



L'étude des gènes et des protéines facilitée par l'utilisation du web service ProteINSIDE

KASPRIC Nicolas

**Thèse débutée en février 2013
Equipe Amuvi**

Encadrants :

Muriel BONNET

Brigitte PICARD

Avec l'appui de Matthieu REICHSTADT

Financement :

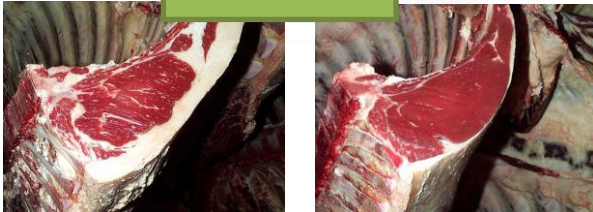
Région Auvergne

Apis-Gene



Contexte

Carcasses



Le muscle pour le rendement et la tendreté

Le tissu adipeux (TA) pour le goût mais diminue le rendement

Age > Race > Elevage

Variabilité importante du rapport masse TA / masse muscle



Des interactions entre TA et muscle suggérées

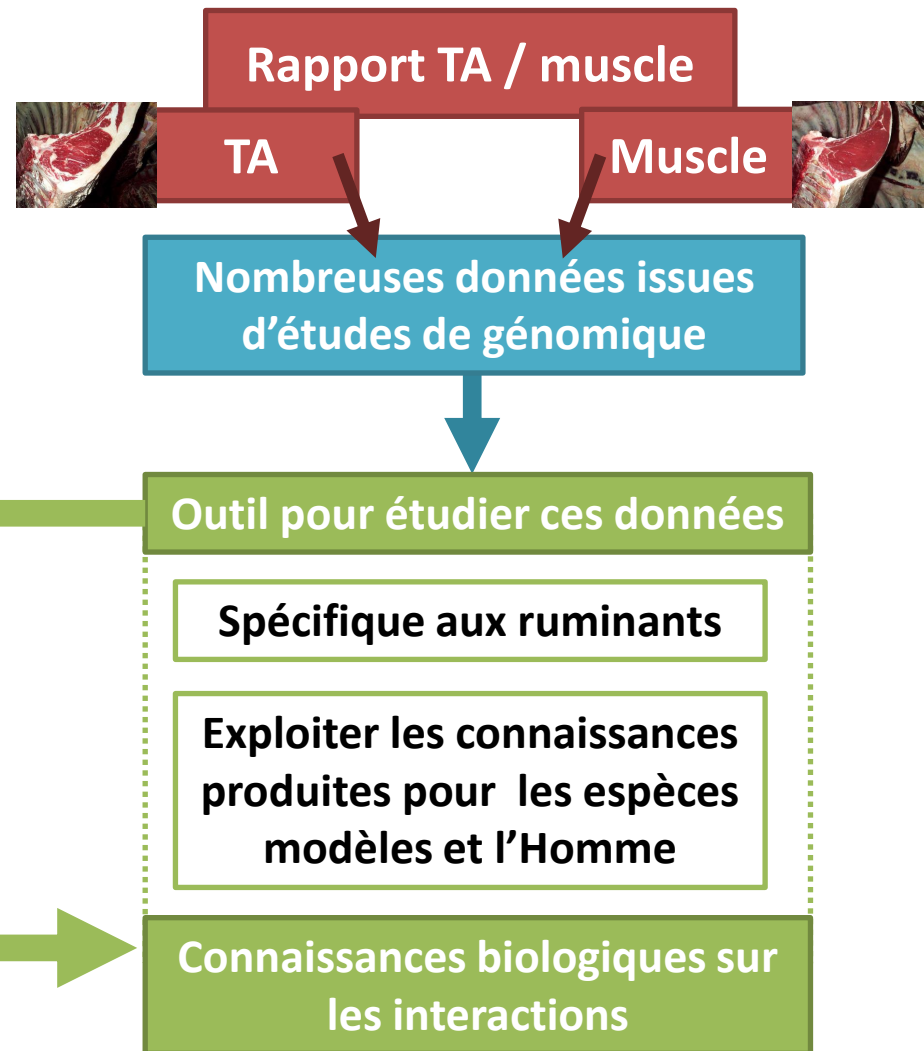
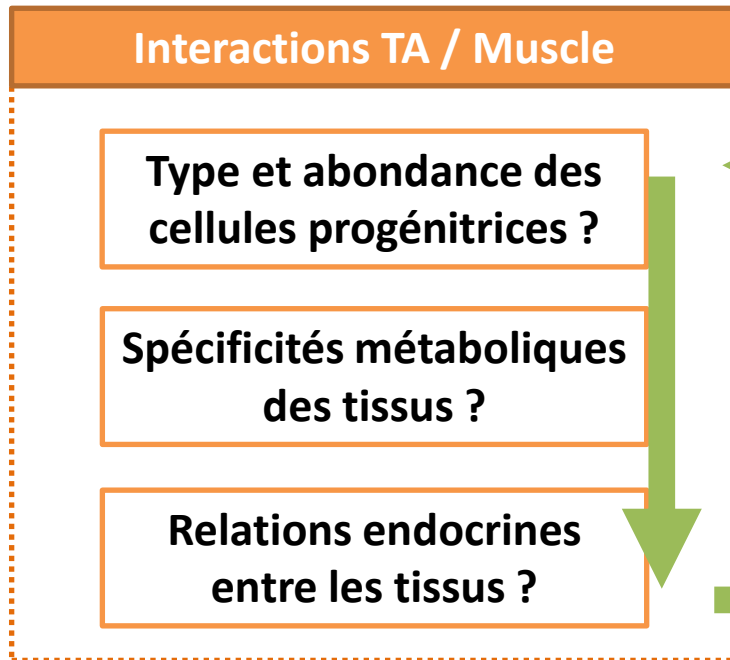
Chronologie dans la croissance des tissus

Variations inverses des masses de tissus: géotypes extrêmes

...et qui commencent à être expliquées



Contexte



Afin de maîtriser ce rapport, il faut mieux connaître les interactions entre le TA et le muscle



