Online Repositories

In the previous chapters of the tutorial, we identified proteins using curated database sequences and enriched our results with knowledge from external resources. In order to allow the community to benefit from your results in turn, online repositories are available to enable the exchange of data. Moreover, making the data public is now required by most journals prior to publication.

**Proteomics Results**

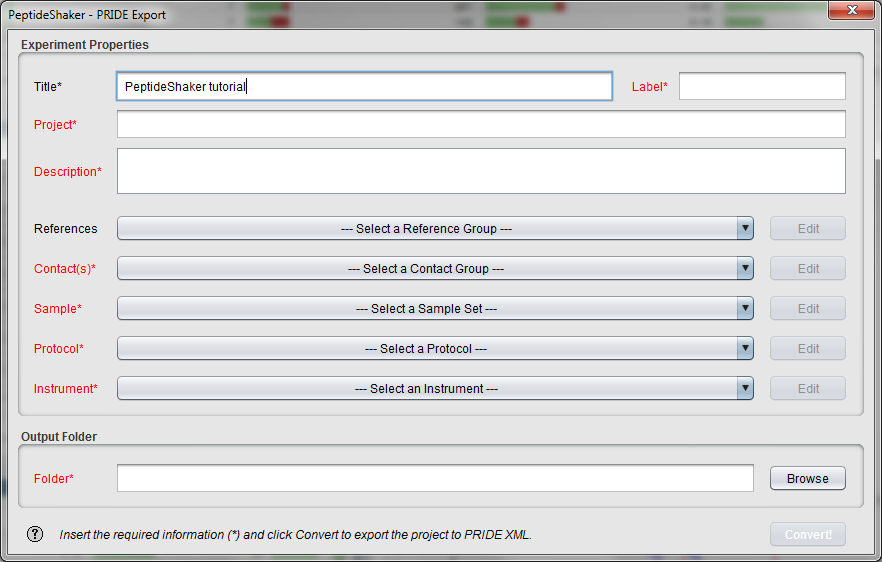
The proteomics identifications database[1](#_ENREF_1) (PRIDE, <http://www.ebi.ac.uk/pride>) and ProteomeXchange (<http://www.proteomexchange.org>) are the repositories of choice for protein identification data. PRIDE stores proteomics identification results while ProteomeXchange allows you to store other information as well. For both of them, the identification results should be converted into the PRIDE XML format. In this section, we will look at how to load a PeptideShaker project into ProteomeXchange and how to visualize PRIDE data using PRIDE Inspector[2](#_ENREF_2).

