Interactive Statistical Resource for Proteomic Analyses

Bobbie-Jo Webb-Robertson, Lisa Bramer, Markus Kobold, Jeffrey Jensen, Kelly Stratton, Mathew Thomas, Amanda White

**Pacific Northwest**
NATIONAL LABORATORY

Proudly Operated by **Battelle** *Since 1965*

P·MART

http://pmart.azurewebsites.net/pmart/index.html

ABSTRACT

One of the major challenges to the discovery of robust biomarker candidates from global shotgun mass spectrometry (MS)-based proteomic experiments is processing these large and complex datasets in a manner that accounts for the underlying statistical hypothesis of interest. We believe for the field of clinical proteomics to successfully identify new prognostics or diagnostics of disease or response to therapies requires not only high quality data with respect to the instrument, but also high quality statistical analysis of the data. This project proposes new informatics technology in the form of a robust, interactive web-based software environment (P-Mart) that will enable biomedical and biological scientists to perform in-depth analyses of global proteomics data from the point of quality assessment and normalization of raw inferred abundances (e.g., peak area) to the identification of protein biomarkers.

P-MART CAPABILITIES

- Datasets collected and released by the Clinical Proteomics Tumor Analysis Consortium (CPTAC) available through https://cptac-data-portal.georgetown.edu/cptacPublic/ at the peptide, gene or protein level. P-Mart allows for the selection of parameter of interest for evaluation (e.g., clinical stage, tumor residual disease) – Figure 1.
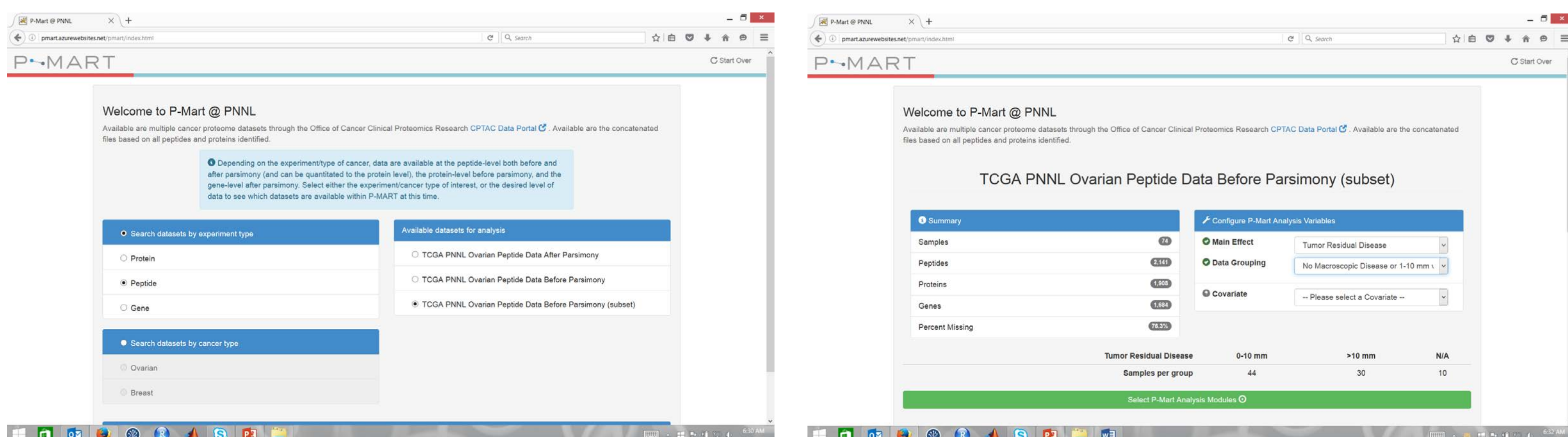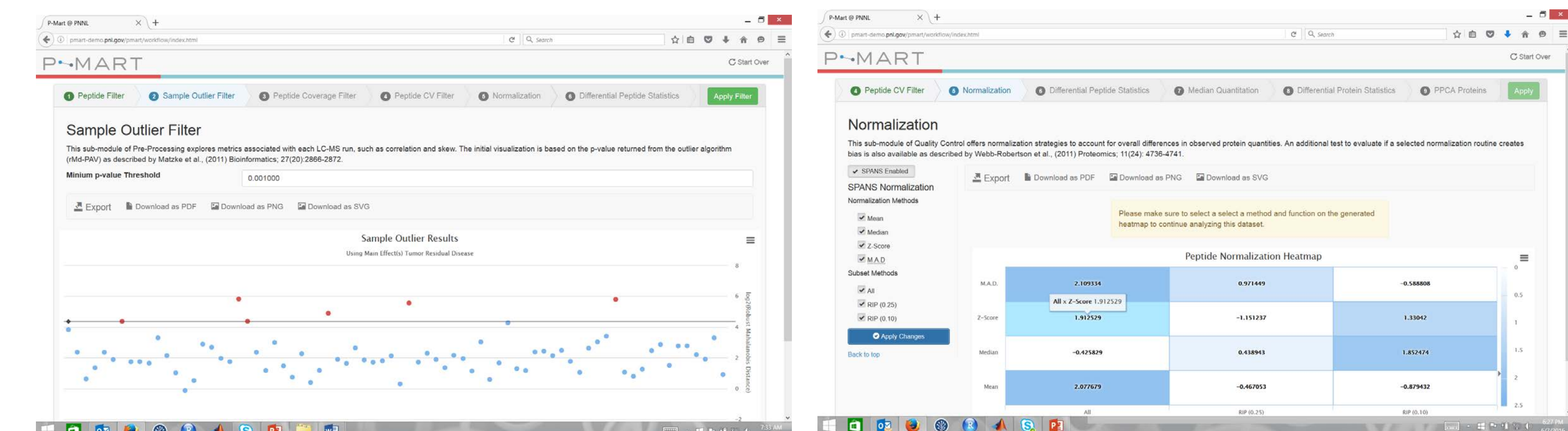

