Journal of proteome research

Technical Note

pubs.acs.org/jpr

Terms of Use

Panorama: A Targeted Proteomics Knowledge Base

Vagisha Sharma,[†] Josh Eckels,[‡] Greg K. Taylor,[‡] Nicholas J. Shulman,[†] Andrew B. Stergachis,[†] Shannon A. Joyner,[§] Ping Yan,[∥] Jeffrey R. Whiteaker,[∥] Goran N. Halusa,[⊥] Birgit Schilling,[#] Bradford W. Gibson,[#] Christopher M. Colangelo,[○] Amanda G. Paulovich,[∥] Steven A. Carr,[▽] Jacob D. Jaffe,[▽] Michael J. MacCoss,[†] and Brendan MacLean*[,†]

[†]University of Washington, Seattle, Washington 98195, United States

[‡]LabKey Software, San Diego, California 92101, United States

[§]Carnegie Melon University, Pittsburgh, Pennsylvania 15213, United States

[∥]Fred Hutchinson Cancer Research Center, Seattle, Washington 98109, United States

[⊥]Leidos Biomedical Research, Inc., Frederick, Maryland 21702, United States

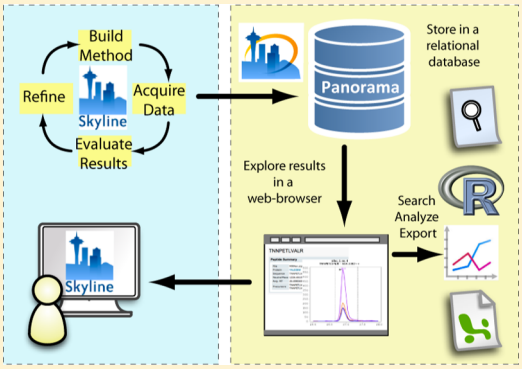
[#]Buck Institute for Research on Aging, Novato, California 94945, United States

[○]Yale University, New Haven, Connecticut 06520, United States

[▽]The Broad Institute, Cambridge, Massachusetts 02142, United States

Supporting Information

ABSTRACT: Panorama is a web application for storing, sharing, analyzing, and reusing targeted assays created and refined with Skyline,[1] an increasingly popular Windows client software tool for targeted proteomics experiments. Panorama allows laboratories to store and organize curated results contained in Skyline documents with fine-grained permissions, which facilitates distributed collaboration and secure sharing of published and unpublished data via a web-browser interface. It is fully integrated with the Skyline workflow and supports publishing a document directly to a Panorama server from the Skyline user interface. Panorama captures the complete Skyline document information content in a relational database schema. Curated results published to Panorama can be aggregated and exported as chromatogram libraries. These libraries can be used in Skyline to pick optimal targets in new experiments and to validate peak identification of target peptides. Panorama is open-source and freely available. It is distributed as part of LabKey Server,[2] an open source biomedical research data management system. Laboratories and organizations can set up Panorama locally by downloading and installing the software on their own servers. They can also request freely hosted projects on https://panoramaweb.org, a Panorama server maintained by the Department of Genome Sciences at the University of Washington.



KEYWORDS: *Software, knowledge base, mass spectrometry, targeted proteomics, SRM, MRM, chromatogram libraries, skyline*

1. INTRODUCTION

In recent years, targeted proteomics has emerged as a mainstream technology to detect and quantify protein and peptide abundances in complex biological samples.[3−5] This popularity has, in part, been fueled by the introduction and broad adoption of the enabling Skyline software tool.[1,3] Skyline is a freely available Windows client application that has become a valuable tool for building and optimizing selected reaction monitoring (SRM) assays. Skyline works across instrument platforms, and its intuitive user interface and rich set of graphs simplifies the development and refinement of targeted methods and also makes it easy for users to visually assess data quality. In addition to being the most commonly used tool for analyzing data acquired on triple quadrupole instruments, Skyline is now widely used for data analysis and method development in experiments that involve TOF and ion trap mass spectrometers, spanning the techniques of data-dependent acquisition (DDA),[6] targeted MS/MS[7] (also referred to as parallel reaction monitoring or PRM),[8] and data-independent acquisition DIA.[9,10] The growth in adoption of Skyline for targeted proteomics experiments has resulted in an increasing accumulation of Skyline documents across various laboratories and organizations. These documents represent a large body of work to develop targeted assays for precise quantitative measurement of proteins and peptides across a range of experimental protocols, which have the potential to become a resource of experimental information. Organizing these docu-