1. Submitting Data to PRIDE

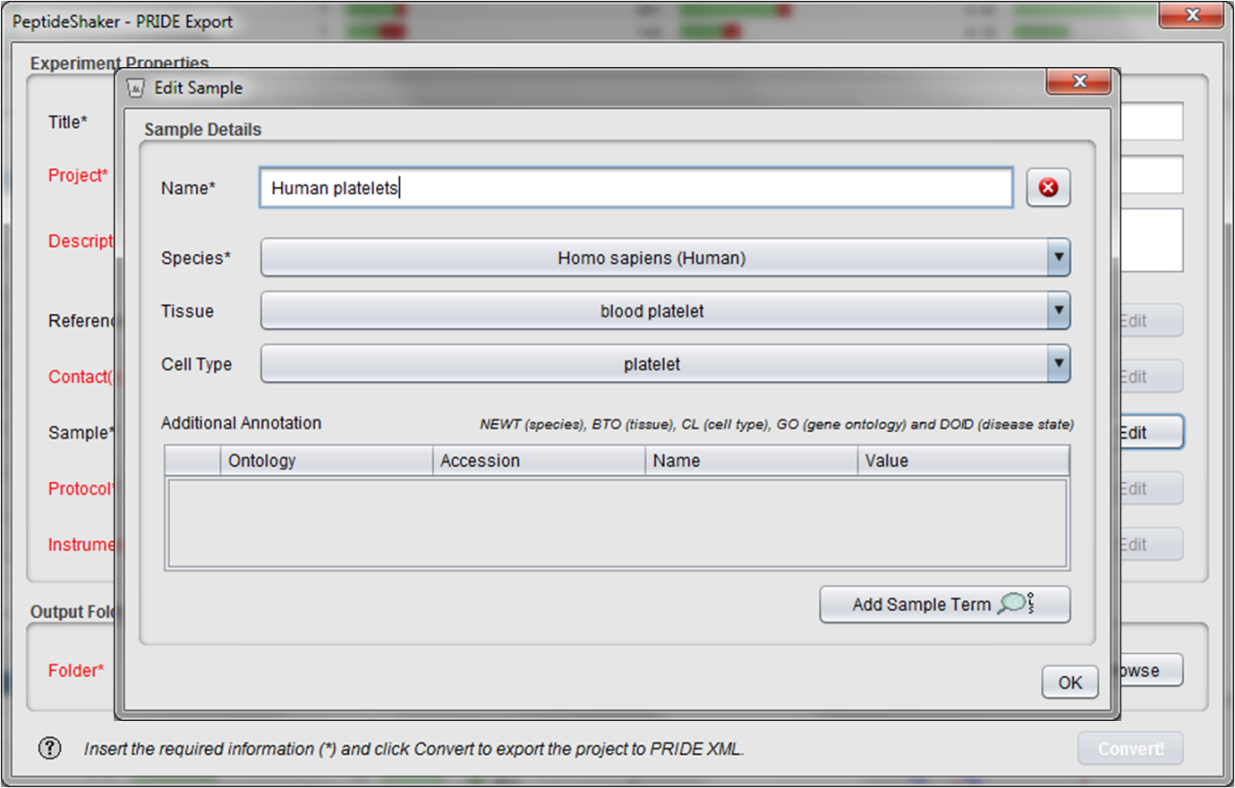
In this section, we will generate a PRIDE XML file from the previously used human dataset. Load the project human.cps located in the resources folder into PeptideShaker:



Now go to the ‘Export’ menu at the top and select ‘PRIDE XML’. The following dialog appears:



The information needed here will be used to reference your dataset in PRIDE. Using the respective fields, create a contact, a sample, a protocol and an instrument for our dataset. Note that all terms are standardized, creating a Human Platelet Sample as detailed below will thus help other platelet interested scientist to find your results straightforwardly.

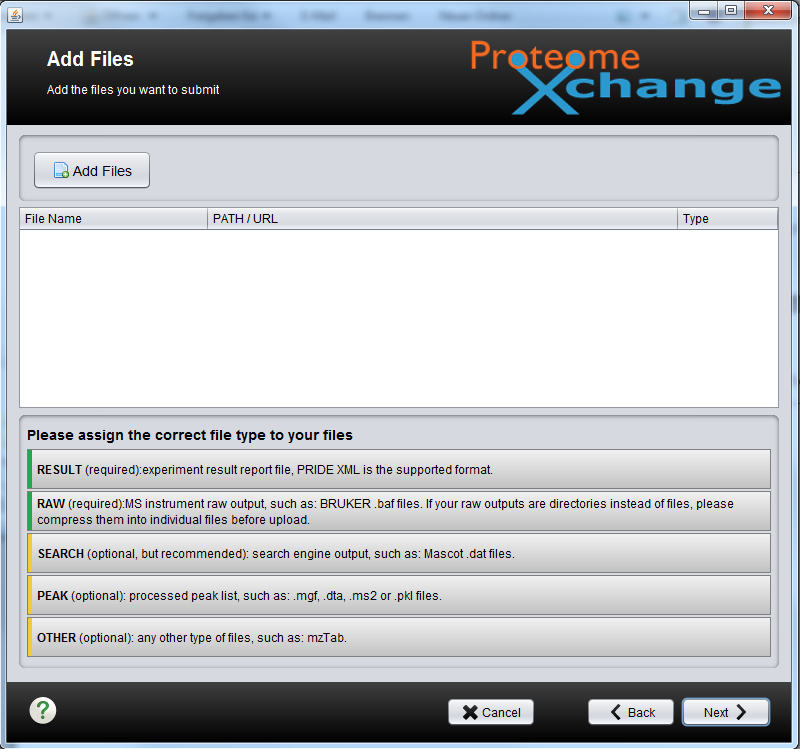


Selecting an output folder and clicking on the 'Convert!' button will start the creation of the PRIDE XML file. In order to save time, the corresponding file has already generated for you and is located in the resources folder.

