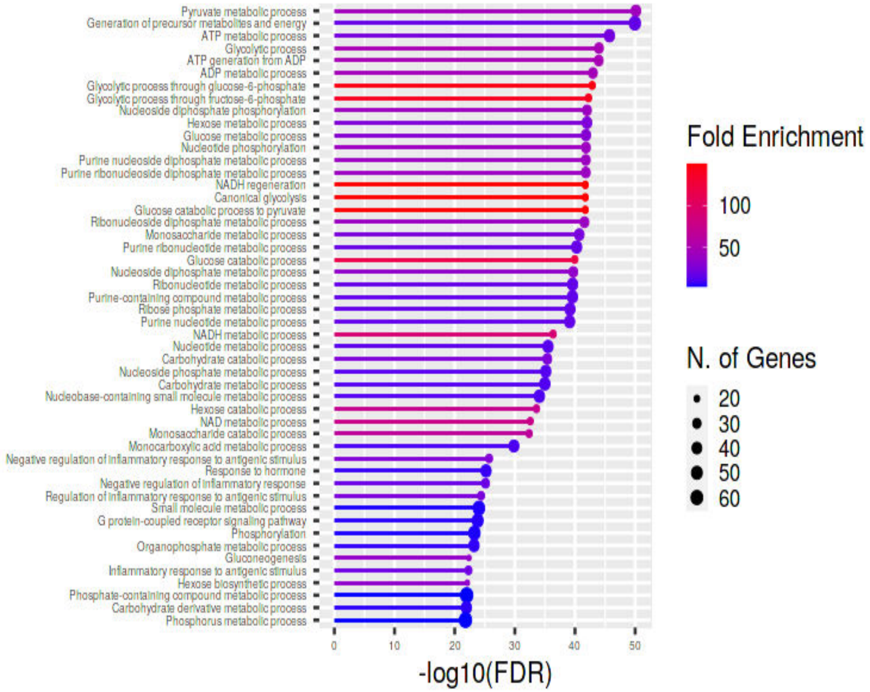
http://bioinformatics.sdstate.edu/go/

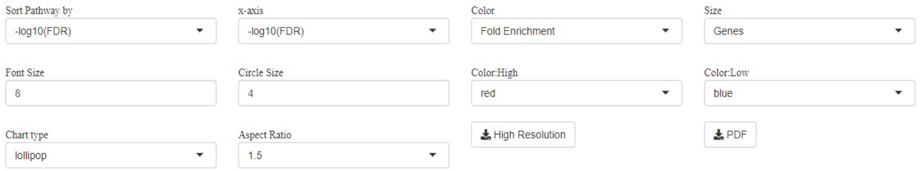
ShinyGO

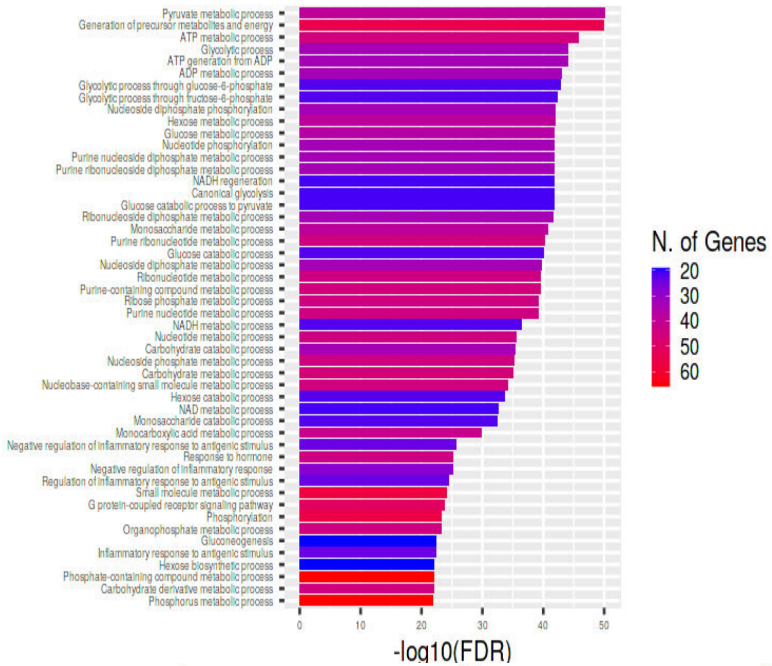
- Publié en 2020 et évolue régulièrement (highly cited)
- Grandes listes (> 3000 ID)
- Nombreuses espèces 🐄 🐐 🐑

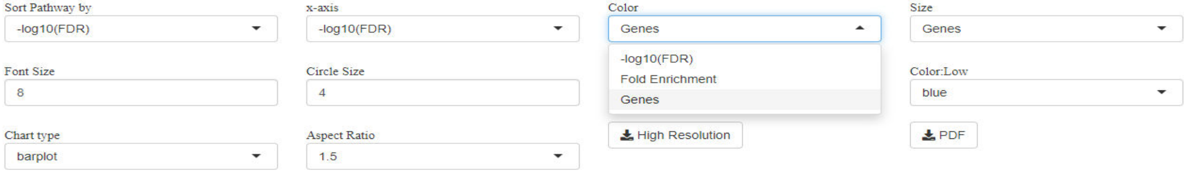
Appli web graphique utilisant package R Shiny : visualisation graphique des résultats d'enrichissements (GO, KEGG, Panther, Reactome, Facteurs de Transcription, miRNAs, CORUM)

BarChart facilement configurables : GO BP :




Sort Pathway by: -log10(FDR)
x-axis: -log10(FDR)
Color: Fold Enrichment
Size: Genes
Font Size: 8
Circle Size: 4
Color:High: red
Color:Low: blue
Chart type: lollipop
Aspect Ratio: 1.5



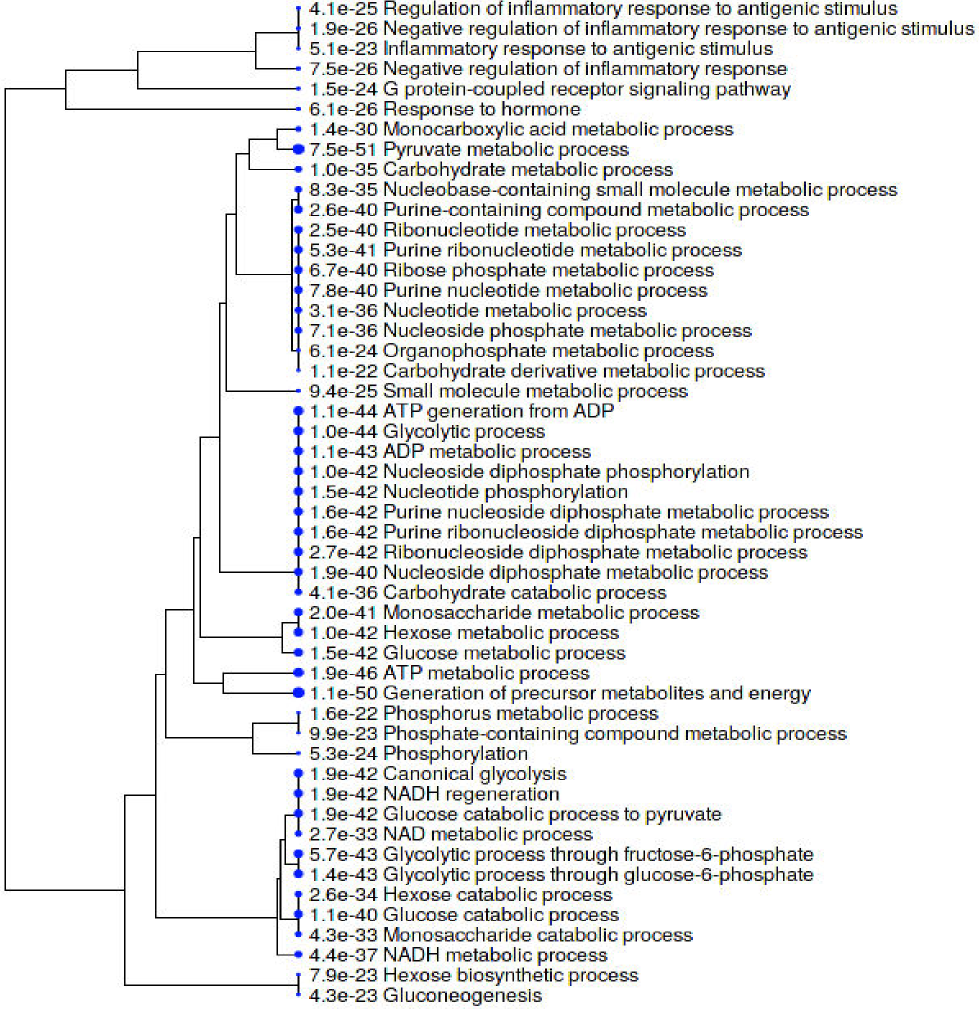

Sort Pathway by: -log10(FDR)
x-axis: -log10(FDR)
Color: Genes
Size: Genes
Font Size: 8
Circle Size: 4
Color:Low: blue
Chart type: barplot
Aspect Ratio: 1.5