Contexte

Milliers de protéines / gènes à analyser

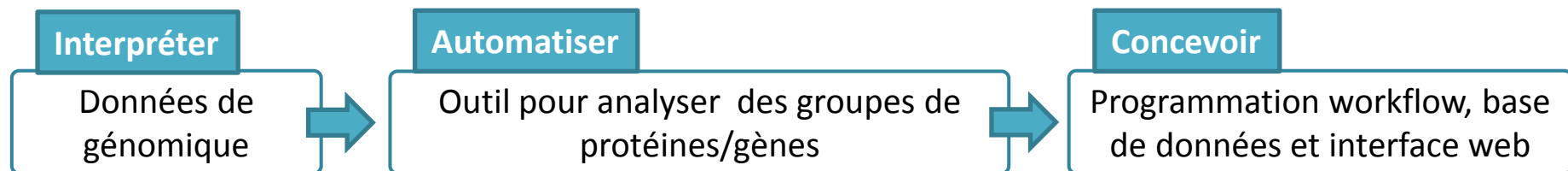
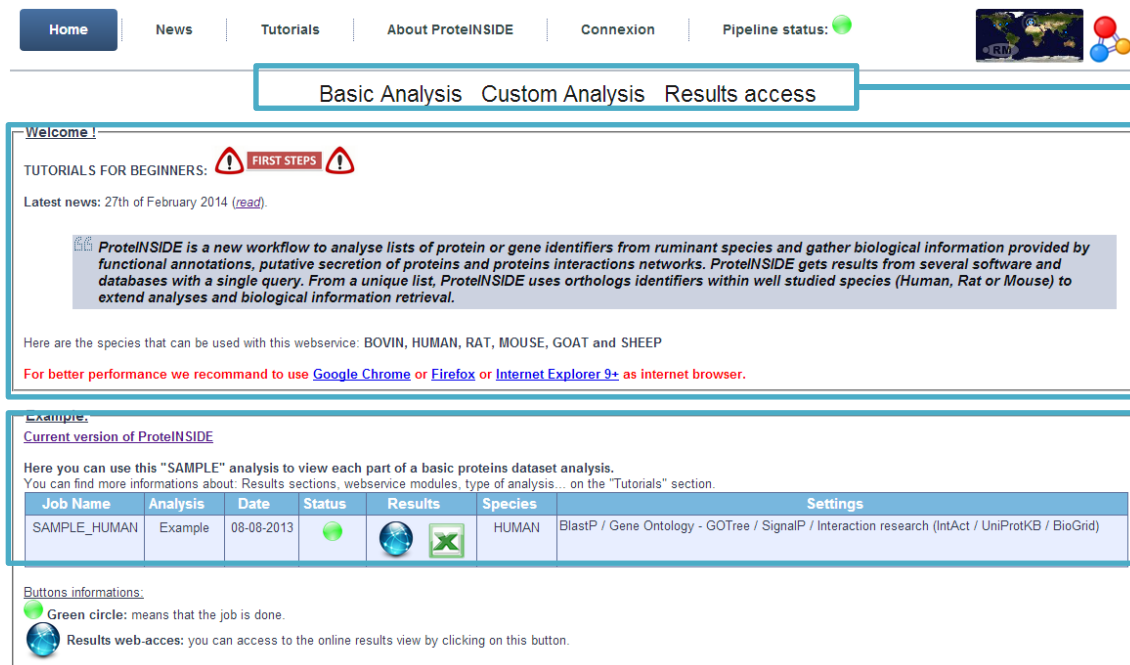
Manuellement

- Recherche d'identifiants biologiques
- Recherche de fonctions biologiques
- Rassemblement et tri des résultats
- Recherche de liens entre les protéines
- Comparaison avec les réseaux biologiques existants
-

Automatiquement

Inexistant

ProteINSIDE : un outil automatisé pour identifier, annoter et lier des protéines/gènes chez le ruminant.

Home | News | Tutorials | About ProteINSIDE | Connexion | Pipeline status: ●

Basic Analysis | Custom Analysis | Results access

Welcome!

TUTORIALS FOR BEGINNERS: ⚠ **FIRST STEPS** ⚠

Latest news: 27th of February 2014 ([read](#)).

ProteINSIDE is a new workflow to analyse lists of protein or gene identifiers from ruminant species and gather biological information provided by functional annotations, putative secretion of proteins and proteins interactions networks. ProteINSIDE gets results from several software and databases with a single query. From a unique list, ProteINSIDE uses orthologs identifiers within well studied species (Human, Rat or Mouse) to extend analyses and biological information retrieval.

Here are the species that can be used with this webservice: BOVIN, HUMAN, RAT, MOUSE, GOAT and SHEEP

For better performance we recommend to use [Google Chrome](#) or [Firefox](#) or [Internet Explorer 9+](#) as internet browser.

Example.

[Current version of ProteINSIDE](#)

Here you can use this "SAMPLE" analysis to view each part of a basic proteins dataset analysis.
You can find more informations about: Results sections, webservice modules, type of analysis... on the "Tutorials" section.

Job Name	Analysis	Date	Status	Results	Species	Settings
SAMPLE_HUMAN	Example	08-08-2013	●	 	HUMAN	BlastP / Gene Ontology - GOTree / SignalP / Interaction research (IntAct / UniProtKB / BioGrid)

Buttons informations:

● Green circle: means that the job is done.

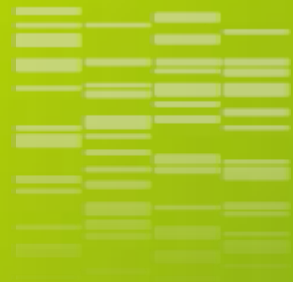
 Results web-access: you can access to the online results view by clicking on this button.