Received: June 29, 2014
Published: August 7, 2014

dx.doi.org/10.1021/pr5006636 | J. Proteome Res. 2014, 13, 4205−4210

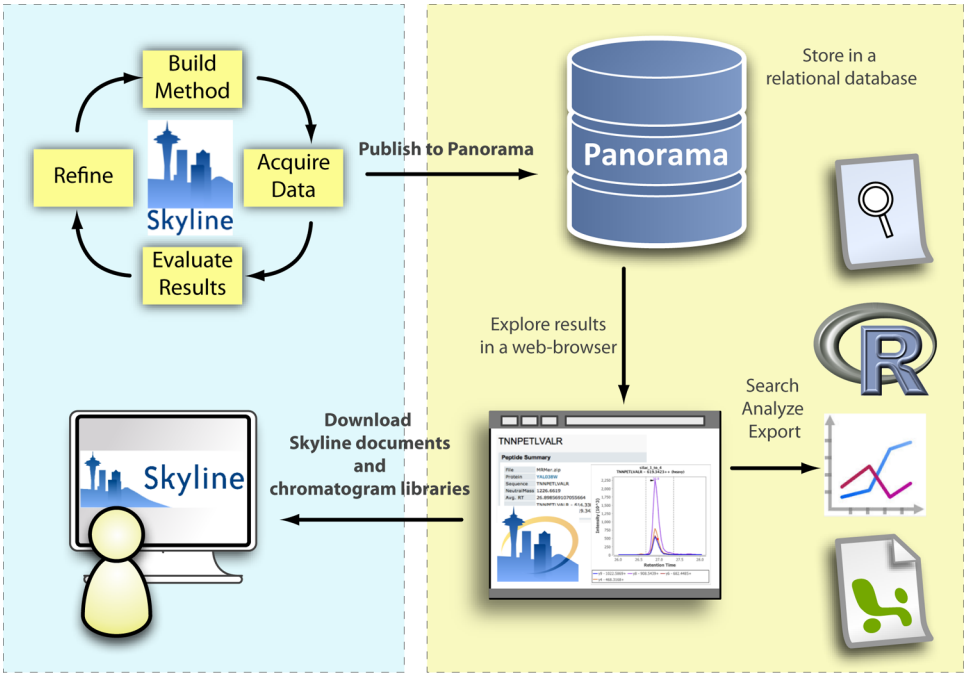


Figure 1. Skyline–Panorama dataflow. Users publish documents containing refined targeted methods from Skyline to a Panorama server, where they are stored in a relational database. Data uploaded to Panorama can be viewed, searched and analyzed in a web-browser interface. Documents published to Panorama as well as chromatogram libraries exported from library folders can be downloaded to view and use in Skyline.

Figure 2. Panorama web-browser interface showing the precursor list in a Skyline document. Clicking on one of the peptide sequences provides a more detailed view for the peptide that includes chromatograms for the precursors in all of the replicates. A graph showing the peak areas for the peptide measured in individual replicates is also displayed. Peak areas can be grouped by a replicate annotation, as shown above for the annotation "Condition" (Disease and Healthy). The source document can be downloaded via a DOWNLOAD link for viewing in Skyline.

dx.doi.org/10.1021/pr5006636 | J. Proteome Res. 2014, 13, 4205–4210

ments into a database is essential to make it possible to review and explore larger sets of well-annotated experimental data in aggregate and to enable use of the accumulated knowledge to inform targeted measurement in future experiments.