Aussi possible avec enrichissement pathways KEGG, Reactome, Panther…

ShinyGO - suite

Clustering hiérarchique: corrélation entre les GO/pathways les plus enrichis
Basé sur le nombre de gènes en commun
Points les + gros : meilleures p-values

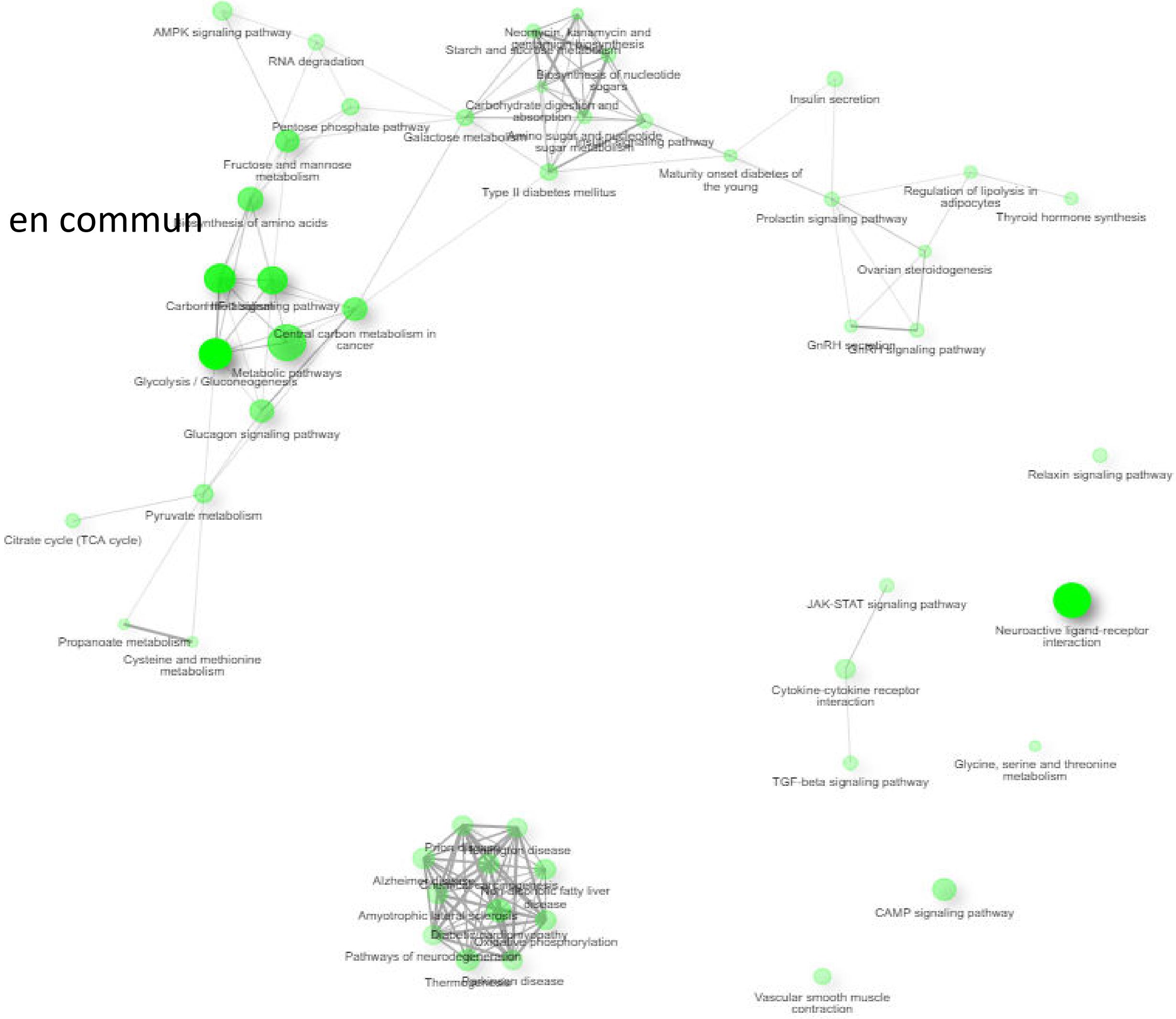
GO BP :



ShinyGO - suite

Network : relation entre les GO/pathways enrichis
2 nœuds reliés si au moins 20% de gènes en communs
Points les + foncés : meilleures p-value
Epaisseur des lignes proportionnelle au nombre de gènes en commun

KEGG pathways :



ShinyGO - suite

✚ **Classification des gènes par groupe** très intéressante :
gènes regroupés par **catégories fonctionnelles** définies par des termes GO de haut niveau.

N	High level GO category	Genes
58	Regulation of biological quality	TG UTS2 GAL PRLH AVP OXT LHB RETN HAMP GCK CALCA ATP5F1B GAPDH STC2 POMC GCG TTR NPPB IAPP COPA INHBA PFKFB2 INHA ADM2 EPO FSHB KL TSHB CGA ADCYAP1 PFKL VIP CRH ADM ALDOA PTH PFKM HK1 GHRL STC1 GIP HK2 HK3 INHBB UCN TRH APLN PRL BPGM NPPA CALCB UCN3 ADIPOQ CCK UTS2B GALP INS SCT
53	Response to external stimulus	PRLH ENO1 PTHLH AVP OXT LHB RETN HAMP RLN2 CALCA GAPDH STC2 POMC GCG IAPP ADM2 EPO COX4I1 FSHB TSHB CGA GH2 ADCYAP1 UCN2 VIP GNRH1 CRH ADM GPHA2 PTH HK1 SST GHRL PFKFB1 STC1 GIP INHBB UCN PRL NPPA CALCB UCN3 GPHB5 ADIPOQ GAST CCK GALP CSHL1 INSL3 INS GH1 SCT RLN3
50	Response to stress	GAL ENO1 PTHLH AVP OXT PGK1 LHB HAMP RLN2 CALCA NDUFS8 GAPDH STC2 POMC GCG IAPP INHBA ADM2 EPO FSHB TSHB CGA ADCYAP1 VIP CRH ADM GPHA2 PTH HK1 SST GHRL PFKFB1 NDUFS2 STC1 GIP HK2 INHBB UCN TRH NPPA CALCB UCN3 GPHB5 ADIPOQ CCK GALP INSL3 INS SCT RLN3
48	Immune system process	PKM GAL PTHLH THPO AVP LHB RETN HAMP RLN2 ALDOC CALCA GAPDH POMC GCG TTR IAPP INHBA INHA ADM2 EPO FSHB TSHB CGA ADCYAP1 PFKL VIP GNRH1 CRH ADM GPHA2 ALDOA PTH HK1 GIP HK3 IGF2 PGAM1 APLN BPGM NPPA CALCB GPHB5 ADIPOQ GALP INSL3 INS SCT RLN3
48	Regulation of response to stimulus	ENO1 PTHLH THPO AVP LHB HAMP RLN2 CALCA POMC GCG IAPP INHBA INHA ADM2 EPO FSHB KL TSHB CGA GH2 INHBE ADCYAP1 VIP CRH ADM GPHA2 PTH GHRL GIP INHBB UCN IGF2 APLN PRL INHBC NPPA CALCB GPHB5 ADIPOQ CCK ENTPD5 CSHL1 INSL3 INS GH1 SCT NDUFS3 RLN3
46	Regulation of signaling	GAL ENO1 THPO AVP OXT RETN GCK CALCA POMC GCG IAPP INHBA PFKFB2 INHA EPO FSHB KL CGA GH2 INHBE ADCYAP1 PFKL CRH ADM PTH PFKM GHRL GIP INHBB UCN IGF2 TRH APLN PRL INHBC NPPA UCN3 GPHB5 ADIPOQ ENTPD5 CSHL1 INSL3 INS GH1 SCT NDUFS3
46	Regulation of multicellular organismal process	TG PKM GAL PRLH ENO1 PTHLH THPO AVP OXT PGK1 RETN HAMP RLN2 CALCA ATP5F1B GAPDH POMC NPPB IAPP INHBA INHA ADM2 FSHB KL CGA ADCYAP1 VIP CRH ADM PTH HK1 GHRL STC1 HK2 INHBB UCN IGF2 APLN PRL NPPA ADIPOQ CCK INSL3 INS GH1 SCT
45	Response to endogenous stimulus	PKM GAL PRLH OXT CGB2 LHB RETN GCK CALCA STC2 GCG INHBA EPO FSHB KL TSHB CGA GH2 ADCYAP1 UCN2 GNRH1 CRH ADM PTH SST GHRL PFKFB1 STC1 GIP INHBB UCN IGF2 TRH APLN PRL NPPA UCN3 GPHB5 ADIPOQ OSTN GALP CSHL1 INS GH1 PKLR
43	Regulation of localization	GAL AVP OXT RETN HAMP GCK CALCA ATP5F1B POMC GCG NPPB INHBA PFKFB2 INHA EPO FSHB CGA ADCYAP1 PFKL VIP GNRH1 CRH PTH PFKM SST GHRL STC1 GIP HK2 INHBB UCN TRH APLN PRL NPPA UCN3 ADIPOQ CCK OSTN INSL3 INS GH1 SCT
39	Immune response	PKM PTHLH AVP LHB RETN HAMP RLN2 ALDOC CALCA GAPDH POMC GCG TTR IAPP ADM2 FSHB TSHB CGA ADCYAP1 PFKL VIP CRH