Accès aux analyses

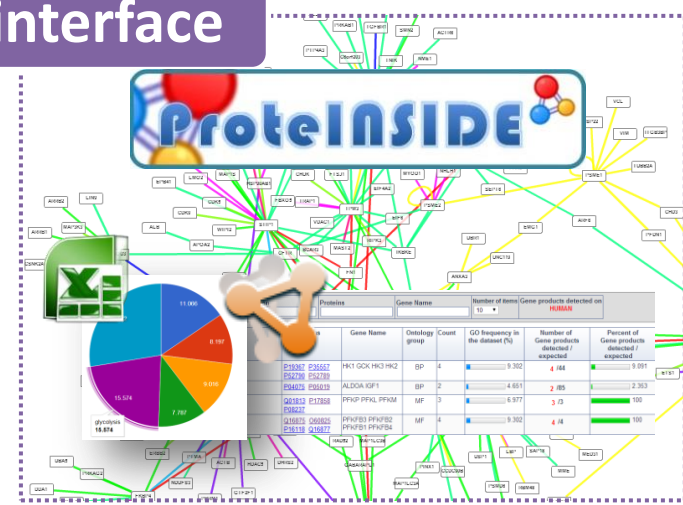
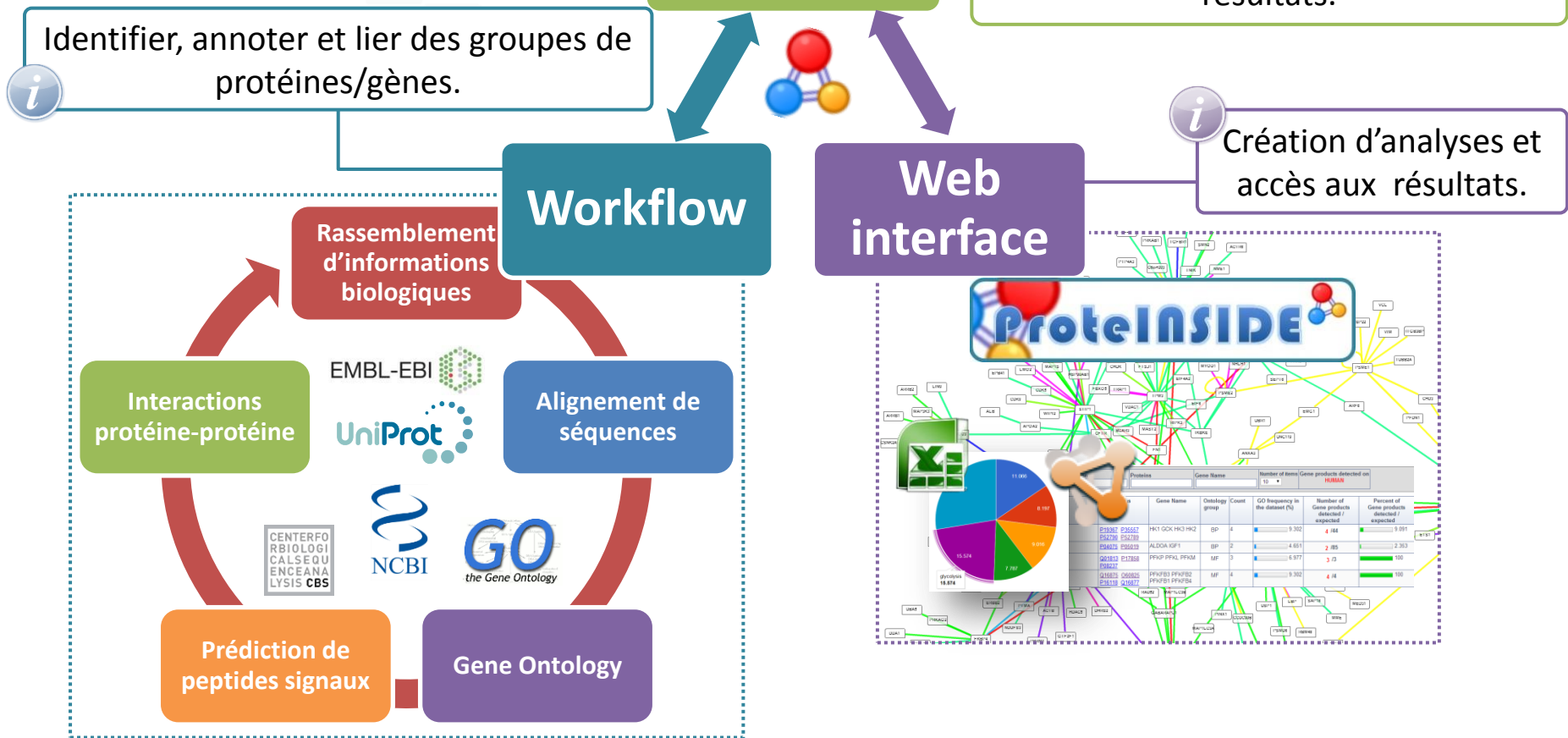
Informations sur ProteINSIDE

