

Proteome data mining using ProteINSIDE online tool

Nicolas Kaspric, Matthieu Matthieu.Reichstadt@inrae.Fr Reichstadt, Brigitte B. Picard, Jérémy Tournayre, Muriel Bonnet

▶ **To cite this version:**

Nicolas Kaspric, Matthieu Matthieu.Reichstadt@inrae.Fr Reichstadt, Brigitte B. Picard, Jérémy Tournayre, Muriel Bonnet. Proteome data mining using ProteINSIDE online tool. JOBIM 2015 - Journées Ouvertes Biologie Informatique Mathématiques, Jul 2015, Clermont-Ferrand, France. 1 p. hal-02740479

HAL Id: hal-02740479
https://hal.inrae.fr/hal-02740479

Submitted on 2 Jun 2020

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.

Proteome data mining using ProteINSIDE online tool

Nicolas KASPRIC[1,2], Matthieu REICHSTADT[1,2], Brigitte PICARD[1,2], Jérémy TOURNAYRE[1,2] and Muriel BONNET[1,2]

[1] INRA, UMR1213 Herbivores, F-63122 Saint-Genès-Champanelle, France

[2] Clermont Université, VetAgro Sup, UMR1213 Herbivores, F-63000, Clermont-Ferrand, France

Corresponding authors: nicolas.kaspric@clermont.inra.fr and muriel.bonnet@clermont.inra.fr

Summary

Genomic and proteomic experiments are widely acknowledged to produce a huge amount of data to be analyzed. The challenge for scientist working on ruminant species is to extract meaningful biological context for proteins or genes to generate new hypothesis. This challenge is currently difficult because of the lack of an integrative workflow that hinders the efficiency and the robustness of data mining, thus scientists must use several tools often not designed for ruminant studies. We designed the online tool ProteINSIDE [1] that provides an overview of the biological information stored in public databases (NCBI and UniProt) or provided by annotations according to the Gene Ontology. It also predicts proteins that are secreted to search for proteins that mediate signalization between cells or tissues, and analyses protein-protein interactions to identify proteins contributing to a process or to visualize functional pathways. Using lists of proteins or genes as a unique input, ProteINSIDE is an original all-in-one tool to mine genomic and proteomic data from cattle, sheep, goat, human, rat, and mouse. A user-friendly web interface provides access to create analyzes and view results that can be sorted online or directly downloaded as several file formats (image, pdf, tabular, spreadsheet, fasta…). ProteINSIDE's database collects and stores the biological information required to the efficient functioning of ProteINSIDE and is monthly updated to gather the last available knowledge. ProteINSIDE was successfully bench tested with 1000 protein identifiers from each species by comparison with DAVID, BioMyn, and AgBase designed for information retrieval and annotation, as well as with PrediSi, and Phobius that predict secreted proteins. ProteINSIDE is freely available using a simple internet browser at www.proteinside.org (an example of results is provided on the home page).

Résumé

Les expérimentations de génomique et de protéomique produisent de grandes quantités de données à analyser. Le défi pour les scientifiques travaillant sur les ruminants est d'extraire de ces expérimentations une information biologique pertinente afin de générer de nouvelles hypothèses de recherche. Actuellement, l'absence d'outils automatisés dédiés aux ruminants entrave l'exploration efficace et robuste de données disponibles. Les scientifiques utilisent donc différents outils qui ne sont pas forcément conçus pour l'analyse des données issues des ruminants. Nous avons conçu l'outil en ligne ProteINSIDE [1] qui synthétise l'information biologique stockée dans les bases de données publiques (NCBI et UniProt) ou fournie par les annotations fonctionnelles issues de l'ontologie des gènes. Il prédit aussi les protéines qui sont sécrétées et qui interviennent dans la signalisation entre les cellules ou tissus, et identifie les interactions entre protéines pour identifier et visualiser celles qui contribuent à un même processus biologique. ProteINSIDE est un outil tout-en-un pour la fouille des données de génomique et de protéomique chez le bovin, l'ovin, le caprin, l'homme, le rat et la souris. L'interface web permet la création d'analyses et la visualisation des résultats en temps réel ou leurs téléchargement dans de nombreux formats (images, pdf, tableaux, données brutes, fasta...). La base de données de ProteINSIDE recueille et stocke l'information biologique nécessaire pour un fonctionnement efficace. Elle est mise à jour mensuellement afin de recueillir les dernières connaissances. ProteINSIDE a été testé avec des jeux de données de 1000 protéines par espèce et a été comparé avec succès à DAVID, BioMyn et AgBase conçus pour la recherche d'information et l'annotation, ainsi qu'à PrediSi et Phobius qui prédisent les protéines sécrétées. ProteINSIDE est disponible gratuitement en utilisant un simple navigateur internet à l'adresse www.proteinside.org (un exemple de résultats est visible sur la page d'accueil).

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* 2015, **10:**e0128086.