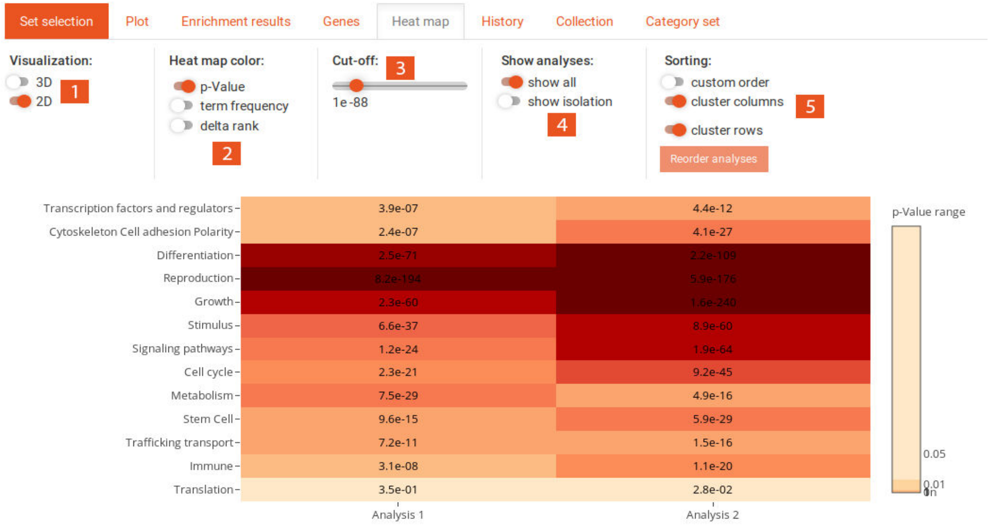
WEADE
Workflow for Enrichment Analysis
and Data Exploration

- Publié en 2018 mais pas mis à jour depuis
- Pas bos taurus
- Grandes listes (> 3000 ID)

http://beta-weade.cos.uni-heidelberg.de/weade

⛔ !!! Prometteur mais pas mis à jour depuis 2018, nouvelle version prévue en 2023

➕ Exploration génomique de jeux de données multiples : enrichissement, comparaison de listes intra ou inter-espèces, analyses de sous-listes.
7 espèces disponibles : h. sapiens, m. musculus, d. rerio, d. melogaster, a. thaliana, c. elegans and h. vulgaris



Visualization: ○ 3D ● 2D

Heat map color: ● p-Value ○ term frequency ○ delta rank

Cut-off: 1e-88

Show analyses: ● show all ○ show isolation

Sorting: ○ custom order ● cluster columns ● cluster rows

	Analysis 1	Analysis 2
Transcription factors and regulators	3.9e-07	4.4e-12
Cytoskeleton Cell adhesion Polarity	2.4e-07	4.1e-27
Differentiation	2.5e-71	2.2e-109
Reproduction	8.2e-194	5.9e-176
Growth	2.3e-60	1.6e-240
Stimulus	6.6e-37	8.9e-60
Signaling pathways	1.2e-24	1.9e-64
Cell cycle	2.3e-21	9.2e-45
Metabolism	7.5e-29	4.9e-16
Stem Cell	9.6e-15	5.9e-29
Trafficking transport	7.2e-11	1.5e-16
Immune	3.1e-08	1.1e-20
Translation	3.5e-01	2.8e-02