Exemple

Utilisation gratuite : www.proteinside.org

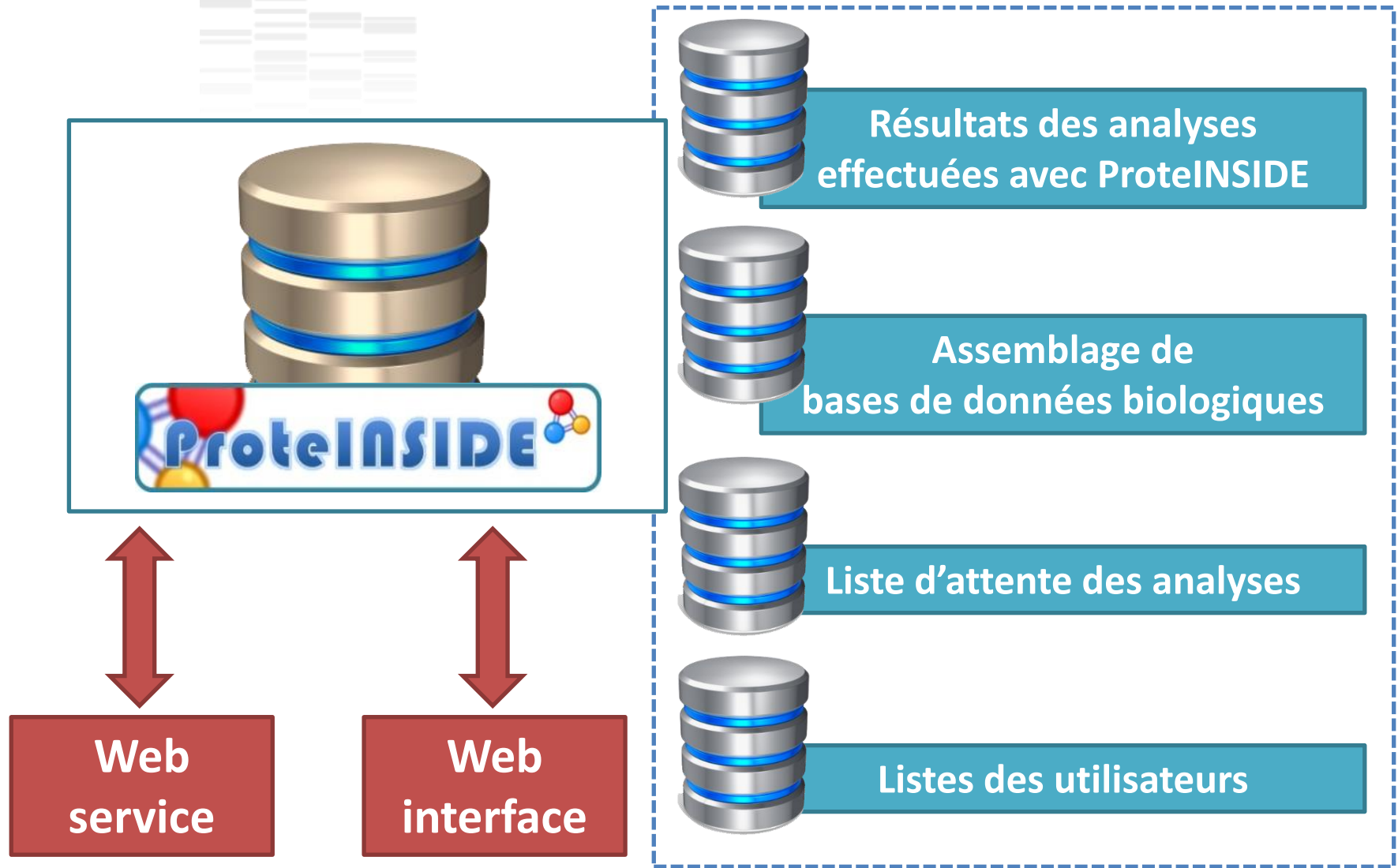


1] Conception de ProteINSIDE





La base de données de ProteINSIDE





La base de données de ProteINSIDE

Informations biologiques



- Genes
- HomoloGene



Ontologie fonctionnelle



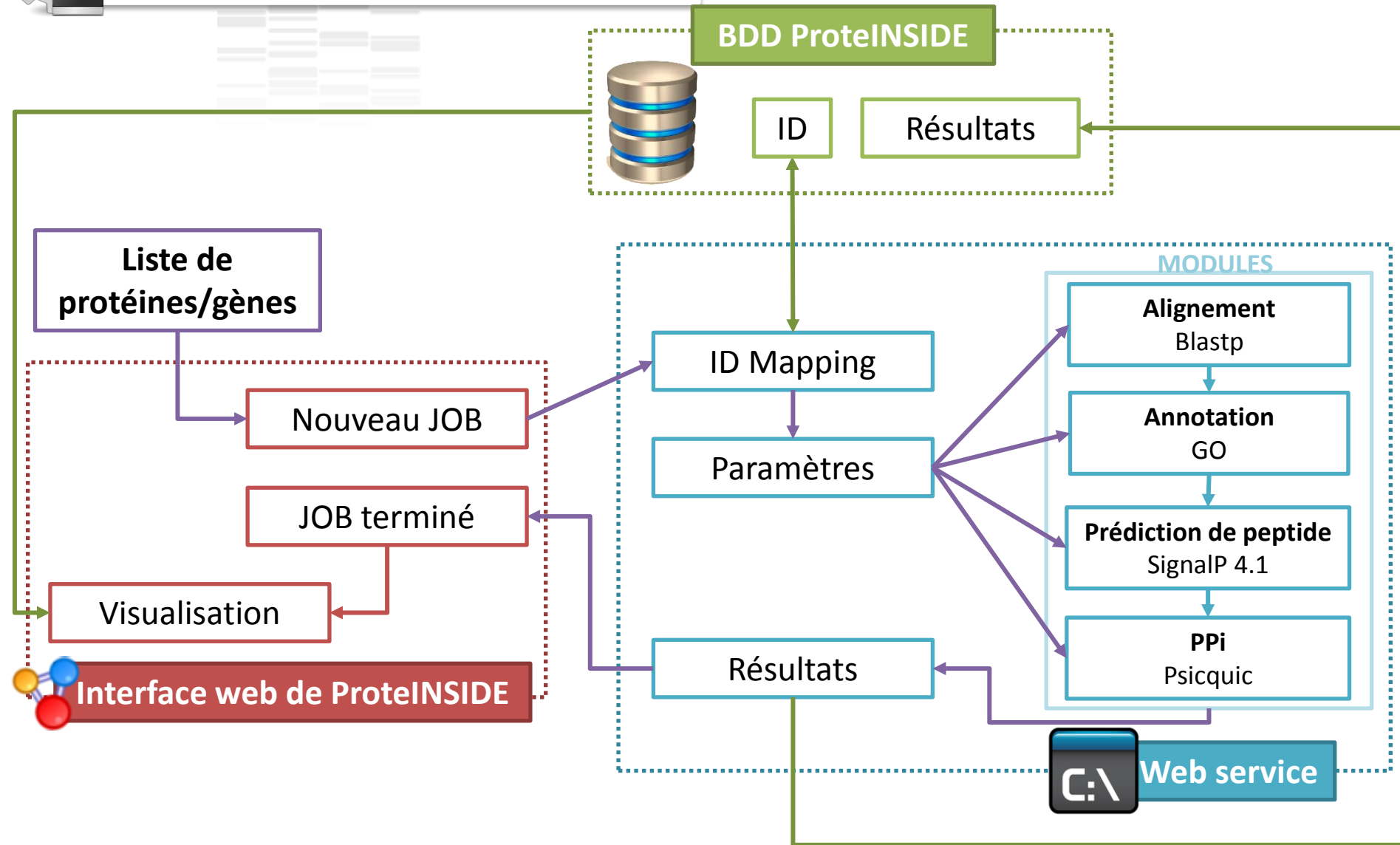
Scripts d'appel, de téléchargement, trie et d'assemblage des données
MAJ mensuelle



Assemblage de
données biologiques



Le web service ProteINSIDE





L'interface web de ProteINSIDE

Utilisateurs



Compte personnel
Manager de travaux
Accès à toutes les analyses
Accès à tous les outils



Visiteurs



Pas de compte
Résultats disponible 1 mois
Accès restreint



Interface Web



ProteINSIDE
ProteCONVERT



Résultats en ligne
Résultats téléchargeable



Tutoriaux
Mise à jours et version du service
Statistiques du service
Citations



L'interface web de ProteINSIDE



IDS de protéines ou de gènes

ProteCONVERT
Convertisseur d'identifiants

Basic Analysis Custom Analysis Preset Analysis
Workflow d'analyse des protéines et des gènes

Job name: (30 characters max)

Ex : job_1

1

IDs from the following species are analysed: BOVIN, HUMAN, RAT, MOUSE, GOAT and SHEEP

Select your species:

2

BOVIN

INPUT

File format (250 ko max / 8000 IDs max):

- + ☐ Microsoft Excel file (.xls or .xlsx)
- + ☐ Tabulation file (.tab or .txt)

OR
OR

PASTE

Paste your IDs (8000 IDs max):

- + ☒ Direct input

P07738
Q14556
O60825
P00338
P00558
P04075
P04406
P05062
P06733
P06744
P07496

3

4

Run the job !