Panorama was developed with the goals of data integrity, security, sharing, scalability, and ease-of-use. It allows laboratories and organizations to aggregate curated results contained in Skyline documents, without loss of information. The Panorama schema provides a relational database storage format for the complete Skyline document data model. Panorama is designed both to facilitate sharing results among lab members, external collaborators, and with the general public and to keep private data secure during multiple experiment studies and the peer-review process. The rich security model of LabKey Server combined with its support for installation on private servers, as well as the availability of professional paid support, has allowed Panorama adoption even among large pharmaceutical companies. Panorama has been tightly integrated with Skyline, allowing users to publish their documents to Panorama directly from Skyline and also to query results exported from Panorama when designing new methods in Skyline (Figure 1).

Existing repositories, such as the Peptide Atlas SRM Experiment Library (PASSEL)[11] and MRMaid-DB,[12] provide a place for collecting and exploring experimental SRM data and transitions that have already been published. These repositories are public and do not provide mechanisms for secure data sharing with internal and external collaborators prior to publication. They also lack a local installation option for organizations and laboratories that may prefer to manage their own repositories. They require the submission of RAW data and transition lists and do not yet have support for importing the rich information about the data curation and analysis process or custom annotations available in Skyline documents. In contrast, Panorama supports a knowledge base environment for continued interaction with and analysis of large sets of targeted proteomics data in public or private circles. Knowledge accumulated over time can be used to create chromatogram libraries tailored to specific biological and experimental conditions. These libraries can be used to select the most suitable peptides and product ions to monitor in a targeted experiment, and the stored chromatograms may also be used to increase confidence in chromatogram peak picking.

2. DESIGN AND IMPLEMENTATION

Panorama has been implemented as a module in LabKey Server,[2] a bioinformatics data management platform with rich support for proteomics data. Both LabKey Server and the Panorama module are available as open-source software under the Apache 2.0 license. The complete source code can be downloaded from https://www.labkey.org/wiki/home/Documentation/page.view?name=sourceCode. LabKey Server is a web application that is implemented in Java and runs on the Apache Tomcat web server. It comes with core customizable features for file management, data organization and security, role-based permissions, integration with analysis tools like R, and client APIs for retrieving and interacting with data.

At the heart of the Panorama module is a comprehensive database schema that closely matches the Skyline document data model for capturing the variety of information contained in Skyline files. A web-browser interface (Figure 2) lets the user explore and interrogate results contained in uploaded Skyline documents, including chromatogram views, peak area graphs, and an interactive viewer for MS/MS spectra (Supporting Information Note 1). Users can upload Skyline documents to a hierarchical folder structure and view the precursors and transitions measured in a Skyline document (Supporting Information Figure 1). More detailed views are available for the proteins (Supporting Information Figure 2) and peptides (Supporting Information Figure 3) contained in each document. An enhanced search interface allows searching for a protein, peptide, or specific modifications contained in the documents published to a folder and, optionally, its subfolders (Supporting Information Figure 4).

2.1. Data Organization, Sharing, and Security

LabKey Server, developed over the past decade primarily by the programmers at LabKey Software, is a data management platform that provides the ability to organize data into a customizable workspace with a hierarchical folder structure. Data can be separated into folders by organism, biological matrix, MS instrumentation, and so on, which allows users to identify and organize experiments over a period of time. Panorama users can control access to the data in each folder by configuring the permission settings to keep the data completely private or can provide access to select collaborators or journal reviewers prior to publication (Supporting Information Note 1). Upon publication, or whenever the data owners choose, making a simple change to the security settings can make the data available to the general public. The comprehensive access control model implemented in LabKey Server makes Panorama a viable option for individual laboratories, companies, and other organizations that want secure and flexible management of their accumulated targeted assay data.