Figure 1: Screen shot of the opening page of P-Mart where the user can select data for evaluation by either data type of cancer type (left). Once the dataset is selected associated meta-data is observed and clinical parameter of interest is selected (right).

- Modules for various analyses are available and user can select workflow or have P-Mart select based on data type – Figure 2.
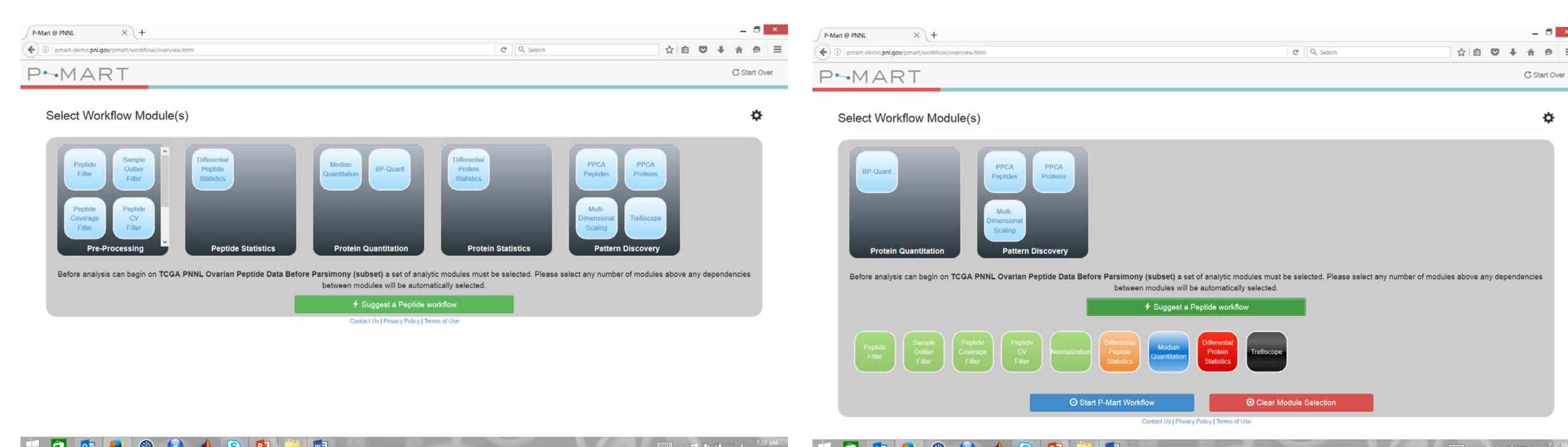
Figure 2: P-Mart is based on customizable workflows. Each capability (left figure) can be added or removed to the workflow (right figure) and checks are implemented that assure specific requirements of modules are met. The user may also let the software propose a workflow.