Start the ProteomeXchange submission tool located in the software folder (PX\_Submission). You should see the following screen:



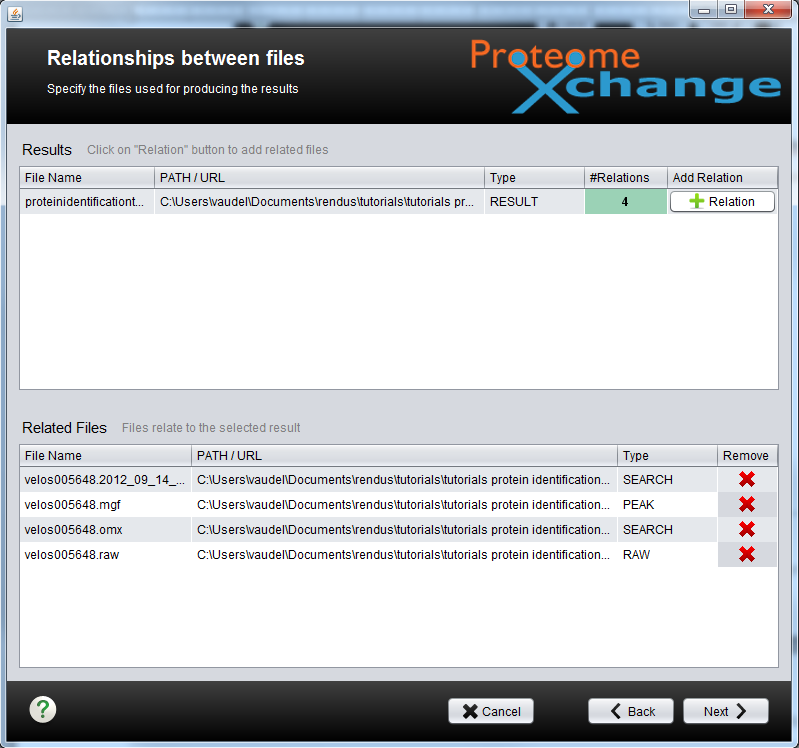
Click on ‘Complete Submission’ then ‘Next’. You will see the following screen:



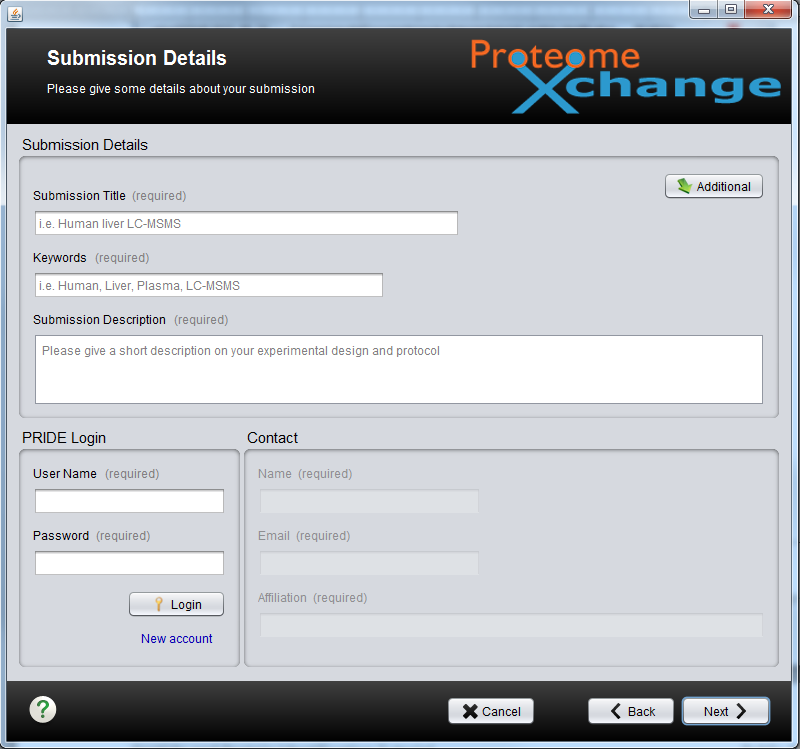
Here, load the Raw file, the pride xml file, the search result files (from OMSSA and X!Tandem) and the peak list. All these files are located in the resources folder. Note that the submission tool recognizes the different file formats, except for the X!Tandem result file.

