An update on MRMAssayDB: a comprehensive resource for targeted proteomics assays in the community

Bhowmick, P.; Roome, S.; Borchers, C.H.; Goodlett, D.R.; Mohammed, Y.

Journal of proteome research

pubs.acs.org/jpr

Technical Note

An Update on MRMAssayDB: A Comprehensive Resource for Targeted Proteomics Assays in the Community

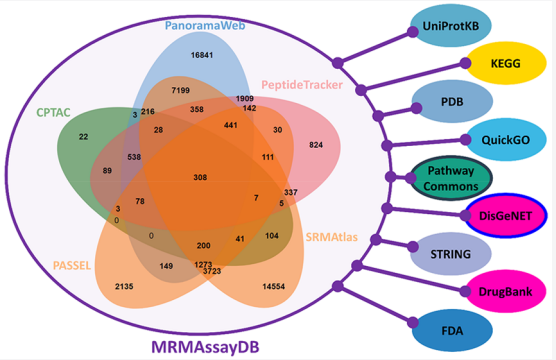
Pallab Bhowmick, Simon Roome, Christoph H. Borchers, David R. Goodlett, and Yassene Mohammed*

Cite This: *J. Proteome Res.* 2021, 20, 2105–2115

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Precise multiplexed quantification of proteins in biological samples can be achieved by targeted proteomics using multiple or parallel reaction monitoring (MRM/PRM). Combined with internal standards, the method achieves very good repeatability and reproducibility enabling excellent protein quantification and allowing longitudinal and cohort studies. A laborious part of performing such experiments lies in the preparation steps dedicated to the development and validation of individual protein assays. Several public repositories host information on targeted proteomics assays, including NCI's Clinical Proteomic Tumor Analysis Consortium assay portals, PeptideAtlas SRM Experiment Library, SRMAtlas, PanoramaWeb, and PeptideTracker, with all offering varying levels of details. We introduced MRMAssayDB in 2018 as an integrated resource for targeted proteomics assays. The Web-based application maps and links the assays from the repositories, includes comprehensive up-to-date protein and sequence annotations, and provides multiple visualization options on the peptide and protein level. We have extended MRMAssayDB with more assays and extensive annotations. Currently it contains >828 000 assays covering >51 000 proteins from 94 organisms, of which >17 000 proteins are present in >2400 biological pathways, and >48 000 mapping to >21 000 Gene Ontology terms. This is an increase of about four times the number of assays since introduction. We have expanded annotations of interaction, biological pathways, and disease associations. A newly added visualization module for coupled molecular structural annotation browsing allows the user to interactively examine peptide sequence and any known PTMs and disease mutations, and map all to available protein 3D structures. Because of its integrative approach, MRMAssayDB enables a holistic view of suitable proteotypic peptides and commonly used transitions in empirical data. Availability: http://mrmassaydb.proteincentre.com.



KEYWORDS: *knowledgebase, targeted proteomics, assay, MRM, multiplexed quantitation, internal standard*

INTRODUCTION

Targeted proteomics quantification of proteins is typically performed using a triple quadrupole operated in multiple reaction monitoring (MRM) mode, or using a mass spectrometer capable of parallel reaction monitoring (PRM) like Orbitraps.[1−7] In both, an initial filtering step isolates the precursor peptide ion typically using a quadrupole filter. After fragmentation, a characteristic fragment ion is isolated using another quadrupole filter in MRM, or all fragment ions are selected for monitoring in PRM.[5,6] In a scheduled LC/MRM-MS or LC/PRM-MS analysis, precursor and product ions are monitored according to the peptide elution time from the liquid chromatography system.[8,9] This allows quantitation of a large number of target peptides and, by inference, the corresponding proteins. Reproducible quantitation of almost 300 proteins in 45 min is possible.[10−15] New developments in ion mobility-mass spectrometry (IM-MS) allow an additional gas-phase separation step that promises even faster quantification.[16]

A crucial difference between targeted and discovery proteomics lies in the scheduled acquisition of the former. Developing a targeted proteomics assay is a laborious multistep process and involves collecting, validating, and documenting various levels of information on each peptide assay.[17] This includes the uniqueness of the peptide (proteotypic surrogate for the protein of interest) within a particular proteome, its retention time under specific LC conditions, the corresponding precursor/fragment ion pairs, and more. Selecting a suitable proxy peptide that can also be chemically synthesized and used as internal standard requires almost 30 rules be reviewed.[18] A crucial, and probably most important, rule is whether a peptide has been previously observed in MS/MS analyses and