- A series of quality control processes are available. Figure 3 shows both sample outlier sample identification [1] and normalization [2] as a visualization that enables users to make statistical-based decisions about the data quality control processing.


Figure 3: P-Mart uses visualization to express statistical analyses to the user. The plot on the left evaluates each sample as a potential outlier based on its data distribution. The user can modify the stringency of the test as appropriate to the analysis of interest. The plot on the right evaluates a large number of normalizations with a heatmap that indicates the most robust approaches as lighter colors.

- A visualization tool called Trelliscope [3] is utilized to sort through the large number of potential biomarker candidates. Figure 4 shows Trelliscope being used to sort proteins by coverage and statistical p-value.

- Figure 5 shows the first plot based on this criteria of 996 plots. Clicking on the link brings you to the NCBI Protein database.
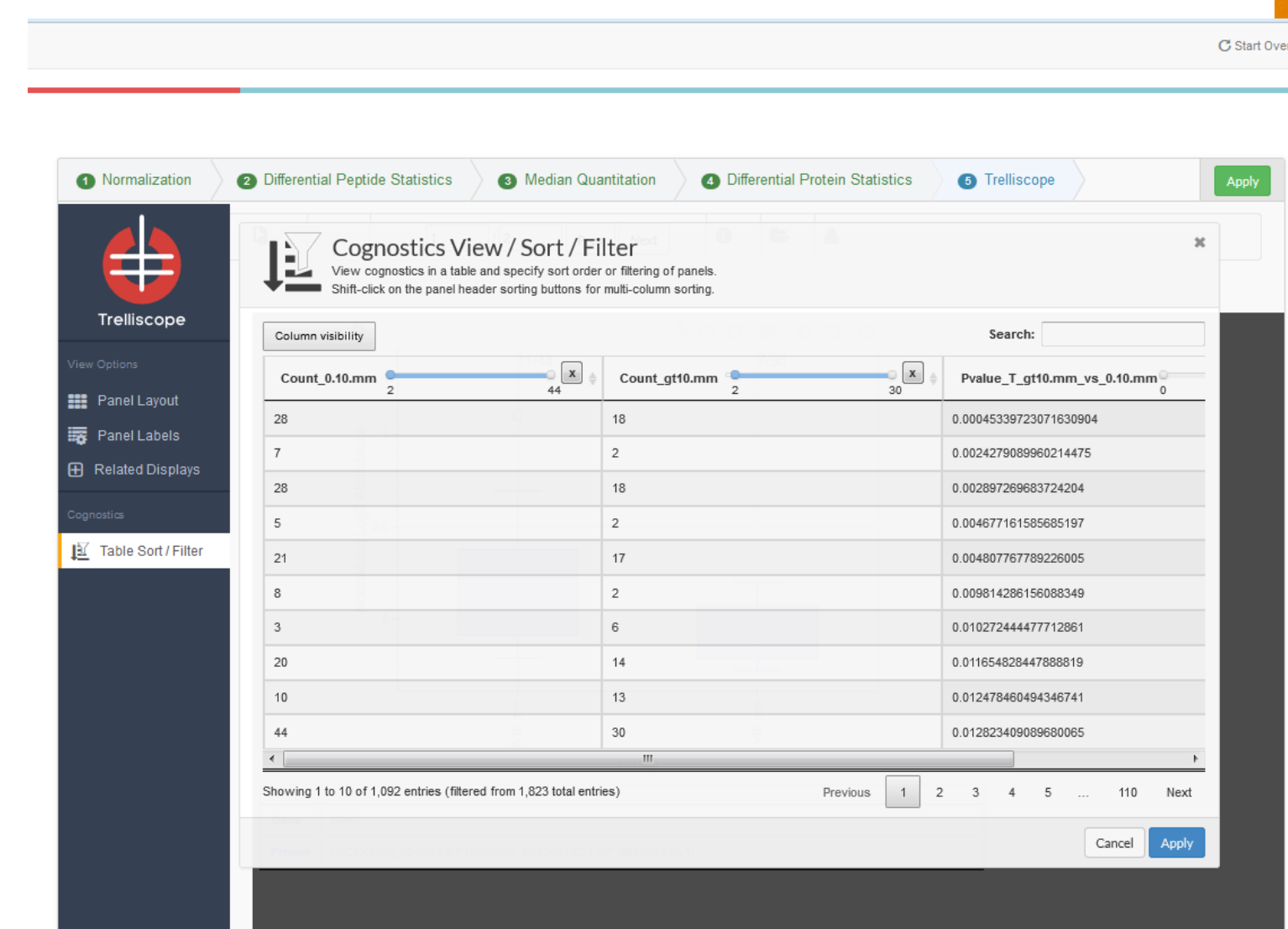