In the last step, we will indicate that all the intermediate files led to the same pride xml result. Add a relation between all files to the pride xml file.

*In which case will you have different PRIDE XML files with different relations?*



In the next and last step, you will reference your experiment and upload it in PRIDE.

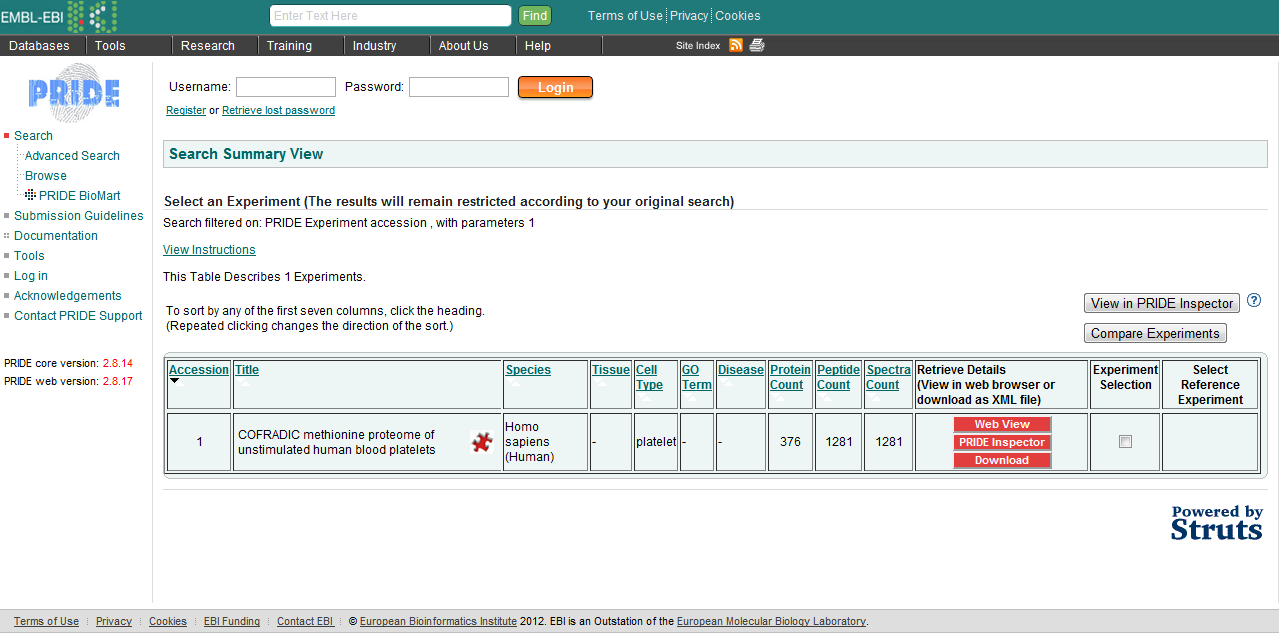


In order to upload a file, you need a PRIDE login which the PRIDE team will provide you on demand. Your dataset will stay private during the review process and a reviewer account will be established so that a reviewer can access your data. The credentials for these accounts should be made available in your manuscript. Once your paper accepted, the data will be made freely accessed to anyone.

Obviously we will not upload the tutorial data online, however, if you do so, all identification results will be available and can be browsed as demonstrated in the following.