Special Issue: Software Tools and Resources 2021

Received: November 30, 2020
Published: March 8, 2021

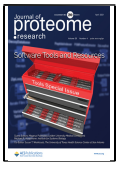


Table 1. MRM Peptides Collected from the Targeted Proteomics Repositories

	CPTAC	PanoramaWeb	PASSEL	PeptideTracker	SRMAtlas
number of MRM peptide assays collected – Jan. 2021	2 388 covering 1642 proteins	212 094 covering 29 683 proteins	25 662 covering 8 641 proteins	9 331 covering 5 208 proteins	684 067 covering 28 905 proteins
number of stripped peptides – Jan. 2021	2 365	149 057	25 085	8 355	661 130
number of peptide assays failed validation	16	512	615	77	4 526
content and assay sources	• hosts assays characterized using detailed guidelines from the CPTAC consortium • assays are from CPTAC consortium	• hosts assays obtained from data submission as part of publication • assays are from proteomics community researchers (usually using Skyline)	• includes datasets from published (71 projects) and unpublished (51 projects)	• hosts assays and information on protein concentrations • most assays are from University of Victoria Proteomics Centre	• hosts assays covering the proteomes of human, *M. tuberculosis* and yeast (99.7%, 97.0%, 97.0% coverage respectively) • data are from three papers
organism coverage	• mainly human assays (1 578) • recently mouse assays were added (798)	• multiple organisms – 92 as of January 2021) • most assays are from mouse (58 031), yeast (53 591), and human (47 920)	• multiple organisms – as of January 2021 17/10 before/after validation • mainly *M. tuberculosis* (12 505), human (7 028), and rat (3 481)	• four organisms: mouse (4 549), human (4 425), rat (267), and wild boar (90)	• three organisms: human (508 594), *M. tuberculosis* (44 269), and yeast (131 204)
raw data hosting	• no	• yes	• yes	• no	• yes
data submission	• does not have an option for external data submission	• accepts external submission by researchers	• accepts external submission by researchers	• accepts external submission by researchers	• does not have an option for external data submission
sample processing protocols	• available from CPTAC assay associated publication	• available from project associated publication	• available for assays associated with publication from related papers • unpublished datasets have associated description on ProteomeXchange	• available on resource and linked to each assay	• available in the three associated papers
open access	• yes	• yes	• yes	• yes	• yes
access method	• data can be downloaded from Web interface	• data can be downloaded from Web interface • data is accessible via API	• data can be downloaded from Web interface • data is accessible via API	• data can be downloaded from Web interface • data is accessible via API	• data can be downloaded from Web interface

therefore is known to be detectable. This type of information is scattered across several online public data repositories hosting raw data and resultant identification information on previous proteomics experiments.[19−22]

Sharing information on existing targeted proteomics assays allows scientists to design their targeted proteomics experiments faster and better. Multiple resources for targeted proteomics data exist and include PeptideAtlas SRM Experiment Library (PASSEL),[23] NCI's Clinical Proteomic Tumor Analysis Consortium (CPTAC),[24] PanoramaWeb,[25] SRMAtlas,[26−28] and PeptideTracker.[29] However, the information hosted in these data repositories and knowledge bases is heterogeneous because they were collected with different goals in mind. To help users address this issue, we previously introduced MRMAssayDB as an integrated Web-resource with comprehensive information on all available targeted proteomics assays in these community-wide online repositories.[30] On its release date, MRMAssayDB contained 168 000 assays covering 34 000 proteins from 63 organisms. We have since added a large number of assays and updated the application interface with various annotations, which prompted us to update the resource and redesign some of its aspects. Beside the more than 4-fold increase in the total number of assays added, various protein annotations related to disease, biological pathway, and interaction associations were also added. Additionally, a new integrated visualization module allows mapping of protein and peptide annotations for a better user experience. The new version of MRMAssayDB is larger in content and scope, making it an excellent starting point for designing targeted proteomics experiments.

MATERIALS AND METHODS

Data Resources

PASSEL is a generic data repository from the Institute for System Biology. MRM experimental results can be submitted

https://dx.doi.org/10.1021/acs.jproteome.0c00961
J. Proteome Res. 2021, 20, 2105−2115

by users along with the corresponding raw data and later are made available to the research community (www.peptideatlas.org/passel/).[23] The assay portal of CPTAC[24] from the National Cancer Institute (NCI) hosts well-characterized targeted proteomic assays (http://assays.cancer.gov). Its goals are centered around standardizing targeted MS-based assays to achieve robust quantification of all human proteins.[31] Submission to the portal is done by consortium partners. PanoramaWeb[25] is a repository for storing, sharing, and analyzing targeted proteomic experiment processed by Skyline software,[32] and therefore is very popular among the targeted proteomics community (https://panoramaweb.org). SRMAtlas is a compendium of targeted proteomics assays for the quantification of annotated human proteins (http://www.srmatlas.org/). It includes assays to quantify spliced variants, nonsynonymous mutations, and post-translational modifications.[26] PeptideTracker[29] was introduced as a knowledge base for collecting and storing information on protein concentration ranges in biological tissues along with the detailed description of the assays that were used (http://peptidetracker.proteincentre.com).

New entries are added continuously to these repositories. Some repositories are specific like CPTAC, while others are more generic like PASSEL. Some repositories store the raw data like PanoramaWeb, while others put emphasis on listing sample preparation protocols like CPTAC and PeptideTracker. PeptideTracker lists the determined protein concentration ranges in samples, as measured by MRM.

Each assay is annotated with various information including UniProtKB[33] accession number, protein name, gene name, organism, peptide sequence, uniqueness in proteome, peptide presence in isoforms, modification, labeled internal standard if used, and a hyperlink to the relevant proteomics resources from which the information were obtained. In addition, assays are annotated with biological pathway associations as present in Pathway Commons[34,35] and KEGG,[36] known protein−protein interactions as present in STRING,[37] disease associations as documented in UniProtKB[33] and DisGeNET,[38] associations with drugs as in DrugBank,[39] available 3D structures from PDB,[40] and PTMs as well as sequence variance known from UniProtKB,[33] community proteomics experiments, and public data sets.[41,42]

FDA Approved Protein Biomarkers

We retrieved the information available from the FDA database on all approved assays, and manually checked the entries for those associated with proteins. Using UniProtKB,[33] KEGG,[36] DrugBank,[39] and NIH's PubChem,[43,44] we manually assigned to each FDA entry the corresponding UniProtKB accession numbers. We then used MRMAssayDB and mapped the protein entries in the FDA approved assays to available targeted proteomics assays in the community.

Software Implementation

MRMAssayDB was written mainly in Python 2.7 (www.python.org). The user Web interface was developed using the Django 1.8 framework (https://djangoproject.com), and plots are generated using JavaScript. MolArt is used for the interactive visualization of protein annotation and 3D structures.[41] Cytoscape.js[45] was used to plot the interactive PPI networks. Data from PeptideTracker,[29] PASSEL,[23] CPTAC,[24] PanoramaWeb,[25] SRMAtlas,[26−28] UniProtKB,[33] PDB,[40] Pathway Commons,[34,35] KEGG,[36] STRING,[37] QuickGO,[46] DisGeNET,[38] DrugBank[39] are automatically retrieved using the APIs of these resources by routines written in Java (www.oracle.com/java/index.html), Python, and Selenium Webdriver (http://www.seleniumhq.org).