Placé en file d'attente
du workflow



L'interface web de ProteINSIDE

User's main page

Name: IWBBIO2014 Tester Job(s) in progress: 0 Change account settings
Account type: user
User's job send: 9 ([All jobs](#))

Latest news: 28th of March 2014 ([read](#)).

[Current version of ProteINSIDE](#)

Utilisateurs



User's job(s)

Job Name Number of rows

Del	Re-Analyze	Job Name	Analysis	Date	Status	Results	Species	Settings
		ANA_bovine_proteconvert	Base	01-27-2014			BOVIN	ID Mapping / Gene Ontology / SignalP / Interaction research (IntAct / UniProtKB / BioGrid)
		SampleCV	Custom	03-20-2014			HUMAN	ID Mapping / Gene Ontology -- GOTree -- IEA active /
		ANA_bovine_nico_custom	Custom	01-27-2014			BOVIN	ID Mapping / Gene Ontology -- IEA active / SignalP / Interaction research -- Enrichment with HUMAN /



Terminée



File d'attente



En cours



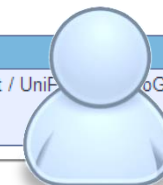
Accès aux résultats

Visitor's job(s) :

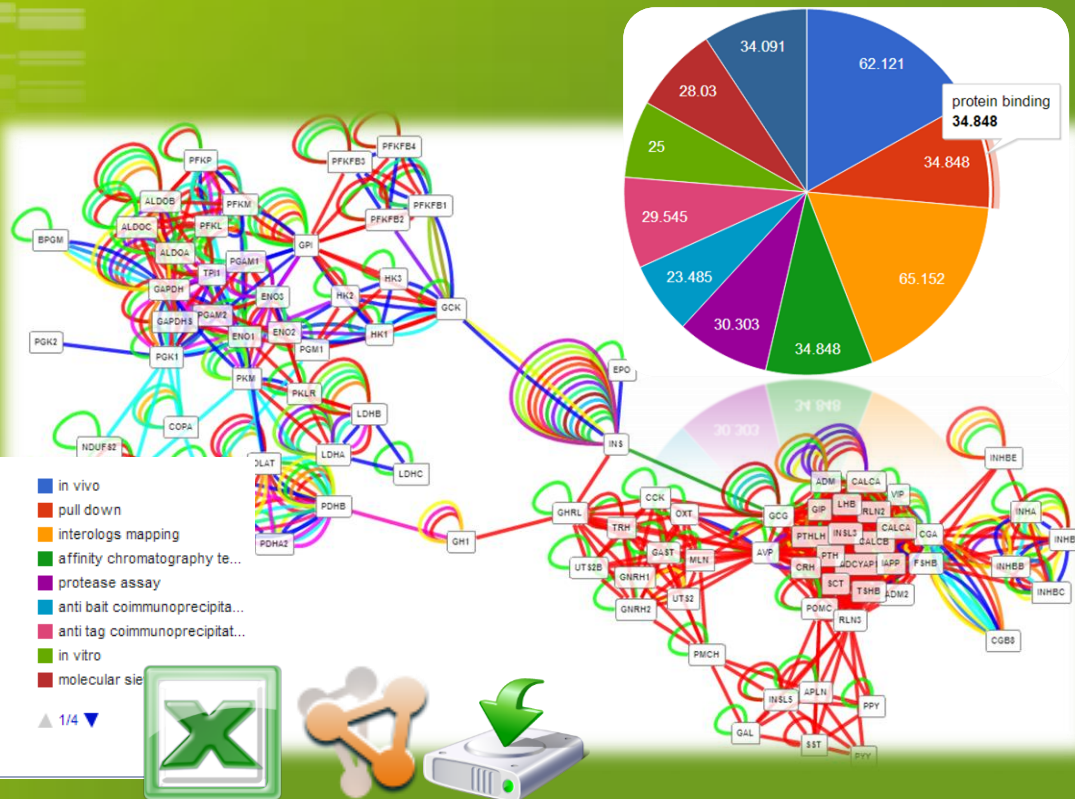
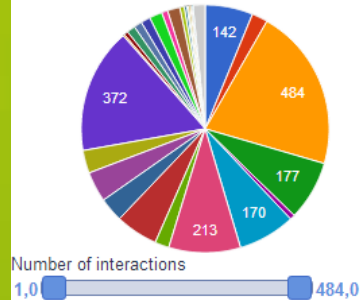
Access code for this job: c9ts6e701882867barb75rjc0j41py
Warning: this job will be remove the 06-11-2014.

Del	Re-Analyze	Job Name	Analysis	Date	Status	Results	Species	Settings
Disable		SampleBV	Base	05-12-2014			HUMAN	ID Mapping / Gene Ontology / SignalP / Interaction research (IntAct / UniProtKB / BioGrid)

Visiteurs



Detection methods for PPI outside the dataset



2] La biologie avec ProteINSIDE



Exemple 133 protéines :

- Glycolyse
- Chaîne respiratoire
- Hormones



Informations
biologiques

ProteINSIDE synthétise automatiquement les informations disponibles pour un gène/protéine dans les bases de données biologiques

Module de récupération d'informations biologique : simple appel à la base de données de ProteINSIDE

Gene Name	Proteins	Function	Tissue	Subcellular location	Number of rows	Download table	Sequences (FASTA)
					10		