Venn diagram showing the uploaded sets

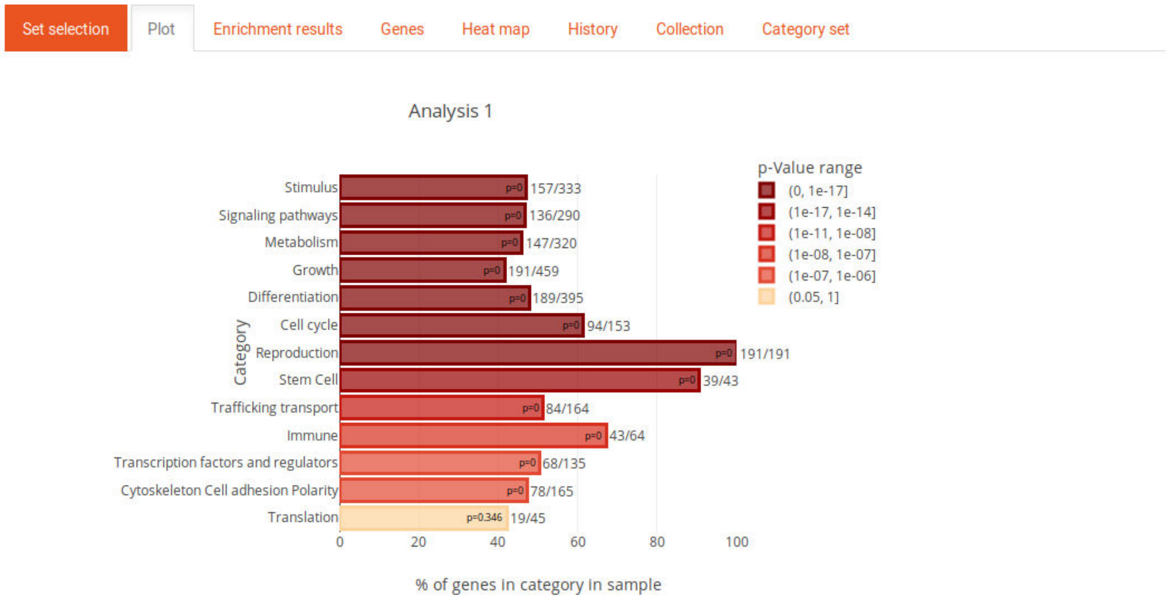
● foreground
○ background

When no background is selected, all annotated GO terms are used

Name for analysis: Analysis 1

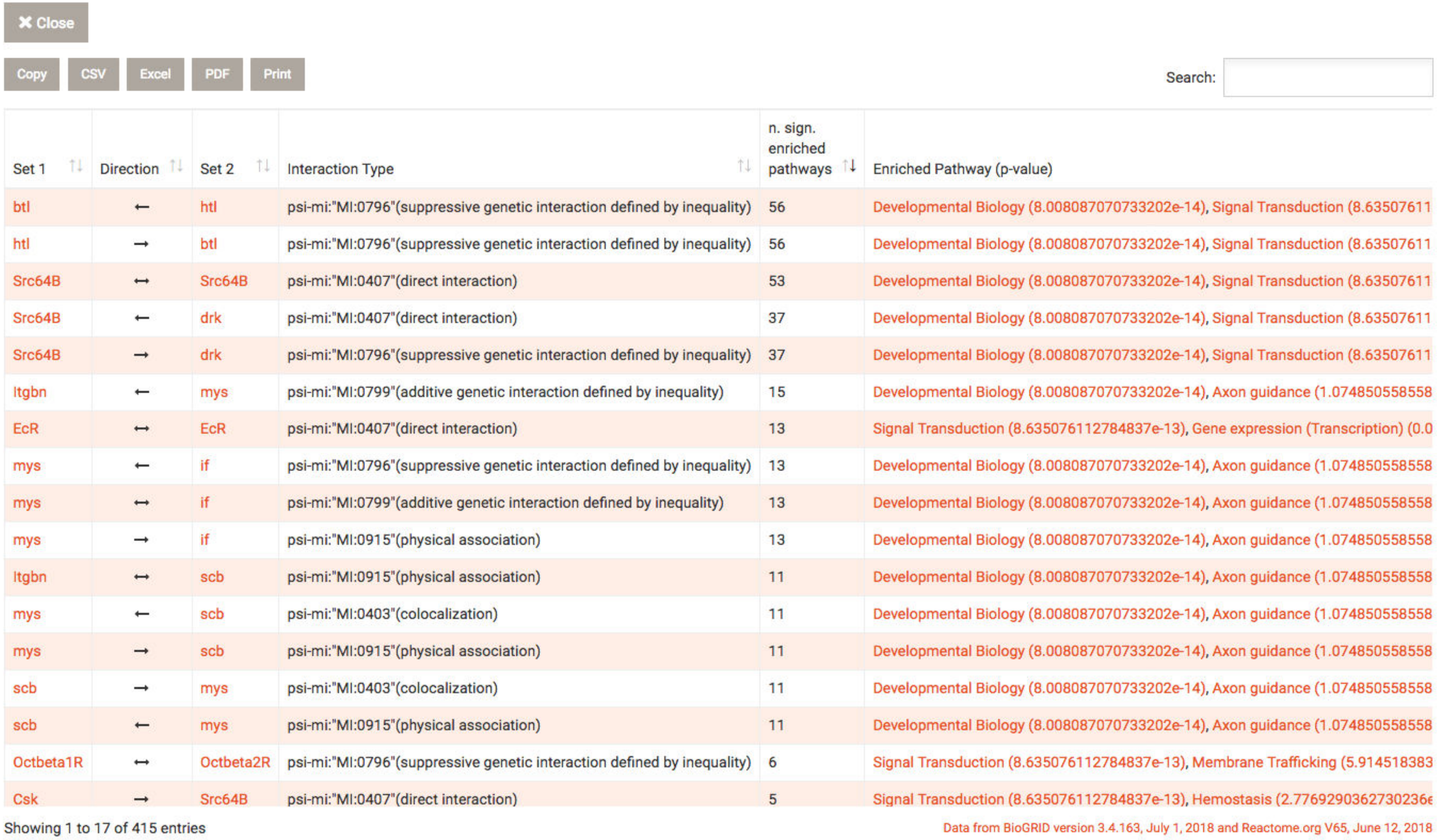
WEADE - suite

Création de GO term category (choix de mots-clés) pour pouvoir faire comparaison inter-espèce
→ clusterise les GO term classiques (espèce dépendants) en catégories plus générales



WEADE - suite

Interactions protéiques entre 2 listes et pathway dans lesquels intervient cette interaction:



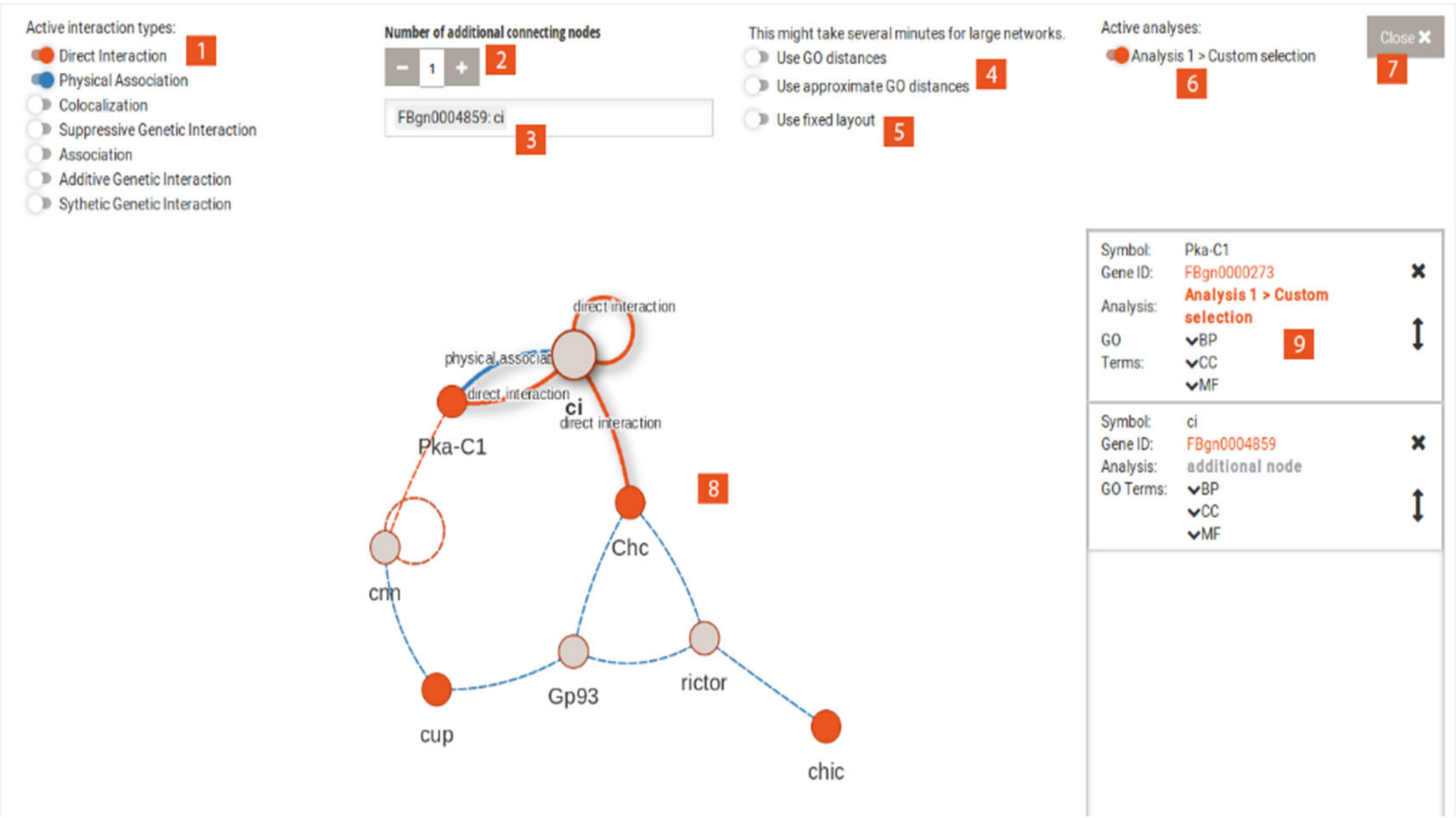
Set 1	Direction	Set 2	Interaction Type	n. sign. enriched pathways	Enriched Pathway (p-value)
btl	←	htl	psi-mi:"MI:0796"(suppressive genetic interaction defined by inequality)	56	Developmental Biology (8.008087070733202e-14), Signal Transduction (8.63507611
htl	→	btl	psi-mi:"MI:0796"(suppressive genetic interaction defined by inequality)	56	Developmental Biology (8.008087070733202e-14), Signal Transduction (8.63507611
Src64B	↔	Src64B	psi-mi:"MI:0407"(direct interaction)	53	Developmental Biology (8.008087070733202e-14), Signal Transduction (8.63507611
Src64B	←	drk	psi-mi:"MI:0407"(direct interaction)	37	Developmental Biology (8.008087070733202e-14), Signal Transduction (8.63507611
Src64B	→	drk	psi-mi:"MI:0796"(suppressive genetic interaction defined by inequality)	37	Developmental Biology (8.008087070733202e-14), Signal Transduction (8.63507611
Itgbn	←	mys	psi-mi:"MI:0799"(additive genetic interaction defined by inequality)	15	Developmental Biology (8.008087070733202e-14), Axon guidance (1.074850558558
EcR	↔	EcR	psi-mi:"MI:0407"(direct interaction)	13	Signal Transduction (8.635076112784837e-13), Gene expression (Transcription) (0.0
mys	←	if	psi-mi:"MI:0796"(suppressive genetic interaction defined by inequality)	13	Developmental Biology (8.008087070733202e-14), Axon guidance (1.074850558558
mys	↔	if	psi-mi:"MI:0799"(additive genetic interaction defined by inequality)	13	Developmental Biology (8.008087070733202e-14), Axon guidance (1.074850558558
mys	→	if	psi-mi:"MI:0915"(physical association)	13	Developmental Biology (8.008087070733202e-14), Axon guidance (1.074850558558
Itgbn	↔	scb	psi-mi:"MI:0915"(physical association)	11	Developmental Biology (8.008087070733202e-14), Axon guidance (1.074850558558
mys	←	scb	psi-mi:"MI:0403"(colocalization)	11	Developmental Biology (8.008087070733202e-14), Axon guidance (1.074850558558
mys	→	scb	psi-mi:"MI:0915"(physical association)	11	Developmental Biology (8.008087070733202e-14), Axon guidance (1.074850558558
scb	→	mys	psi-mi:"MI:0403"(colocalization)	11	Developmental Biology (8.008087070733202e-14), Axon guidance (1.074850558558
scb	←	mys	psi-mi:"MI:0915"(physical association)	11	Developmental Biology (8.008087070733202e-14), Axon guidance (1.074850558558
Octbeta1R	↔	Octbeta2R	psi-mi:"MI:0796"(suppressive genetic interaction defined by inequality)	6	Signal Transduction (8.635076112784837e-13), Membrane Trafficking (5.914518383
Csk	→	Src64B	psi-mi:"MI:0407"(direct interaction)	5	Signal Transduction (8.635076112784837e-13), Hemostasis (2.7769290362730236