RESULTS

Thousands of targeted proteomics assays have been used previously in various experiments. The information on these assays are largely available in different repositories including PanoramaWeb,[25] CPTAC,[24] SRMAtlas,[26−28] PASSEL,[23] and PeptideTracker.[29] Each one of these resources was developed with a specific goal in mind; however, the information and entries hosted in each are complementary (Table 1). Together they form a valuable foundation in the targeted proteomics community. MRMAssayDB integrates this disperse information into a single Web-based application, provides up-to-date annotations on the thousands of these assays, and makes all available for researchers. Currently, 828 974 assay entries corresponding to 732 132 unique stripped peptides are listed, covering 51 668 proteins from 94 organisms (Table 2, Supporting Information, Table S1). These entries are based on 834 132 assays corresponding to 736 391 unique stripped peptides currently available in the targeted proteomics community repositories[23−29] (a peptide assay associates a peptide with a protein; same peptide can however be associated with proteins from multiple organisms). The difference, 5710 assays corresponding to 5601 stripped peptide failed during validation against the reference protein sequences (based on UniProtKB release 2020_6), i.e., they are found not to be part of the associated protein. However, if an organism is not present in MRMAssayDB or if researchers are interested in a specific set of proteins in a sample, they can upload their own proteome, and in this case, all 736 000 unique stripped peptides will be searched for suitable assays that enable targeted proteomics experiment in that specific proteome as provided by the researchers. Uniqueness of a peptide assay in the proteome is documented in MRMAssayDB and is always displayed in the results. It is defined with the purpose of quantifying of the protein in the specific organism background; i.e., a peptide assay will be flagged as not unique if it is present in more than one protein in the organism proteome. An exception to this is if the peptide is present in multiple isoforms; in that case, it is still considered unique.

Table 2. Top 5 Organisms with Largest Number of Entries in MRMAssayDB[a]

	number of stripped peptide entries (March 2018)	number of stripped peptide entries (January 2021)
Homo sapiens	101 613 (19 777 proteins)	494 917 (21 699 proteins)
Saccharomyces cerevisiae	47 946 (6478 proteins)	137 003 (6617 proteins)
Escherichia coli	3420 (2363 proteins)	5211 (2530 proteins)
Mus musculus	5327 (2304 proteins)	52 342 (7399 proteins)
Mycobacterium tuberculosis	5892 (1681 proteins)	46 367 (3825 proteins)
Others	4211 (2137 proteins)	28 334 (9598 proteins)
Total	168 409 (34 740 proteins)	732 132 (51 668 proteins)

[a]A full list with all organisms is included in Supporting Information Table S1.

A protein may have multiple proteotypic peptide assays already in use. If one's interest is a simple quantification of proteins, a logical choice is to consider those peptides used independently by different researchers. This also applies for the

https://dx.doi.org/10.1021/acs.jproteome.0c00961
J. Proteome Res. 2021, 20, 2105−2115