Data in a Panorama database can be accessed through client libraries available in several programming languages, including R, JavaScript, Java, Perl, and Python. Programmatic access to the data facilitates building external applications for data analysis and designing custom user interfaces.

2.2. Chromatogram Libraries

Selected reaction monitoring (SRM) has been recognized as a method for detecting and quantifying proteins of interest in a given sample with high sensitivity and specificity.[3] A key requirement in developing an SRM assay is selecting peptides that uniquely identify a protein and have physicochemical properties that make them readily detectable in a mass spectrometer. Also important is the selection of candidate fragment ions for each of the peptides of interest. Current approaches use a combination of spectral libraries from shotgun experiments[13] and de novo prediction algorithms[14] or broad measurement and subsequent method refinement[15] to guide the selection of peptides. The former assumes that peptides most frequently observed in shotgun experiments will produce a good response in a targeted proteomics setting, and the latter necessitates concerted effort for each new method developed. Although spectrum libraries can accurately predict which fragment ions will produce the most intense signal on a triple quadrupole instrument,[13] these libraries do not always have adequate coverage of the entire proteome, especially low-abundance proteins such as human transcription factors and kinases.[16] Even the fragment ion measurements in these libraries may differ from targeted measurements due to instrument differences or interference that goes undetected in spectrum measurement. To overcome these limitations, assay development efforts may measure all precursor and product ions for proteins of interest, under experimental conditions

dx.doi.org/10.1021/pr5006636 | J. Proteome Res. 2014, 13, 4205−4210

with safeguards like reference standard peptides, and then iteratively refine assays.[15,17] Once an assay has been established, it can be used to guide the selection of peptides and product ions to measure in any new experiment under similar conditions. This benefit, however, is frequently lost or cumbersome without organized storage and recall.

Panorama solves this problem by providing the option of organizing assay results in special chromatogram library folders (Supporting Information Note 2). Assays that have been curated in Skyline can be published to designated library folders in Panorama, where they may be subsequently downloaded as a chromatogram library file. The library download is updated with new assays each time a Skyline document is published to the library folder. Users can create library folders for each biological source, sample matrix, experimental protocol, and mass spectrometer type used in any experiment. This makes these libraries as specific to a given experimental setup as necessary and the assays reproducible under similar settings. Chromatogram libraries can be downloaded for use in Skyline to pick optimal peptides for proteins and optimal product ions for precursors in a new experiment. Chromatogram libraries also contain representative chromatograms and peak areas for each precursor that can be compared with newly acquired data for validation of peptide identification (Supporting Information Figure 5 and Supporting Information Note 2).

2.3. Integration with Skyline

Skyline captures extensive information about the targeted proteomics methods it is used to design as well as the acquired results imported and curated within its user interface. This information is stored in the Skyline document, a human readable XML format file, and a companion binary data file. The relational database schema in which Panorama stores data was designed to mirror closely the Skyline data model and to include the full information content it contains, such as complex peptide modification information, normalized retention times (iRT),[18] optimized collision energy,[19] and declustering potential. Panorama is fully integrated into the Skyline user interface, allowing users to configure one or more Panorama servers in Skyline and to publish documents to a folder within a server of choice through a point-and-click user interface (Supporting Information Figure 6 and Supporting Information Note 1). All documents published to Panorama may be downloaded again through a web-browser link and reviewed within Skyline. Documents published to designated library folders in Panorama can be combined into library files and used in Skyline to pick optimal peptides and product ions, as well as to compare acquired data to results contained in Panorama for peak identification confidence. Panorama development keeps pace with Skyline development to include new features added to Skyline such as replicate annotations and mass errors in chromatograms extracted from high-resolution mass spectra.