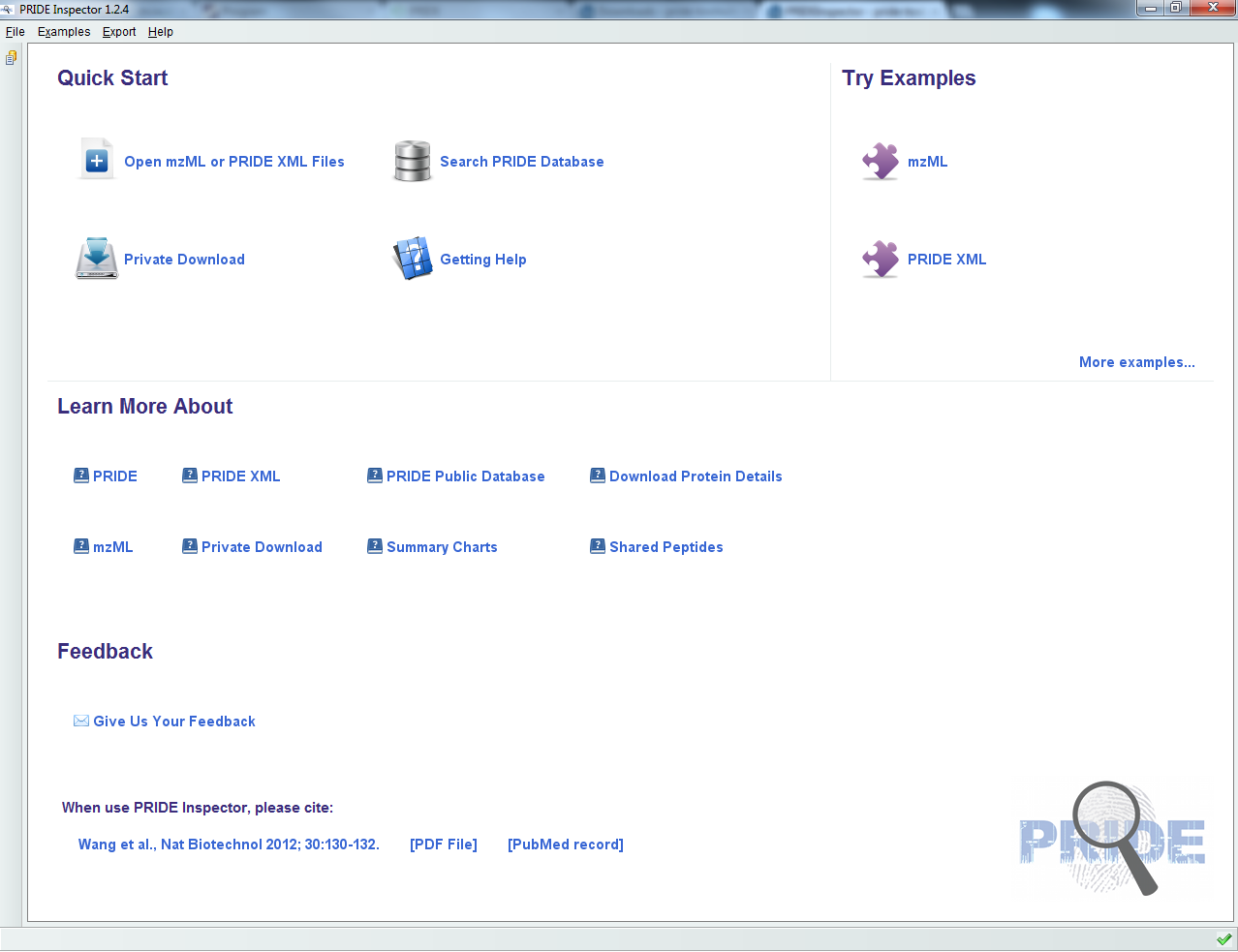
1. Browse PRIDE

Go on the PRIDE website (<http://www.ebi.ac.uk/pride>) and search PRIDE project number 1, you should see the following screen:

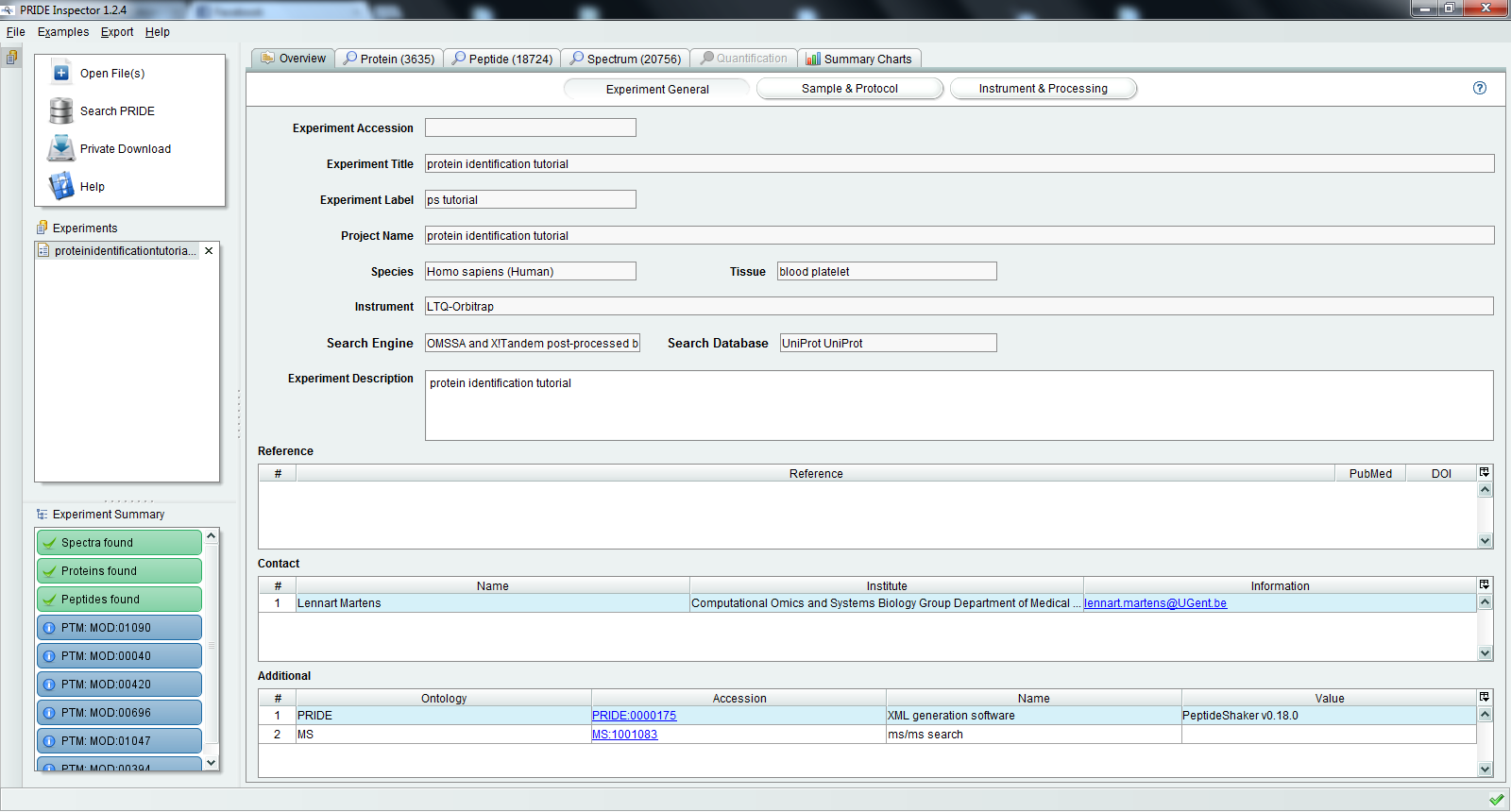


Click on ‘Web View’ to browse this project. *What information can you access on this experiment? What is your opinion about making all data available?*

It is possible to browse all online PRIDE datasets using PRIDE Inspector (<http://pride-toolsuite.googlecode.com>), available in the software folder. Starting PRIDE Inspector, you should see the following:



Select ‘Open mzML or PRIDE XML Files’ and open proteinidentificationtutorial.xml located in the resources folder. You should see the following:



Note that all spectrum annotation (modifications, ions, etc.) have been passed by PeptideShaker to PRIDE Inspector as standardized terms and will thus be available for all other online resources. *What difference do you see compared to the PeptideShaker results?*

References

(1) Martens, L.; Hermjakob, H.; Jones, P.; Adamski, M.; Taylor, C.; States, D.; Gevaert, K.; Vandekerckhove, J.; Apweiler, R. PRIDE: the proteomics identifications database. *Proteomics* **2005**, *5*, 3537.

(2) Wang, R.; Fabregat, A.; Rios, D.; Ovelleiro, D.; Foster, J. M.; Cote, R. G.; Griss, J.; Csordas, A.; Perez-Riverol, Y.; Reisinger, F.; Hermjakob, H.; Martens, L.; Vizcaino, J. A. PRIDE Inspector: a tool to visualize and validate MS proteomics data. *Nat Biotechnol* **2012**, *30*, 135.