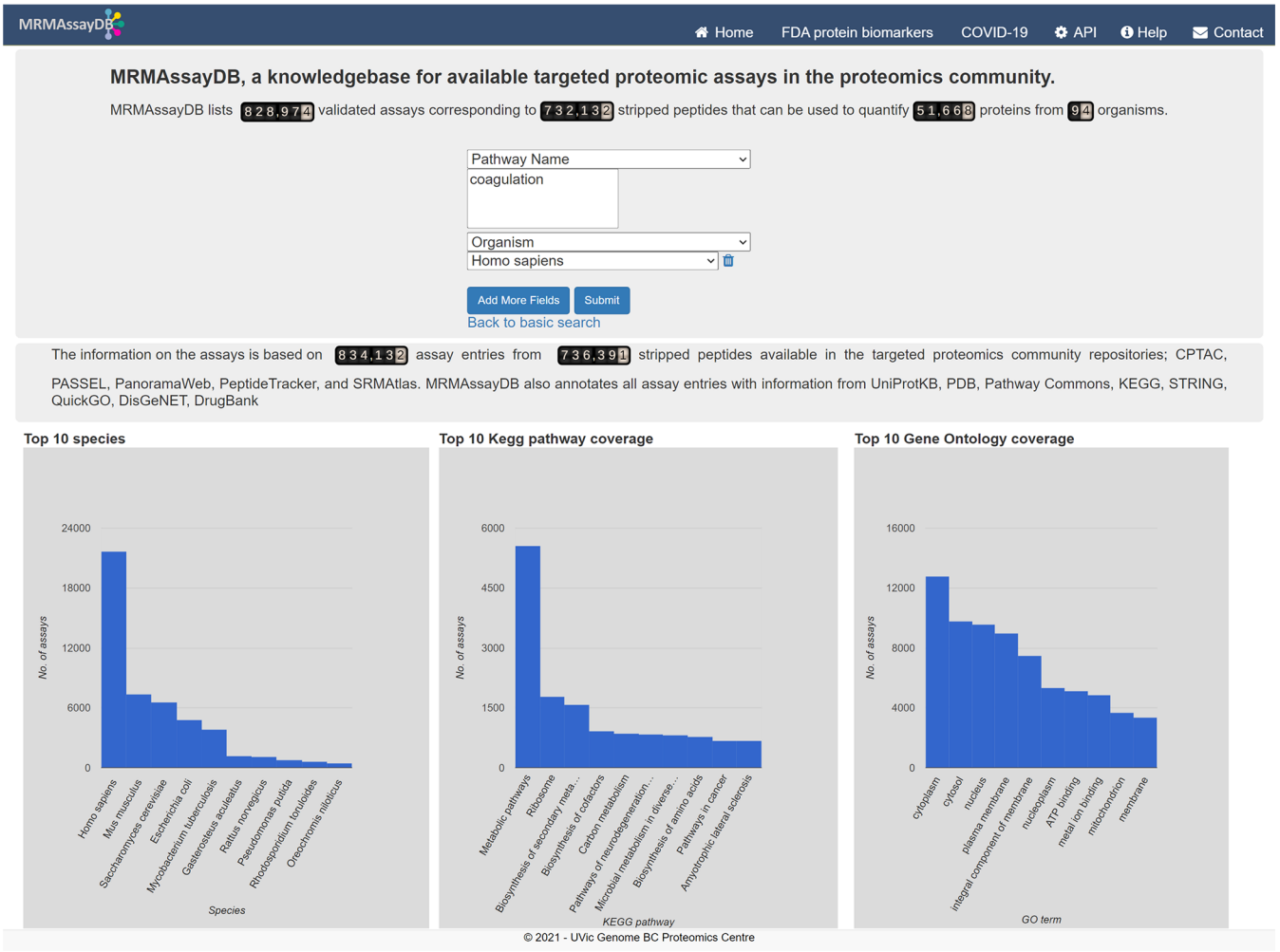


Figure 1. Screenshot of MRMAssayDB home page with simple and advanced search functionalities. Statistics on assays are also displayed on the home page.

transitions of each surrogate peptide. MRMAssayDB lists the most common transitions for each peptide MRM assay as reported in the different repositories. Because all transitions are originated from actual assays used in one or more laboratory or experiment, good transitions are those appearing most frequently in the documented assays. There might be alternative methods for determining goodness of transitions—for example, based on signal intensity, best peak shape, those with no interferences, or a combination thereof. However, if researchers have followed different paths to determine the best set of transitions, validated, and applied them experimentally, then those transitions used most frequently in experiments are most likely to work.

Besides the main goal of having an extensive resource for available targeted proteomics assays, MRMAssayDB provides numerous protein and assay annotations that makes it a resource for researchers looking to design a targeted proteomics experiment. Currently, as of January 2021, 87% of all assays are based on unique peptides. From all of these assays, we were able to map 17 219 proteins to 2399 biological pathways, 48 225 proteins to 21 424 GO terms, 3381 proteins to 8126 drugs, and 4128 proteins to 25 543 diseases. These numbers change over time with the periodic update; however, they provide a good overview on the scope of the assays in the community. In the following, we describe some of the resource features and included annotations.

Advanced Search and Result Viewing Functionalities

Users can benefit from simple or advanced search modes as well as post search filtering. The search can be performed from the home page using protein name, protein UniProtKB accession, partial peptide sequence, gene name, organism, as well as protein annotation like association with biological pathways or diseases (Figure 1). For specific results, a combination of two or more aspects is also possible. Once the search results are present on the screen, the user can add filter terms to each column individually. Columns can also be blended and viewed based on the user interest. Search results can always be downloaded as a spreadsheet document for further local analysis. MRMAssayDB has also an application programming interface (API) allowing developers to send queries in an automated manner and enabling them to incorporate MRMAssayDB information in own data- and knowledge-bases.

Annotation and Visualization

In the new release of MRMAssayDB we have incorporated a molecular structural annotation of each protein and peptide assay using the MolArt package.[41] This allows users to interactively explore the protein annotations in the context of

https://dx.doi.org/10.1021/acs.jproteome.0c00961
J. Proteome Res. 2021, 20, 2105−2115

