



Update of ProteINSIDE: an online tool for proteome data mining

Jérémy Tournayre, Céline Boby, Matthieu Reichstadt, Muriel Bonnet

► **To cite this version:**

Jérémy Tournayre, Céline Boby, Matthieu Reichstadt, Muriel Bonnet. Update of ProteINSIDE: an online tool for proteome data mining. JOBIM, Jul 2021, Paris - (Congrès virtuel), France. hal-03283757

HAL Id: hal-03283757

<https://hal.inrae.fr/hal-03283757>

Submitted on 12 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Update of ProteINSIDE: an online tool for proteome data mining

Jérémy Tournayre¹, Céline Boby¹, Matthieu Reichstadt¹ and Muriel Bonnet¹

¹ INRAE, Université Clermont Auvergne, Vetagro Sup, UMRH, 63122, Saint-Genès-Champanelle, France

Corresponding Authors: jeremy.tournayre@inrae.fr and muriel.bonnet@inrae.fr

1. Introduction

We previously presented the ProteINSIDE tool to the scientific community [1,2]. ProteINSIDE was developed to mine results from large lists of proteins or genes, and thus to extract meaningful biological knowledge from “omics” datasets. The first version of ProteINSIDE gave 4 types of analyses or results: (1) identifiers conversion plus an overview of the biological information stored in public databases (NCBI and UniProt), (2) Gene Ontology enrichment analysis, (3) proteins that are predicted as secreted by mammalian cells, (4) protein protein interactions. Since then, we have improved this tool on many points, including an increase in the number of organisms considered, new functional enrichments (in addition to the GO ones) and the search for quantitative trait loci.

This new version of ProteINSIDE is available at the following address:

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

2. Improvements

The workflow now uses the g:profiler API (Application Programming Interface) [3] for the conversion module (g:convert) and enrichment module (g:GOS). The second version of ProteINSIDE analyzed lists of identifiers from more than 600 organisms rather than those from the 6 species previously targeted by the first version. Functional enrichment analysis previously focused on GO is now complemented by functional enriched analysis to find over-representation of information from several databases: (Human Proteome Atlas, Human Phenotype Ontology, Kegg, miRTarBase, Transfac, Reactome and WikiPathways).

A new functionality for protein protein interactions was added: the comparison between multiple lists experiments. For this, a user sends several lists of proteins. Then, ProteINSIDE, with the help of Psicquic [4], searches for any interactions between the proteins in these lists. The results are downloadable and can also be displayed on a network.

A complete overhaul of the tool was carried out in order to have a simplest interface and best optimization, in particular by parallelizing the calculations. Now the modules are independent of each other to allow the user to choose which one to launch, in order to access to the results as soon as possible without waiting for all the modules to finish running. For the visualization of networks, cytoscape web Flash was no longer supported and has been replaced by its Javascript version [5].

Finally, we added a fifth module, to search for quantitative trait loci using AnimalQTLdb [6] for the bovine species only.

References

1. Kaspric N, Picard B, Reichstadt M, Tournayre J, Bonnet M. ProteINSIDE to Easily Investigate Proteomics Data from Ruminants: Application to Mine Proteome of Adipose and Muscle Tissues in Bovine Foetuses. PLOS ONE. 22 mai 2015;10(5):e0128086.
2. Kaspric N, Reichstadt M, Picard B, Tournayre J, Bonnet M. Protein function easily investigated by genomics data mining using the proteINSIDE online tool. Genomics Comput Biol. 2015;1(1):e16.
3. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. g:Profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 2 juill 2019;47(W1):W191- 8.
4. Aranda B, Blankenburg H, Kerrien S, Brinkman FSL, Ceol A, Chautard E, et al. PSICQUIC and PSIScore: accessing and scoring molecular interactions. Nat Methods. juill 2011;8(7):528- 9.
5. Franz M, Lopes CT, Huck G, Dong Y, Sumer O, Bader GD. Cytoscape.js: a graph theory library for visualisation and analysis. Bioinformatics. 15 janv 2016;32(2):309- 11.

6. Hu Z-L, Park CA, Reecy JM. Building a livestock genetic and genomic information knowledgebase through integrative developments of Animal QTLdb and CorrDB. *Nucleic Acids Res.* 8 janv 2019;47(D1):D701- 10.