2.4. Installation Options

Panorama can be installed on a server in a lab by following one of the installation options available for LabKey Server. The latest release of LabKey Server, with the Panorama module, can be downloaded from http://labkey.com/download-labkey-server, and commercial support is available to assist server setup. For laboratories without the resources, expertise, or interest to maintain their own Panorama server, a public server, PanoramaWeb, is hosted at the University of Washington by the MacCoss lab (https://panoramaweb.org). Anyone can request a free project in which they will have full administrative rights. These rights includes the ability to configure the project workspace with the desired folder hierarchy, project membership, and access control policies, as one would on any private installation. PanoramaWeb currently hosts projects for 66 different laboratories, and a total of 779 Skyline documents have been uploaded to the server. Latest figures on PanoramaWeb usage can be viewed at the following link: https://panoramaweb.org/labkey/wiki/home/page.view?name=Dashboard.

3. USE CASES AND PUBLIC IMPLEMENTATIONS THAT ENABLE COLLABORATION

Panorama facilitates collaborative, targeted proteomics projects that use the Skyline software. In addition to the ability to configure access permissions to suit the needs of a project, Panorama leverages a variety of components and features available in LabKey Server to provide an environment that facilitates collaboration and data sharing. This includes the availability of client libraries in several programming languages that allow external applications to programmatically access the data in a Panorama database. These libraries can also be used on the Panorama server to design custom interfaces. The comprehensive Panorama database schema is exposed to the user to enable creating custom views of the data or even writing custom SQL scripts. Data analysis can be performed within Panorama using familiar tools such as the R software environment for statistical computing. PanoramaWeb (https://panoramaweb.org), the public Panorama server maintained by the Department of Genome Sciences at the University of Washington, hosts several multilab projects that showcase the suitability of the platform as a solution for collaborative targeted proteomics efforts.

The Clinical Proteomic Tumor Analysis Consortium (CPTAC) Assay Portal (https://assays.cancer.gov/),[20] designed to disseminate high-quality proteomics assays, is tied to the CPTAC Assay Portal project (https://daily.panoramaweb.org/labkey/CPTAC_Assays.url) on PanoramaWeb. CPTAC members have published detailed assay characterization data, contained in Skyline documents, to this project. The CPTAC-hosted web portal uses the LabKey Server programmatic interfaces to request data from the PanoramaWeb database. Several analysis scripts, written in the R programming language, were developed by CPTAC members to ensure data quality and to evaluate assay performance. These scripts were integrated into their PanoramaWeb project (Supporting Information Figure 7), and the results from running these scripts are displayed in the CPTAC-hosted portal. As new assays are submitted, these scripts are executed on PanormaWeb via programmatic requests, and the results are imported into the assay portal database.

PanoramaWeb was chosen as the repository for the American Biomedical Research Foundation (ABRF) Standards Proteomics Research Group (sPRG) 2013 study in which over 40 anonymized laboratories published their results to a PanoramaWeb project. Panorama also has the ability to import Excel spreadsheets, allowing results from participants that did not use Skyline for data processing to be added to the same project (Supporting Information Figure 8). In addition, a custom summary table was uploaded to PanoramaWeb (Supporting Information Figure 9), enabling results from all of the participants to be captured in one overview table with hyperlinks leading to the results for individual participants.

Results from analysis scripts, developed by the sPRG members to assess data quality, were added to each participant's folder on PanoramaWeb (Supporting Information Figure 10). The project benefitted from the data management and organization features of Panorama, and data analysis was greatly simplified by the ability to easily access results from multiple participating laboratories. Processed data as well as analysis scripts and their results are available in the ABRF sPRG project on PanoramaWeb (https://panoramaweb.org/labkey/ABRF-sPRG-2013.url).

PanoramaWeb also hosts targeted phosphoproteomics data from an assay developed under the auspices of the Library of Integrated Network-based Cellular Signatures (LINCS) project (www.lincsproject.org). The LINCS consortium engages several centers that generate molecular phenotypic data for cellular perturbations using different assay technologies. It is extremely important that data provenance, traceability, and interoperability are maintained and available to consortium members and other consumers of the data. PanoramaWeb was the logical choice for deposition of all LINCS targeted proteomics data.