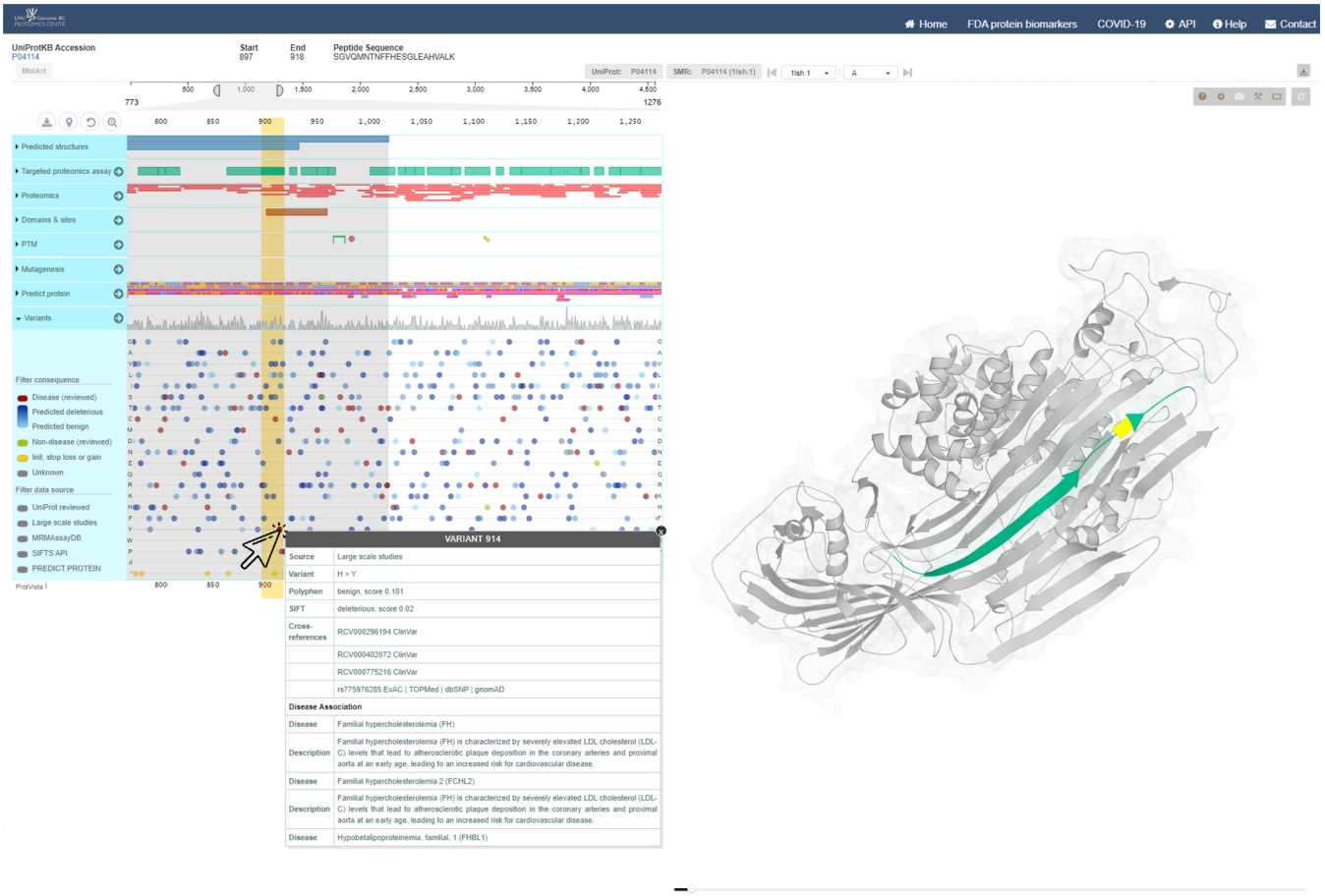


Figure 2. An MRM assay along with sequence annotations visualized in MRMAssayDB. The example here is from Apolipoprotein B-100 protein (APOB, P04114) showing its annotations and features. Molecular structural annotations of the protein and peptide assays are visualized based on adapted MolArt. Users can interactively browse the sequence features and annotations in ProtVista (example of H914Y variant is shown), obtain detailed information on each variant or PTM, and map these to their positions on available 3D structures of the protein visualized by LiteMol.

known experimental or predicted protein structures (Figure 2). Users can navigate a screen divided between the sequence annotation visualized by ProtVista[42] and the 3D structural data of the protein visualized by LiteMol.[47] Both views are coupled allowing intuitive interactive exploration of the protein, proteotypic peptide assays, annotations, and their positions in the available 3D structures. The annotations include post-translational modifications, protein domains and chains, as well as known amino acid variations and their relation to disease. The proteotypic peptides available as targeted proteomics assays are mapped so users can easily view the molecular annotation and structural location of all surrogate peptides with assays. This provides a simple yet powerful visualization of the various information relevant to a targeted proteomics experiment. The annotations are retrieved live from UniProtKB, community proteomics experiments, public data sets.[41] This is important as information on protein annotation is in flux; i.e., new annotations are continuously becoming available on the positions of disease-related mutations or PTMs on the protein.[48,49] While the majority of targeted proteomics assays are based on nonmodified peptides, it is important to have updated information on whether a selected surrogate peptide carries a mutation or PTM or not. MRMAssayDB provides this information in few clicks.

Protein−Protein Interaction and Biological Pathways

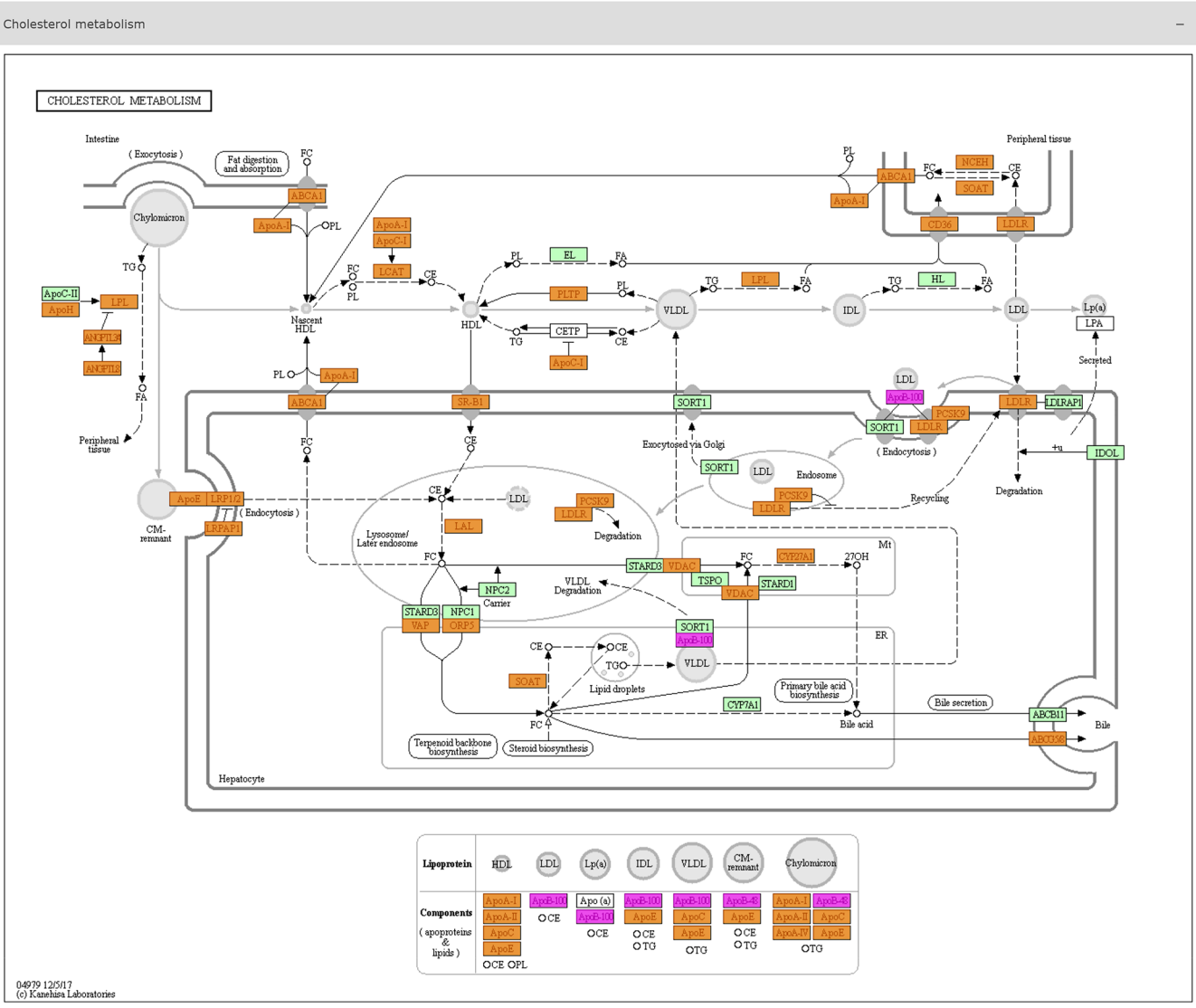
Assays in MRMAssayDB are linked to documented biological pathways and known protein−protein interactions. Multiple resources host these annotations, and although sometimes redundant, complementarity between resources is an essential aspect. In our original release in 2018, we have included KEGG[36] and STRING.[37] In order to extend on the annotation, assay entries are now also linked to Pathway Commons annotations.[34,35] Pathway Commons is an extensive integrated resource of biological pathways and includes annotations from 22 databases with more than 5700 pathways.