Showing 1 to 17 of 415 entries

Data from BioGRID version 3.4.163, July 1, 2018 and Reactome.org V65, June 12, 2018

WEADE - suite

Réseau d'interactions fonctionnel (biogrid):



Active interaction types:
- ● Direct Interaction
- ● Physical Association
- ○ Colocalization
- ○ Suppressive Genetic Interaction
- ○ Association
- ○ Additive Genetic Interaction
- ○ Sythetic Genetic Interaction

Number of additional connecting nodes: 1

FBgn0004859: ci

This might take several minutes for large networks.
- ○ Use GO distances
- ○ Use approximate GO distances
- ○ Use fixed layout

Active analyses:
- ● Analysis 1 > Custom selection

Symbol: Pka-C1
Gene ID: FBgn0000273
Analysis: **Analysis 1 > Custom selection**
GO Terms: ✓BP ✓CC ✓MF

Symbol: ci
Gene ID: FBgn0004859
Analysis: **additional node**
GO Terms: ✓BP ✓CC ✓MF

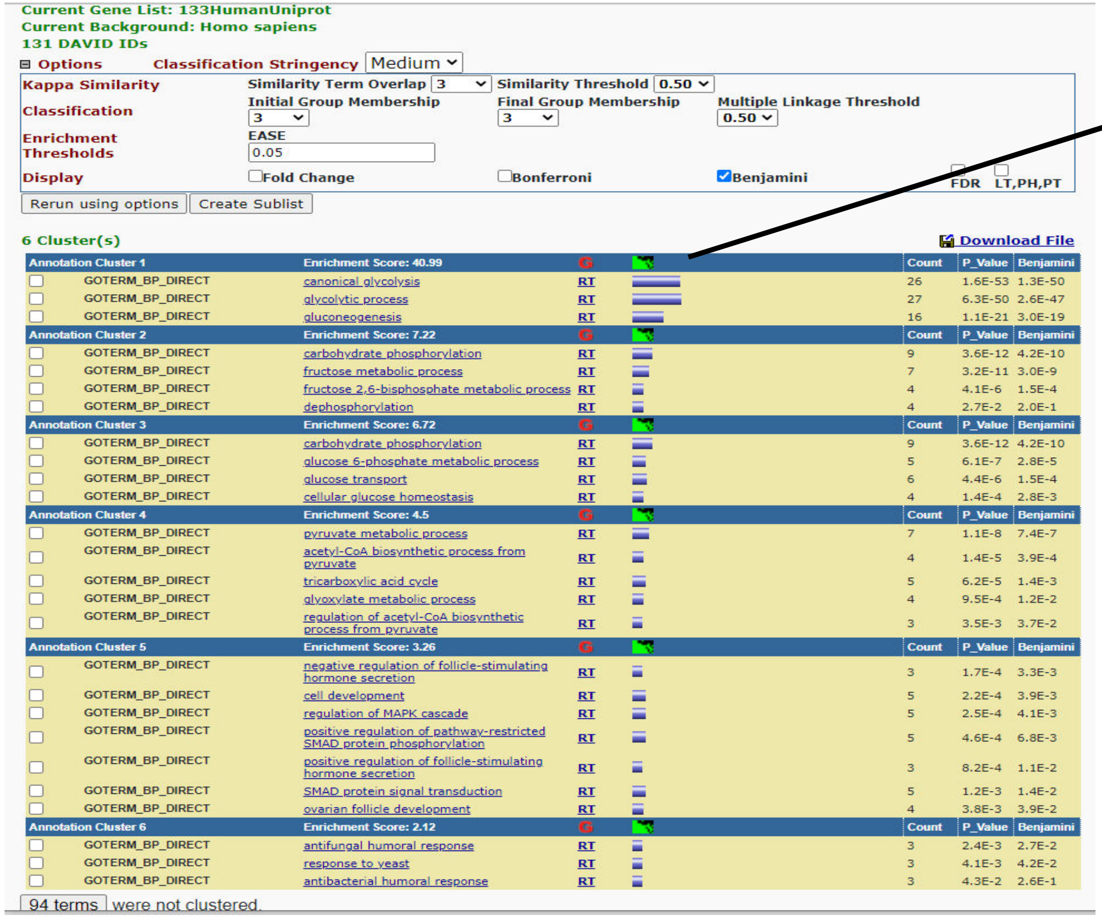
Possibilité de rajouter des gènes

https://david.ncifcrf.gov/tools.jsp

- Publié en 2009, dernière mise à jour 2022
- Nombreuses espèces
- Listes < 3000 ID

DAVID Functional Annotation Clustering Tool :

- condense les annotations redondantes pour faciliter l’interprétation biologique (compare gènes en commun).
- Calcul un score d’enrichissement pour chaque cluster (moyenne géométrique des p-values associées)



Current Gene List: 133HumanUniprot
Current Background: Homo sapiens
131 DAVID IDs

Options — Classification Stringency: Medium

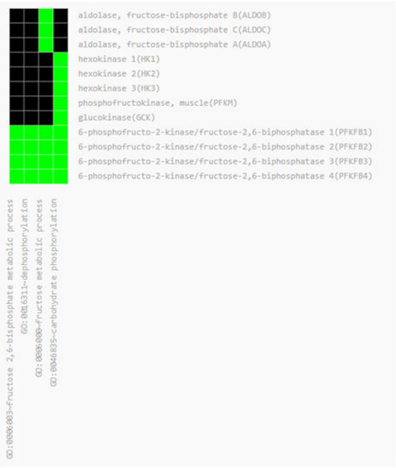
Kappa Similarity: Similarity Term Overlap 3; Similarity Threshold 0.50
Classification: Initial Group Membership 3; Final Group Membership 3; Multiple Linkage Threshold 0.50
Enrichment Thresholds: EASE 0.05
Display: ☐ Fold Change ☐ Bonferroni ☑ Benjamini ☐ FDR ☐ LT,PH,PT