This particular use case illustrates the flexibility and system interoperability that PanoramaWeb creates. Execution of a multicenter consortium project requires rigorous attention to sample and assay annotation with appropriate metadata to promote reuse of the data that is generated. The Skyline document model allows faithful assignment of dozens of pieces of metadata to hundreds of samples in individual studies. After publishing LINCS data to PanoramaWeb at https://daily.panoramaweb.org/labkey/LINCS.url, the combination of quantitative MS data with sample metadata annotations allow automated creation of data formats that can then be leveraged by other components of the consortium. Using the R programming language interface, Panorama enables complete data processing from a Skyline document, including normalization, outlier removal, probe failure detection, and summarization, to exporting results in the Gene Cluster Text (GCT) file format (Supporting Information Figure 11), a recognized standard for analysis of gene expression data. This allows compatibility with a number of other downstream tools such as the GENE-E visualization and analysis platform (http://www.broadinstitute.org/cancer/software/GENE-E/index.html) and allows direct import into the LINCS Connectivity Map portal (http://www.lincscloud.org).

4. CONCLUSIONS

Panorama addresses the need to archive and share an increasing number of documents being generated by the growing adoption of Skyline in targeted proteomics workflows. Over time, experiments stored in a Panorama server can serve as a knowledge base of curated targeted assays for the proteomics research community. Development of Panorama has been a collaborative process to enable the implementation of components and features useful to end users and to identify and explore its potential beyond a data archival system. We have described several multilab collaborative projects hosted on PanoramaWeb that are using Panorama not only for storing and sharing Skyline documents but also for analyzing large data sets and presenting results to the broader community.

■ ASSOCIATED CONTENT

Supporting Information

Supporting Information provides more details of the Panorama user interface as well as tutorials describing a broad range of Panorama features and use cases. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: brendanx@uw.edu. Phone: (206) 616-9023. Fax: (206) 685-7301.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors would like to acknowledge financial support from National Institute of Health grants R01 GM103551 (to M.J.M.), U01 CA164186 (to J.D.J.), U24 CA160034 from the National Cancer Institute Clinical Proteomics Tumor Analysis Consortium Initiative (to S.A.C. and A.G.P.) and grants HHSN268201000033C and R01HL096738 from the National Heart, Lung, and Blood Institute (to S.A.C.), as well as funding from the Broad Institute of MIT and Harvard.

■ REFERENCES

(1) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966−968.

(2) Nelson, E. K.; Piehler, B.; Eckels, J.; Rauch, A.; Bellew, M.; Hussey, P.; Ramsay, S.; Nathe, C.; Lum, K.; Krouse, K.; et al. LabKey Server: an open source platform for scientific data integration, analysis and collaboration. *BMC Bioinf.* **2011**, *12*, 71.

(3) Marx, V. Targeted proteomics. *Nat. Methods* **2013**, *10*, 19−22.

(4) Picotti, P.; Bodenmiller, B.; Aebersold, R. Proteomics meets the scientific method. *Nat. Methods* **2013**, *10*, 24−27.

(5) Gillette, M. A.; Carr, S. A. Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nat. Methods* **2013**, *10*, 28−34.

(6) Schilling, B.; Rardin, M. J.; MacLean, B. X.; Zawadzka, A. M.; Frewen, B. E.; Cusack, M. P.; Sorensen, D. J.; Bereman, M. S.; Jing, E.; Wu, C. C.; et al. Platform-independent and label-free quantitation of proteomic data using MS1 extracted ion chromatograms in skyline: application to protein acetylation and phosphorylation. *Mol. Cell. Proteomics* **2012**, *11*, 202−214.