A visualization of the pathways and protein−protein interaction is also supported. The proteins in the pathway or interaction network are color-coded showing and linking those proteins that have available targeted proteomics assays. If a researcher is interested in designing a targeted proteomics experiment to infer information on specific biological pathway or the interactome of a specific protein, they are able to investigate the network and possible coverage using available assays in one single interactive plot (Figure 3, Figure 4, and Supporting Information Figure S1).

Functional, Disease, and Drug Associations

The targeted protein assays are annotated with functional data as represented by Gene Ontology (GO) annotations[46] (Supporting Information Figure S2). This includes information on the three GO aspects: Biological Process, Molecular

https://dx.doi.org/10.1021/acs.jproteome.0c00961
J. Proteome Res. 2021, 20, 2105−2115



Cholesterol metabolism

Violet indicates the protein or gene that the user searched for in MRMAssayDB, while orange indicates that the protein or gene is present in MRMAssayDB.

All color-coded boxes are linked to the data table in MRMAssayDB via hyperlinks. Rectangular green boxes represent non-mapped genes/proteins in the pathway.

Figure 3. Available assays for mouse in the Cholesterol Metabolism pathway as presented in MRMAssayDB and data obtained from KEGG. Larger coverage in humans as can be seen in Supporting Information, Figure S1.

Function, and Cellular Component.[50] The assays are also annotated with known disease associations (Supporting Information Figure S3). The original application linked to the curated disease annotations as present in UniProtKB. To allow a wider view on expected and machine-curated associations, we added in the new version of MRMAssayDB disease annotations as present in DisGeNET.[38] DisGeNET is an integrated resource with disease associations inferred from expert curated repositories, GWAS, animal models, and scientific literature. Furthermore, annotations for protein−drug association are also provided in MRMAssayDB (Supporting Information Figure S4). We continue to link the assays based on the target protein associations as in DrugBank.[39] Having targeted proteomics assays annotated with their known functional, disease, and drug associations can be very helpful for design experiments to study disease, protein−drug interaction, or novel unknown protein function.[14,51]

FDA Protein Biomarkers and COVID-19 Related Proteins

The US Food and Drug Agency (FDA) lists 1140 approved clinical tests, 208 of which are associated with 253 proteins. Using MRMAssayDB we have data-mined all available proteomics assays to determine those, which can be used for the FDA-approved protein markers. We identified quantitative proteomics assays for 222 proteins of the 241 FDA approved protein markers, 109 of the 222 are known to be secreted, with 64 of these proteins listed in MRMAssayDB with information on their concentration ranges in human plasma (Figure 5). A selection of current assays and annotations (as of January 2021) is included in Supporting Information Table S2. The selection was based on the assays that are most frequent among the original resources, whether possible protein concentration values are available, and whether previous information on assay suitability for targeted proteomics experiment is available.[52] Updated information on these protein assays and their annotations is available on

https://dx.doi.org/10.1021/acs.jproteome.0c00961
J. Proteome Res. 2021, 20, 2105−2115