6 Cluster(s)

				Count	P_Value	Benjamini
Annotation Cluster 1		**Enrichment Score: 40.99**		**Count**	**P_Value**	**Benjamini**
☐	GOTERM_BP_DIRECT	canonical glycolysis	RT	26	1.6E-53	1.3E-50
☐	GOTERM_BP_DIRECT	glycolytic process	RT	27	6.3E-50	2.6E-47
☐	GOTERM_BP_DIRECT	gluconeogenesis	RT	16	1.1E-21	3.0E-19
Annotation Cluster 2		**Enrichment Score: 7.22**		**Count**	**P_Value**	**Benjamini**
☐	GOTERM_BP_DIRECT	carbohydrate phosphorylation	RT	9	3.6E-12	4.2E-10
☐	GOTERM_BP_DIRECT	fructose metabolic process	RT	7	3.2E-11	3.0E-9
☐	GOTERM_BP_DIRECT	fructose 2,6-bisphosphate metabolic process	RT	4	4.1E-6	1.5E-4
☐	GOTERM_BP_DIRECT	dephosphorylation	RT	4	2.7E-2	2.0E-1
Annotation Cluster 3		**Enrichment Score: 6.72**		**Count**	**P_Value**	**Benjamini**
☐	GOTERM_BP_DIRECT	carbohydrate phosphorylation	RT	9	3.6E-12	4.2E-10
☐	GOTERM_BP_DIRECT	glucose 6-phosphate metabolic process	RT	5	6.1E-7	2.8E-5
☐	GOTERM_BP_DIRECT	glucose transport	RT	6	4.4E-6	1.5E-4
☐	GOTERM_BP_DIRECT	cellular glucose homeostasis	RT	4	1.4E-4	2.8E-3
Annotation Cluster 4		**Enrichment Score: 4.5**		**Count**	**P_Value**	**Benjamini**
☐	GOTERM_BP_DIRECT	pyruvate metabolic process	RT	7	1.1E-8	7.4E-7
☐	GOTERM_BP_DIRECT	acetyl-CoA biosynthetic process from pyruvate	RT	4	1.4E-5	3.9E-4
☐	GOTERM_BP_DIRECT	tricarboxylic acid cycle	RT	5	6.2E-5	1.4E-3
☐	GOTERM_BP_DIRECT	glyoxylate metabolic process	RT	4	9.5E-4	1.2E-2
☐	GOTERM_BP_DIRECT	regulation of acetyl-CoA biosynthetic process from pyruvate	RT	3	3.5E-3	3.7E-2
Annotation Cluster 5		**Enrichment Score: 3.26**		**Count**	**P_Value**	**Benjamini**
☐	GOTERM_BP_DIRECT	negative regulation of follicle-stimulating hormone secretion	RT	3	1.7E-4	3.3E-3
☐	GOTERM_BP_DIRECT	cell development	RT	5	2.2E-4	3.9E-3
☐	GOTERM_BP_DIRECT	regulation of MAPK cascade	RT	5	2.5E-4	4.1E-3
☐	GOTERM_BP_DIRECT	positive regulation of pathway-restricted SMAD protein phosphorylation	RT	5	4.6E-4	6.8E-3
☐	GOTERM_BP_DIRECT	positive regulation of follicle-stimulating hormone secretion	RT	3	8.2E-4	1.1E-2
☐	GOTERM_BP_DIRECT	SMAD protein signal transduction	RT	5	1.2E-3	1.4E-2
☐	GOTERM_BP_DIRECT	ovarian follicle development	RT	4	3.8E-3	3.9E-2
Annotation Cluster 6		**Enrichment Score: 2.12**		**Count**	**P_Value**	**Benjamini**
☐	GOTERM_BP_DIRECT	antifungal humoral response	RT	3	2.4E-3	2.7E-2
☐	GOTERM_BP_DIRECT	response to yeast	RT	3	4.1E-3	4.2E-2
☐	GOTERM_BP_DIRECT	antibacterial humoral response	RT	3	4.3E-2	2.6E-1

94 terms were not clustered.

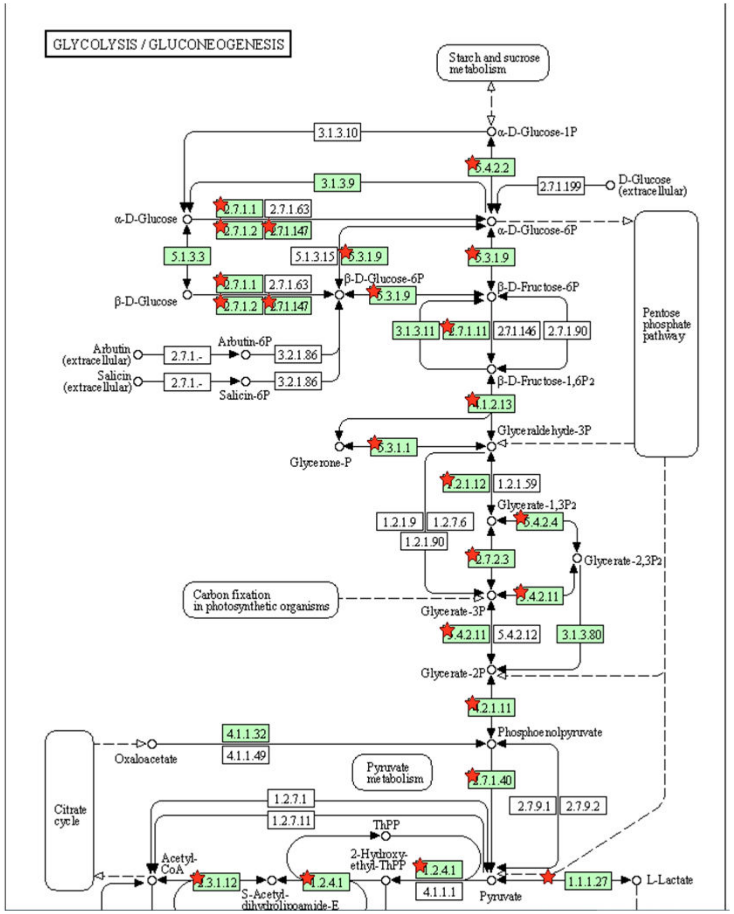
Pour chaque cluster : heatmap avec gènes impliqués



	GO:0006003~fructose 2,6-bisphosphate metabolic process	GO:0016311~dephosphorylation	GO:0006000~fructose metabolic process	GO:0046835~carbohydrate phosphorylation
aldolase, fructose-bisphosphate B(ALDOB)	⬛	⬛	🟩	⬛
aldolase, fructose-bisphosphate C(ALDOC)	⬛	⬛	🟩	⬛
aldolase, fructose-bisphosphate A(ALDOA)	⬛	⬛	🟩	⬛
hexokinase 1(HK1)	⬛	⬛	⬛	🟩
hexokinase 2(HK2)	⬛	⬛	⬛	🟩
hexokinase 3(HK3)	⬛	⬛	⬛	🟩
phosphofructokinase, muscle(PFKM)	⬛	⬛	⬛	🟩
glucokinase(GCK)	⬛	⬛	⬛	🟩
6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 1(PFKFB1)	🟩	🟩	🟩	🟩
6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2(PFKFB2)	🟩	🟩	🟩	🟩
6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 3(PFKFB3)	🟩	🟩	🟩	🟩
6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 4(PFKFB4)	🟩	🟩	🟩	🟩

DAVID - suite

Visualisation pathways KEGG et Biocarta
avec gènes/protéines d’intérêts mis en évidence



Metascape
A Gene Annotation & Analysis Resource

https://metascape.org

⛔
- Publié en 2015, mis à jour régulièrement
- Listes < 3000 ID

⊕ Interroge un très grand nombre de bases de données (>40) mises à jour tous les mois

10 espèces modèles : H. sapiens, M. musculus, R. norvegicus, D. rerio, D. melanogaster, C. elegans, S. cerevisiae, A. thaliana, S. Pombe, and P. falciparum **(mais pas bos taurus)**

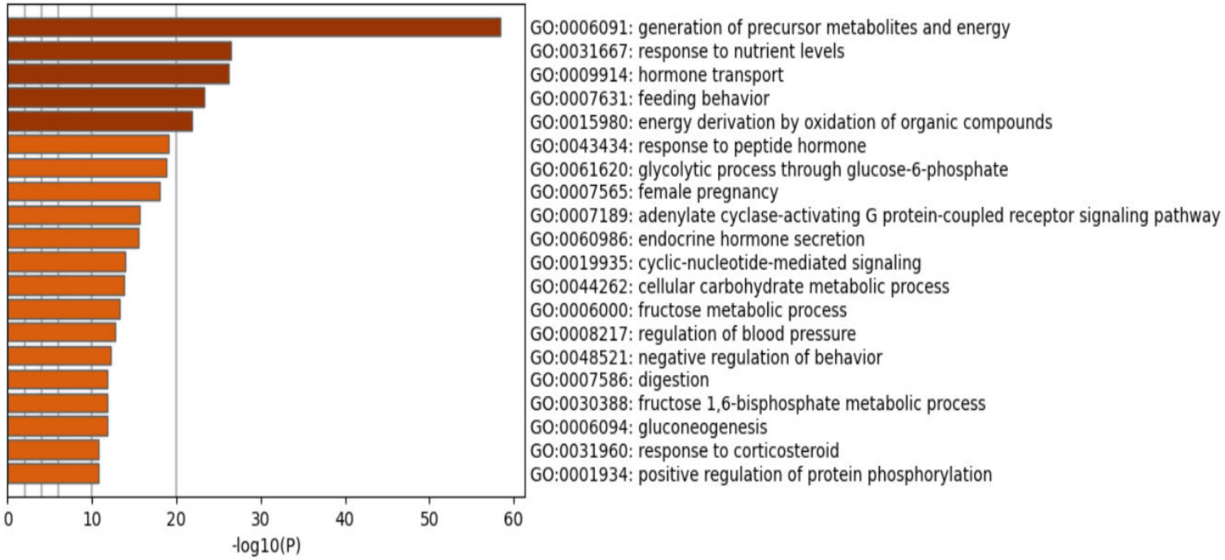
\+ recherche d’orthologues

⊕ **Clustering des termes enrichis : idem DAVID**

⚠ !!! Metascape garde le terme du cluster ayant la plus petite p-value pour nommer le cluster dans les graphs !!!
Bien vérifier que ce terme soit représentatif de l’ensemble du cluster