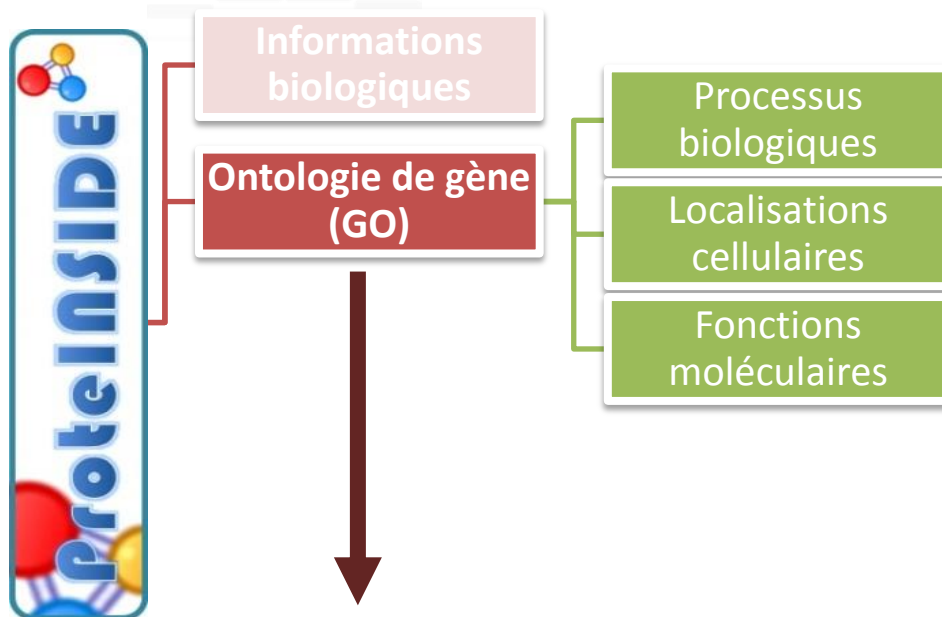
Query	Proteins	Protein ID	Gene Name	Gene Entrez	Function	Chromosome	Tissue	Subcellular location
50	Q9Y5Q6	INSL5_HUMAN	INSL5	10022	May have a role in gut contractility or in thymic development and regulation. Activates RXFP4 with high potency and appears to be the endogenous ligand for this receptor.	1p31.3	Highly expressed in rectum with lower levels in uterus and ascending and descending colon.	Secreted.
51	Q9Y581	INSL6_HUMAN	INSL6	11172	May have a role in sperm development and fertilization.	9p24	Testis specific.	Secreted (By similarity).
3	Q9ULZ1	APEL_HUMAN	APLN	8842	Endogenous ligand for APJ, an alternative coreceptor with CD4 for HIV-1 infection. Inhibits HIV-1 entry in cells coexpressing CD4 and APJ. Apelin-36 has a greater inhibitory activity on HIV infection than other synthetic peptides. It also stimulates milk ejection and the milk ejection reflex in neonates. It also stimulates body fluid homeostasis and behavior.	Xq25	Expressed in the brain with highest levels in the frontal cortex, thalamus, hypothalamus and midbrain. Secreted by the mammary gland into the colostrum and the milk.	Secreted.
52				165	The Klotho peptide generated by cleavage of the membrane-bound isoform may be an anti-aging circulating hormone which would extend life span by inhibiting insulin/IGF1 signaling (By similarity).	13q12	placenta, small intestine and prostate. Down regulated in renal cell carcinomas, hepatocellular carcinomas, and in chronic renal failure kidney.	Klotho peptide: Secreted.
7	Q9UBU3	GHRL_HUMAN	GHRL	51738	Obestatin may be the ligand for GPR39. May have an appetite-reducing effect resulting in decreased food intake. May reduce gastric emptying activity and jejunal motility (By similarity).	3p26-p25	Highest level in stomach. All forms are found in serum as well. Other tissues compensate for the loss of ghrelin synthesis in the stomach following gastrectomy.	Secreted.

ProteINSIDE synthétise automatiquement les informations disponibles pour un gène /une protéine dans les bases de données NCBI et UniProtKB

NCBI - Coordinators NR 2013
UniProtKB - Magrane et al. 2011

Exemple des protéines :

- Glycolyse
- chaîne respiratoire
- hormone



Dictionnaire hiérarchisé de termes GO

ProteINSIDE recherche pour des protéines ou des gènes les termes de la « Gene Ontology (GO) » (on parle d'annotations).



Module de Gene Ontology

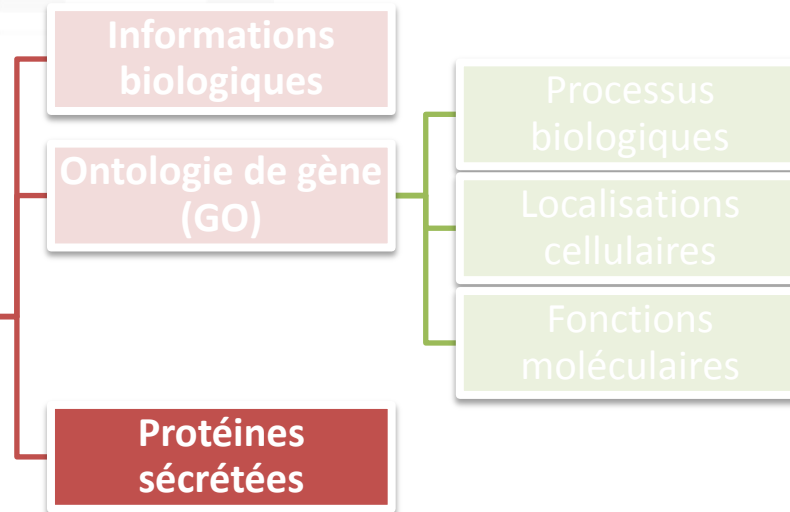
- Appel à la BDD pour annoter les protéines
- Calcul de la représentativité des fonctions biologique
- Calcul des probabilité qu'une protéine agisse bien dans une fonction biologique
- Calcul du lien entre les fonctions
- ...

Multiple annotations GOs

No of annotated protein by GO		Function	Proteins	Gene Name	Number of items	Gene products detected on		
7.0 63.0					10	HUMAN		
GO	Function	Proteins	Gene Name	Ontology group	Count	GO frequency (%)	Number of Gene products detected / expected	Percent of Gene products detected / expected
GO_0006096	glycolysis	Q01813 Q16877 P09104 P16118 P00338 P00558 P15259 P08237 P04406 P14618 P30613 P07205 P18669 P52789 P60174 P36871 Q60825 P04075 P19367 Q75356 Q16875 P09972 P06733 Q14556 P13929 P05062 P17858 P06744	PFKP PFKFB4 ENO2 PFKFB1 LDHA PGK1 PGAM2 PFKM GAPDH PKM PKLR PGK2 PGAM1 HK2 TP11 PGM1 PFKFB2 ALDOA HK1 ENTPD5 PFKFB3 ALDOC ENO1 GAPDHS ENO3 ALDOB PFKL GPI	BP	28	22.764	28 / 45	62.222
GO_0005184	neuropeptide hormone activity	P15089 P10092 P01282 P06850 P11509 P01185 P22466 P06307	UCN CALCB VIP CRH ADCYAP1 AVP GAL CCK	MF	8	6.504	8 / 15	53.333
GO_0005179	hormone activity	P11148 P06850 P01298 P06307 P81277 P09529 P12272 P01242 P01282 P61278 Q9Y581 P04808 P01350 Q14406 Q95399 P01225 Q43555 P01189 P10082 P09681 Q15848 P01308 P09683 P01233 P52823 P08476 P01270 P05111 P16860 Q76061 P01215 P01222	GHRH1 CRH PPY CCK PRLH INHBB PTHLH GH2 VIP SST INSL6 RLN1 GAST CSHL1 UTS2 FSHB GNRH2 POMC PPY GIP ADIPOQ INS SCT CGB8 STC1 INHBA PTH INHA NPPB STC2 CGA TSHB	MF	32	26.016	32 / 62	51.613