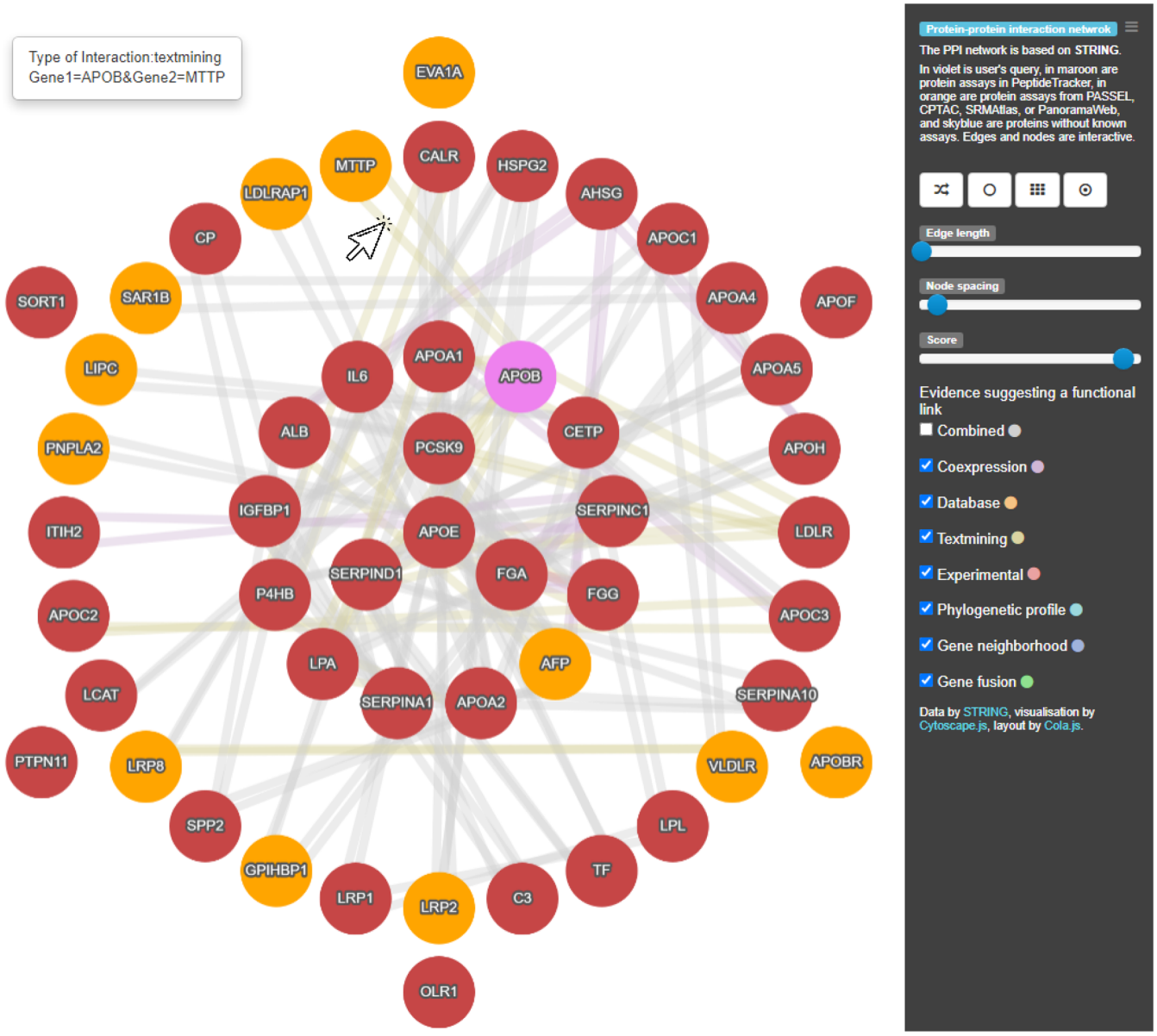


Figure 4. Protein−protein interaction network of apolipoprotein B-100 (APOB) as shown ion MRMAssayDB.

MRMAssayDB home page (from the navigation bar at the top of the home screen or by using the following link: http://mrmassaydb.proteincentre.com/fdaassay/).

In the wake of the COVID-19 pandemic and the increased focus of research on SARS-CoV-2, we maintain a list of proteins associated with the infection as reported in UniProtKB. Currently, there are 60 human proteins curated to be associated with COVID-19, and there are assays in MRMAssayDB for all 60 proteins (as of January 2021). These assays can also be viewed directly from the navigation bar at the top of the home screen or by using the following link: http://mrmassaydb.proteincentre.com/covid19/.

CONCLUSIONS

MRMAssayDB is an integrated Web-based resource for available targeted proteomics assays in the community. It integrates targeted assay information from the five major MRM/PRM assay portals and data repositories including PASSEL, PanoramaWeb, CPTAC assay portal, SRMAtlas, and PeptideTracker. Information on which data repository reported each assay allows sorting search results to determine most frequent assays for each protein. For each assay, the most frequent transitions among repositories and instruments are annotated. This allows informed design of targeted proteomics experiments.

The assays are linked to available annotations on the protein sequence, function, and interactions. These annotations include associations with biological pathways, protein−protein interactions, Gene Ontology terms, known PTMs, disease-associated mutations, protein domains, disease involvements, and drug associations. Advanced synchronized visualization of the annotations and the 3D structure of the target protein, as well as how the proteotypic surrogate peptides map to all, empower the user to carry out a comprehensive study of their protein of interest and choose most suitable assay. As targeted proteomics data sets are continuously shared by scientists and researchers through the data repositories, we intend to

https://dx.doi.org/10.1021/acs.jproteome.0c00961
J. Proteome Res. 2021, 20, 2105−2115