Figure 4: The table-sort capability in Trelliscope to visualize proteins based on defined criteria.
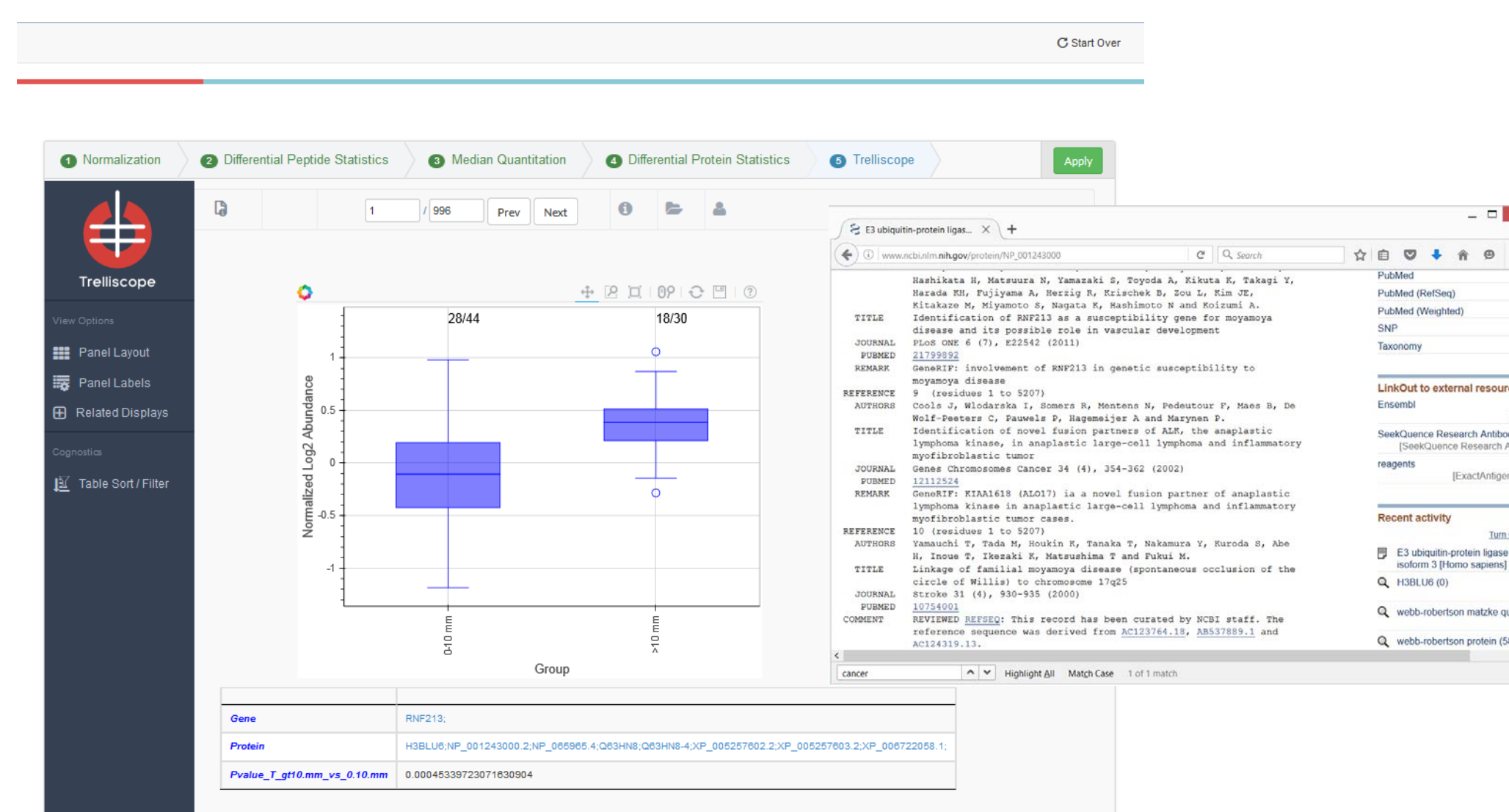
Figure 5: The first gene identified by the table-sort is H3BLU6. The link from the protein identifier brings up a protein associated with cancer based on publications cited.