Fonctions dans lesquelles les protéines vont entrer en jeu (activation / inactivation de réactions biologiques)

QuickGO - Binns et al. 2009
GO.org - Ashburner et al. 2000



Exemple des protéines :

- Glycolyse
- chaîne respiratoire
- hormone



ProteINSIDE prédit la sécrétion d'une protéine en :

- Identifiant un peptide signal sur la séquence d'acides aminés
- Recherchant des termes GO liés à la localisation cellulaire

Module de Protéines sécrétée

- Alignement de séquences
- Calcul de la probabilité d'apparition d'un peptide signal
- Vérification du peptide signal
- Calcul de la probabilité que la protéine soit bien sécrété par rapport à son background
- Comparaison avec les fonction biologique GO

Proteins	GO related to secretion	Gene Name	Number of rows	Signal peptides detected	Download table
			10	85 (on 133 proteins imported)	

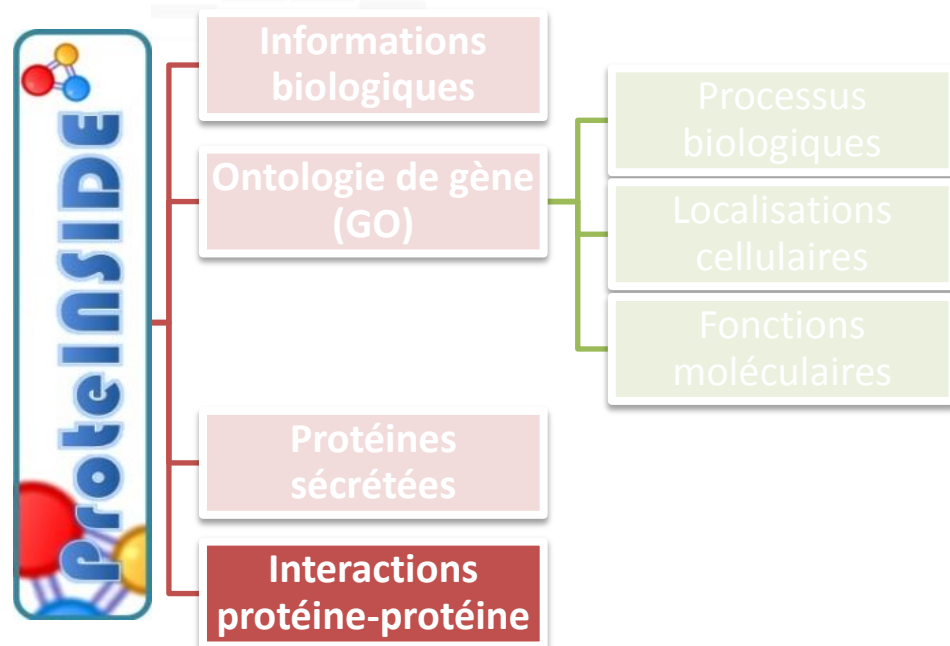
Proteins	Protein ID	Gene Name	Peptide	GO related to secretion	Number of GO
Q9UBU3	GHRL_HUMAN	GHRL	noTM	GO:0043400 GO:0032024 GO:0051461 GO:0030252 GO:0005615 GO:0034774 GO:0005576 GO:0060124 GO:0051464	9
P01308	INS_HUMAN	INS	noTM	GO:0050708 GO:0090277 GO:0005615 GO:0034774 GO:0050715 GO:0005576 GO:0050796	7
Q15848	ADIPO_HUMAN	ADIPOQ	noTM	GO:0034383 GO:0045715 GO:0005615 GO:0005576	4
P01189	COLI_HUMAN	POMC	noTM	GO:0030141 GO:0005615 GO:0034774 GO:0005576	4
P08476	INHBA_HUMAN	INHBA	noTM	GO:0046880 GO:0042701 GO:0005576 GO:0046881	4
P16860	ANFB_HUMAN	NPPB	noTM	GO:0007589 GO:0005615 GO:0005576	3
P06850	CRF_HUMAN	CRH	noTM	GO:0005615 GO:0005576 GO:0051464	3
P09681	GIP_HUMAN	GIP	noTM	GO:0034774 GO:0005576 GO:0050796	3
P01275	GLUC_HUMAN	GCG	noTM	GO:0034774 GO:0005576 GO:0050796	3
Q9UEF7	KLOT_HUMAN	KL	noTM	GO:0005615 GO:0005576 GO:0070062	3

prev next 1 2 3 4 5 6 7 8 9

Protéines potentiellement sécrété et GO liés à la sécrétion des protéines

ProteINSIDE prédit la sécrétion d'une protéine en utilisant l'algorithme SignalP et les termes GO reliés à la sécrétion

SignalP (v 4.1) - Petersen et al. 2011



Exemple des protéines :

- Glycolyse
- chaîne respiratoire
- hormone

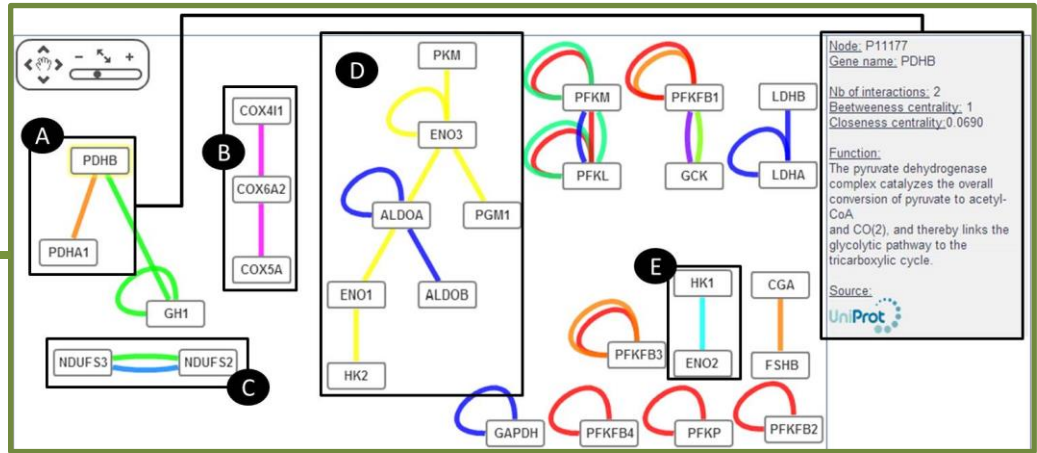


L'analyse de réseaux d'interactions (interactome) pour identifier la fonction et/ou le rôle central de protéines dans des mécanismes biologiques