GO-BP : TOP 20 clustering Barchart :

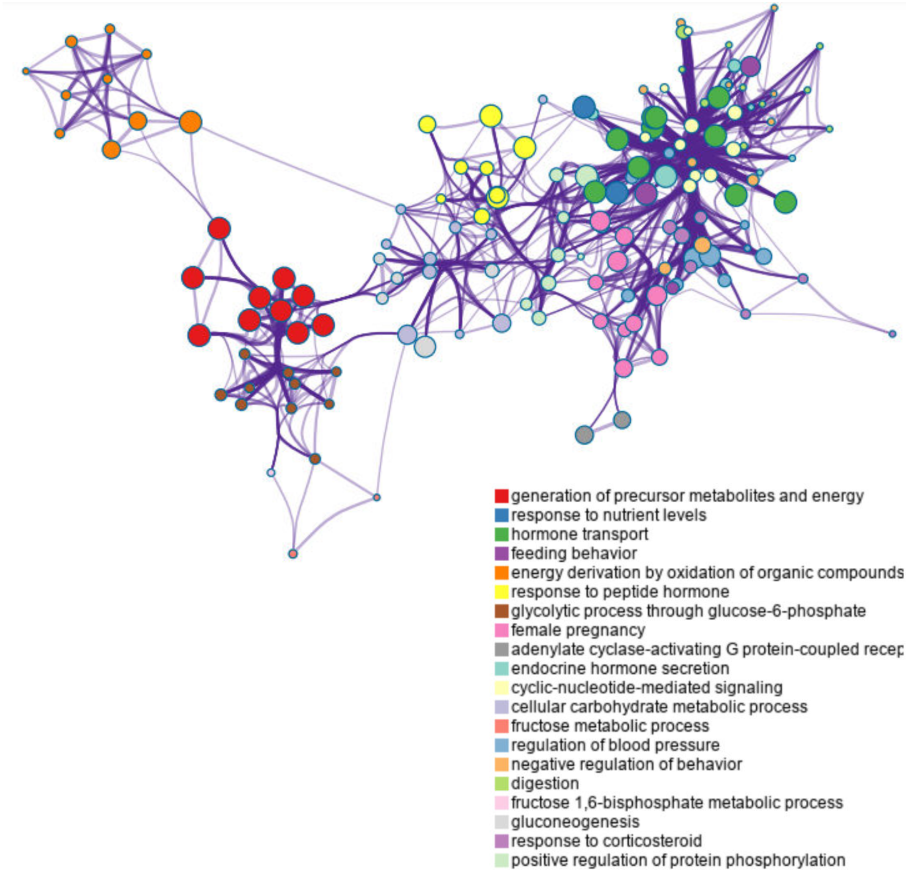
Top 100 disponible



Metascape - suite

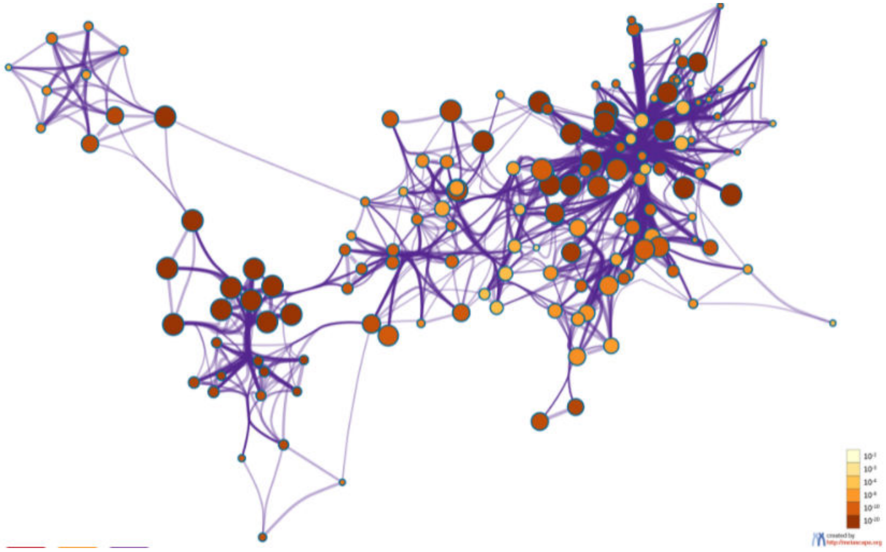
GO-BP : Clustering network :

Couleur par cluster :



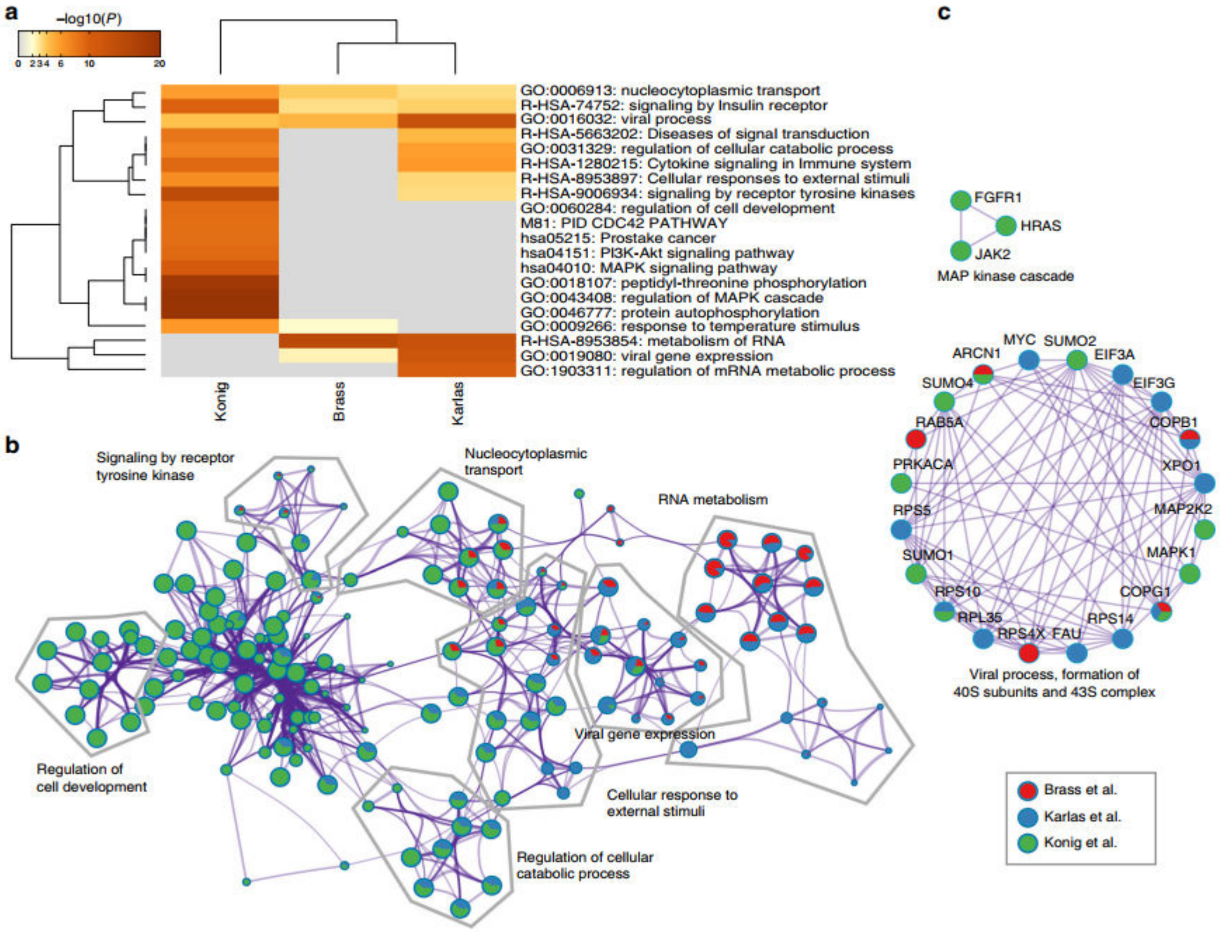
GO-BP : Clustering network :