(7) Sherrod, S. D.; Myers, M. V.; Li, M.; Myers, J. S.; Carpenter, K. L.; Maclean, B.; Maccoss, M. J.; Liebler, D. C.; Ham, A.-J. L. Label-free quantitation of protein modifications by pseudo selected reaction monitoring with internal reference peptides. *J. Proteome Res.* **2012**, *11*, 3467−3479.

(8) Peterson, A. C.; Russell, J. D.; Bailey, D. J.; Westphall, M. S.; Coon, J. J. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. *Mol. Cell. Proteomics* **2012**, *11*, 1475−1488.

(9) Venable, J. D.; Dong, M.-Q.; Wohlschlegel, J.; Dillin, A.; Yates, J. R. Automated approach for quantitative analysis of complex peptide mixtures from tandem mass spectra. *Nat. Methods* **2004**, *1*, 39−45.

(10) Gillet, L. C.; Navarro, P.; Tate, S.; Rost, H.; Selevsek, N.; Reiter, L.; Bonner, R.; Aebersold, R. Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **2012**, *11*, O111.016717.

(11) Farrah, T.; Deutsch, E. W.; Kreisberg, R.; Sun, Z.; Campbell, D. S.; Mendoza, L.; Kusebauch, U.; Brusniak, M.-Y.; Hüttenhain, R.;

Schiess, R.; et al. PASSEL: the PeptideAtlas SRM experiment library. *Proteomics* **2012**, *12*, 1170−1175.
(12) Cham, J. A.; Bianco, L.; Barton, C.; Bessant, C. MRMaid-DB: a repository of published SRM transitions. *J. Proteome Res.* **2010**, *9*, 620−625.
(13) Prakash, A.; Tomazela, D. M.; Frewen, B.; Maclean, B.; Merrihew, G.; Peterman, S.; Maccoss, M. J. Expediting the development of targeted SRM assays: using data from shotgun proteomics to automate method development. *J. Proteome Res.* **2009**, *8*, 2733−2739.
(14) Mallick, P.; Schirle, M.; Chen, S. S.; Flory, M. R.; Lee, H.; Martin, D.; Ranish, J.; Raught, B.; Schmitt, R.; Werner, T.; et al. Computational prediction of proteotypic peptides for quantitative proteomics. *Nat. Biotechnol.* **2007**, *25*, 125−131.
(15) Bereman, M. S.; MacLean, B.; Tomazela, D. M.; Liebler, D. C.; MacCoss, M. J. The development of selected reaction monitoring methods for targeted proteomics via empirical refinement. *Proteomics* **2012**, *12*, 1134−1141.
(16) Stergachis, A. B.; MacLean, B.; Lee, K.; Stamatoyannopoulos, J. A.; MacCoss, M. J. Rapid empirical discovery of optimal peptides for targeted proteomics. *Nat. Methods* **2011**, *8*, 1041−1043.
(17) Abbatiello, S. E.; Mani, D. R.; Keshishian, H.; Carr, S. A. Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. *Clin. Chem.* **2010**, *56*, 291−305.
(18) Escher, C.; Reiter, L.; MacLean, B.; Ossola, R.; Herzog, F.; Chilton, J.; MacCoss, M. J.; Rinner, O. Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **2012**, *12*, 1111−1121.
(19) Maclean, B.; Tomazela, D. M.; Abbatiello, S. E.; Zhang, S.; Whiteaker, J. R.; Paulovich, A. G.; Carr, S. A.; Maccoss, M. J. Effect of collision energy optimization on the measurement of peptides by selected reaction monitoring (SRM) mass spectrometry. *Anal. Chem.* **2010**, *82*, 10116−10124.
(20) Whiteaker, J. R.; Halusa, G. N.; Hoofnagle, A. N.; Sharma, V.; MacLean, B.; Yan, P.; Wrobel, J. A.; Kennedy, J.; Mani, D. R.; Zimmerman, L. J.; et al. CPTAC Assay Portal: a repository of targeted proteomic assays. *Nat. Methods* **2014**, *11*, 703−704.