- Trelliscope can also be used to search for genes of interest (e.g., BRAF) and the visualization of the data associated with the proteomics data can be visualized along with meta-data, such as p-value – Figure 6.
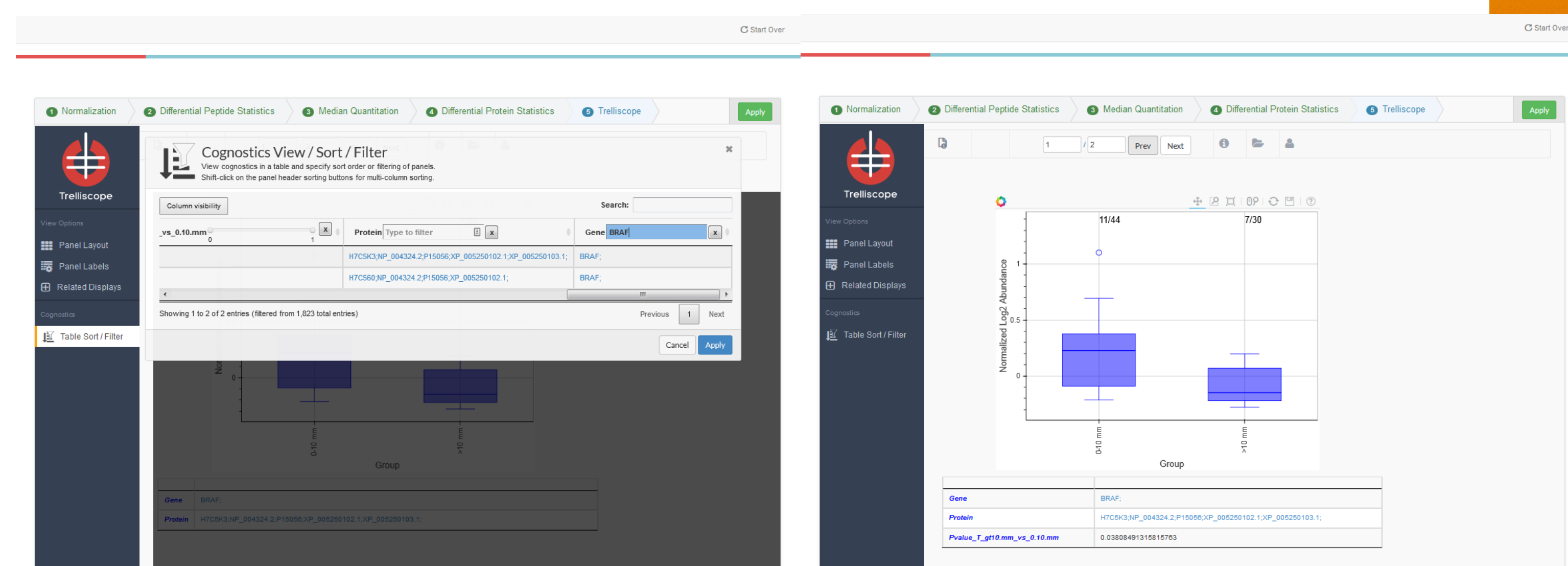
Figure 6: The cognostics view is used to search for a protein of interest to ovarian cancer (BRAF) to easily view the resulting data in a boxplot on the right.

ABOUT

Pacific Northwest National Laboratory

The Pacific Northwest National Laboratory, located in southeastern Washington State, is a U.S. Department of Energy Office of Science laboratory that solves complex problems in energy, national security, and the environment, and advances scientific frontiers in the chemical, biological, materials, environmental, and computational sciences. The Laboratory employs nearly 5,000 staff members, has an annual budget in excess of $1 billion, and has been managed by Ohio-based Battelle since 1965.

For more information on the science you see here, please contact:
Bobbie-Jo Webb-Robertson
bj@pnnl.gov

REFERENCES

[1] Matzke et al., (2011) Bioinformatics. 27(20): 2866-72.
[2] Webb-Robertson et al., (2011) Proteomics. 11(24): 4736-41.
[3] Hafen et al., (2013) LDAV http://tessera.io/#

U.S. DEPARTMENT OF
ENERGY

www.pnnl.gov