Couleur par p-value :



Metascape - suite

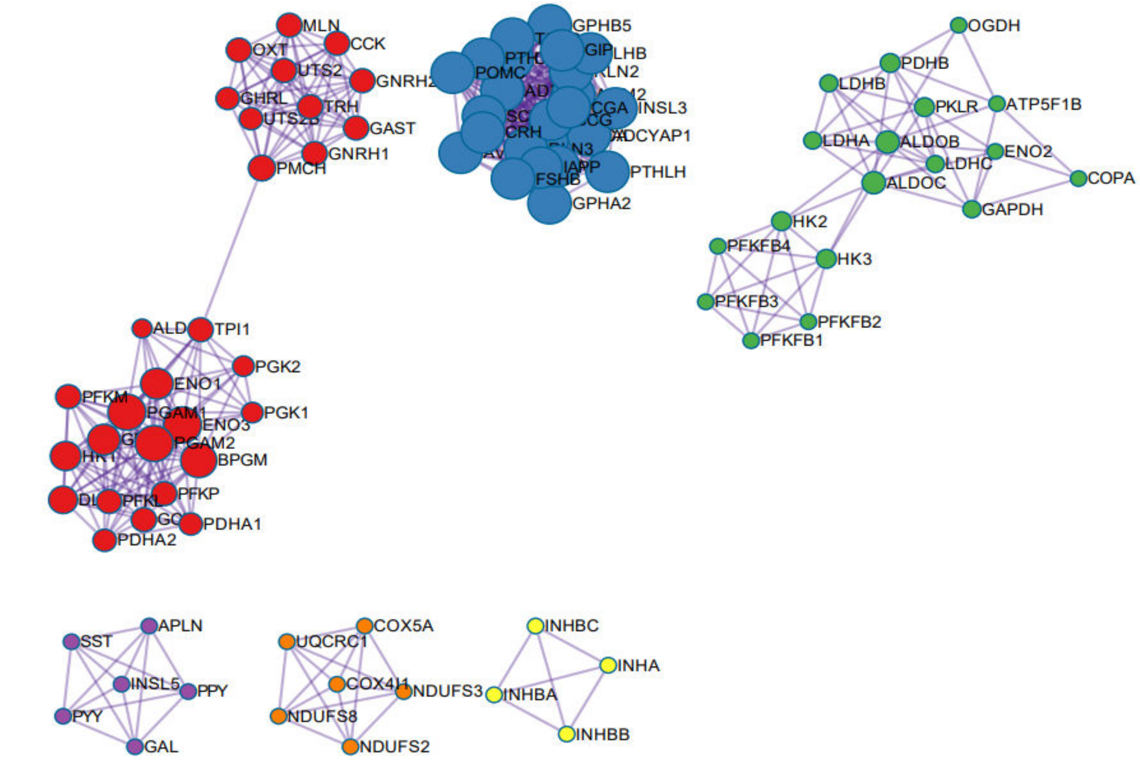
Comparaison de listes
Analyse de **données multi-omiques**



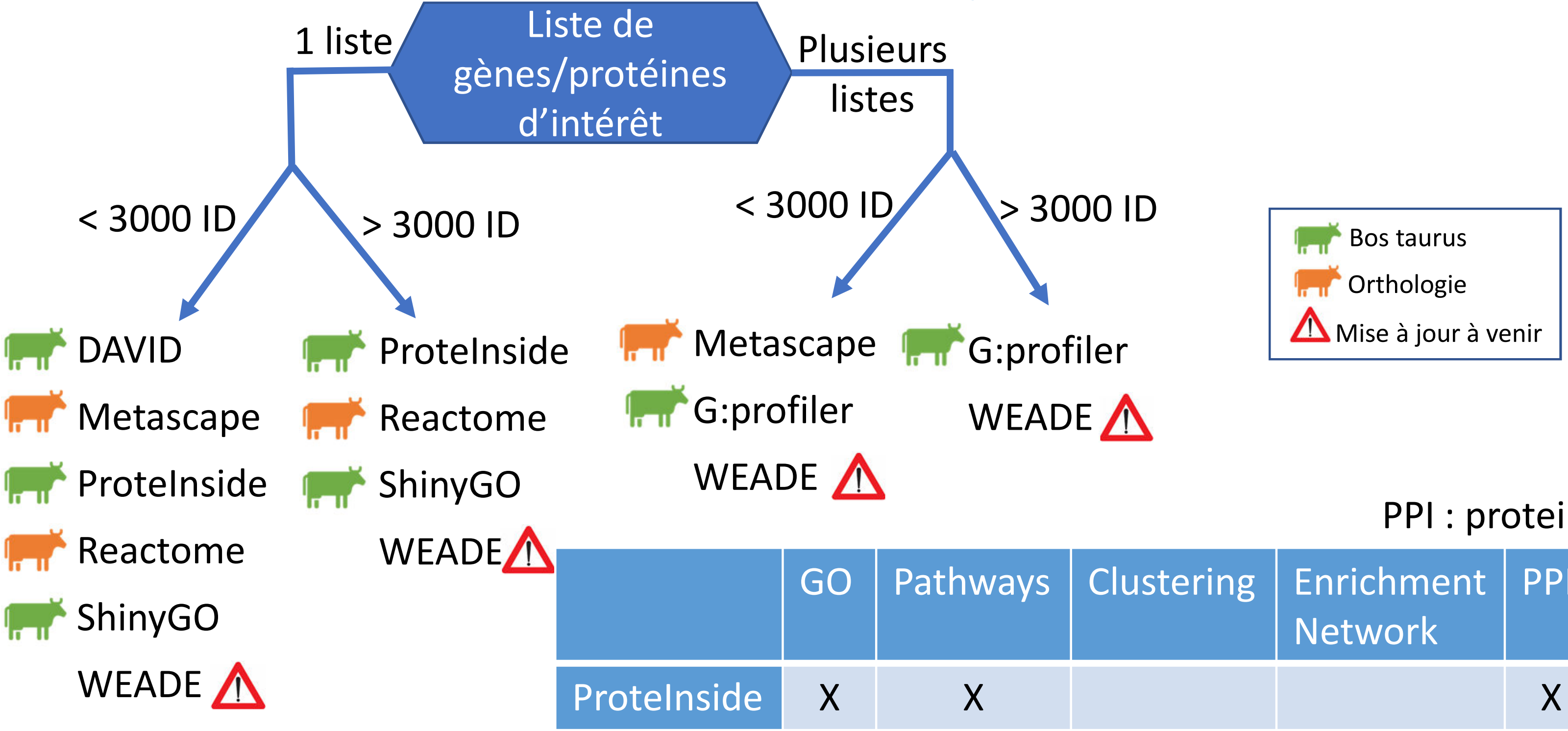
Circos plot

Metascape - suite

Interactions protéiques :



Synthèse



Liste de gènes/protéines d'intérêt

- 1 liste
  - < 3000 ID
    - 🟩 DAVID
    - 🟧 Metascape
    - 🟩 ProteInside
    - 🟧 Reactome
    - 🟩 ShinyGO
    - WEADE ⚠
  - > 3000 ID
    - 🟩 ProteInside
    - 🟧 Reactome
    - 🟩 ShinyGO
    - WEADE ⚠
- Plusieurs listes
  - < 3000 ID
    - 🟧 Metascape
    - 🟩 G:profiler
    - WEADE ⚠
  - > 3000 ID
    - 🟩 G:profiler
    - WEADE ⚠

🟩 Bos taurus
🟧 Orthologie
⚠ Mise à jour à venir

PPI : protein-protein interactions

	GO	Pathways	Clustering	Enrichment Network	PPI	Multi-list	>3000 ID	Bos taurus
ProteInside	X	X			X	PPi only	X	X
G:profiler	X	X				X	X	X
Reactome	X	X				X	X	orthologs
ShinyGO	X	X	X	X			X	X
DAVID	X	X	X					X
Metascape	X	X	X	X	X	X		orthologs
WEADE	X	X	X		X	X	X	

THANK YOU FOR

your attention