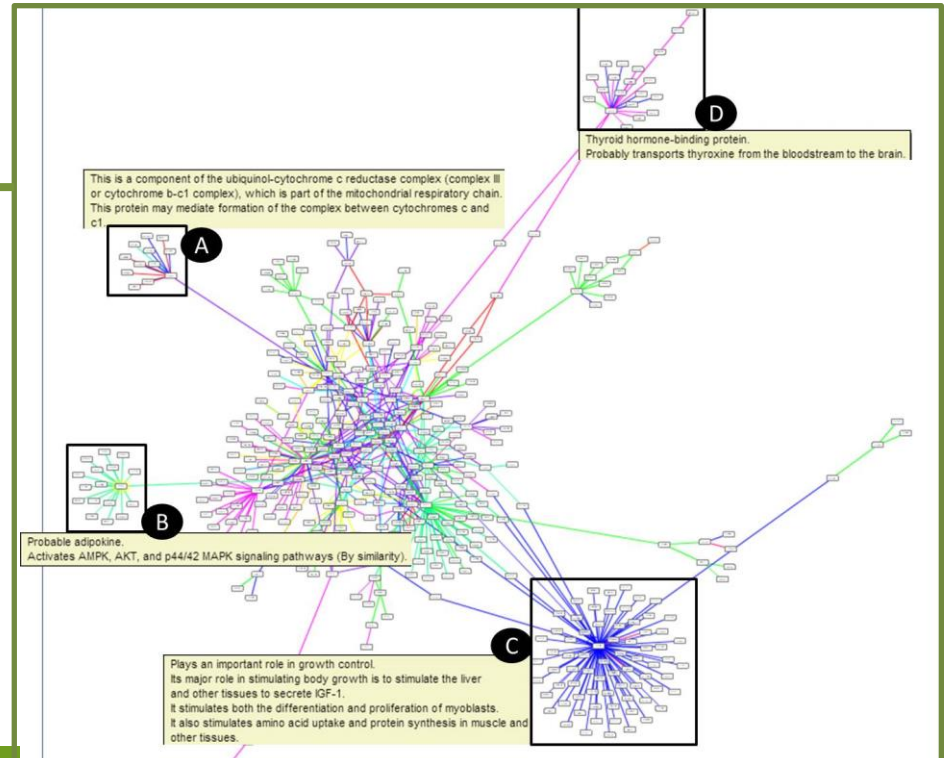


ProteINSIDE

Entre protéines du
jeu de données
(36 interactions)



Entre protéines du jeu de
données et extérieures
(616 interactions)

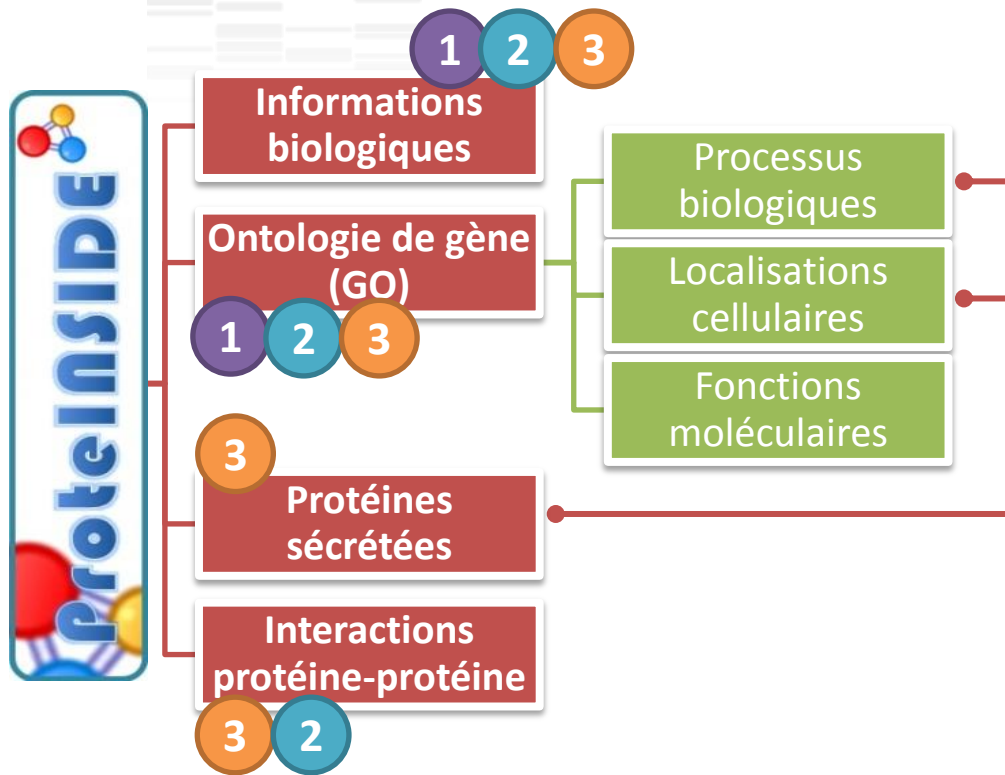


ProteINSIDE réalise des réseaux de
protéines pour mettre en évidence
des mécanismes biologiques

PsicQuick - Aranda et al. 2011
Cytoscape web - Lopes et al. 2010



Application



Interactions TA / Muscle

- 1 **Type et abondance des cellules progénitrices ?**
- 2 **Spécificités métaboliques des tissus ?**
- 3 **Relations endocrines entre les tissus ?**

ProteINSIDE sera utilisé pour identifier des gènes et des protéines spécifiques de cellules musculaires ou adipeuses, de voies métaboliques, des adipokines ou des myokines...



Application

Milliers de protéines / gènes à analyser



Manuellement

- Recherche d'identifiants biologique
- Recherche de fonctions biologique
- Rassemblement et trie des résultats
- ...

Plusieurs mois d'étude



Automatiquement



2 min d'analyse pour 1000 protéines

1 jour d'analyse pour génome bovin (33 000 protéines)

www.proteinside.org