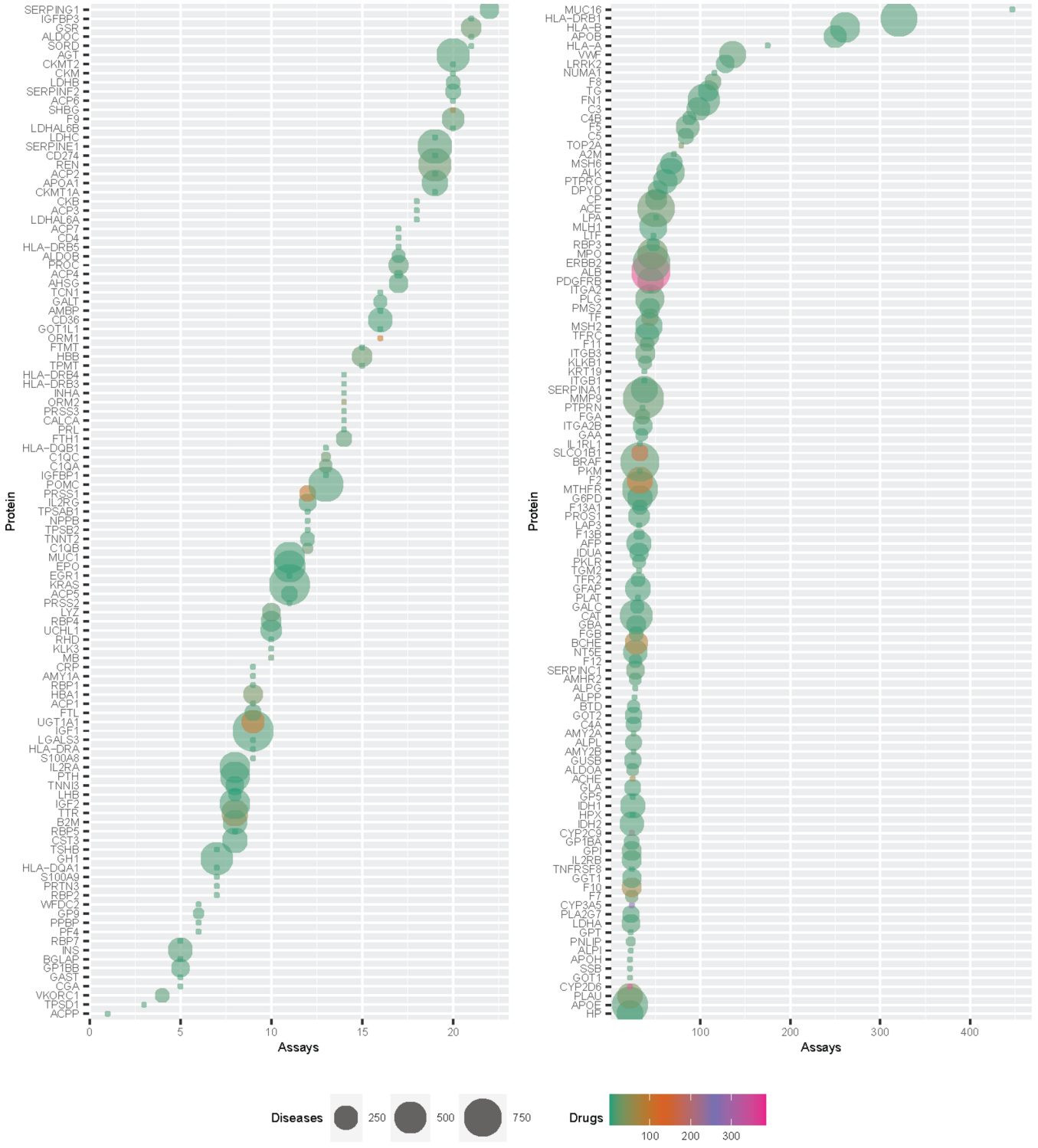


Figure 5. Targeted proteomics assays associated with FDA-approved protein markers as present in MRMAssayDB. The *x*-axis indicates number of assays, the *y*-axis lists the assays, and size and color indicate the number of drugs and diseases, respectively, associated with the protein marker.

continue integrating these into our knowledgebase. Currently, the assay entries are compiled automatically once a month.

A targeted proteomics experiment starts always with a planning phase about which targets to measure. Using MRMAssayDB researchers can streamline this process by finding suitable assays that were previously applied. The large number of entries and annotations available to researchers in this single resource are a result of long-term archiving of experimental information that has been carried out in the targeted-proteomics community. We hope by maintaining MRMAssayDB researchers can easier design and perform targeted proteomics experiments, and by sharing their experimental raw data in public repositories, the assay information in MRMAssayDB will continue to be up-to-date. Access to the software is available free of charge at http://MRMAssayDB.proteincentre.com/.

https://dx.doi.org/10.1021/acs.jproteome.0c00961
J. Proteome Res. 2021, 20, 2105−2115

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.0c00961.

Figure S1: Available assays for human in the Cholesterol Metabolism pathway as presented in MRMAssayDB; Figure S2: GO annotations for Coagulation Factor IX (F9) as presented in MRMAssayDB; Figure S3: Disease associations, here of F9, are presented in searchable and linked terms with tabs for DesGeNET and for UniProtKB; Figure S4: Drug associations as shown in MRMAssayDB search results of P00740 (coagulation factor IV, F9); Table S1: The organisms and number of associated assays in MRMAssayDB (as of January 2021) (PDF)

Table S2: Selected assays and annotations for FDA approved protein biomarkers; the selection was based on including assays that are most frequent in the resources and whether they have protein concentration values available; the list was generated in January 2021; however, updated list and annotations are available on the MRMAssayDB Web site and can be accessed directly at http://mrmassaydb.proteincentre.com/fdaassay/ (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Yassene Mohammed − *University of Victoria − Genome BC Proteomics Centre, Victoria, British Columbia V8Z 7X8, Canada; University of Victoria, Victoria, British Columbia V8P 5C2, Canada; Center for Proteomics and Metabolomics, Leiden University Medical Center, 2333 ZA Leiden, Netherlands;* orcid.org/0000-0003-3265-3332; Email: yassene@proteincentre.com, y.mohammed@lumc.nl

Authors

Pallab Bhowmick − *University of Victoria − Genome BC Proteomics Centre, Victoria, British Columbia V8Z 7X8, Canada; University of Victoria, Victoria, British Columbia V8P 5C2, Canada*

Simon Roome − *University of Victoria − Genome BC Proteomics Centre, Victoria, British Columbia V8Z 7X8, Canada; University of Victoria, Victoria, British Columbia V8P 5C2, Canada*

Christoph H. Borchers − *Proteomics Centre, Segal Cancer Centre, Lady Davis Institute, Jewish General Hospital, McGill University, Montreal, Quebec H3T 1E2, Canada; Gerald Bronfman Department of Oncology, Jewish General Hospital, Montreal, Quebec H3T 1E2, Canada; Department of Data-Intensive Science and Engineering, Skolkovo Institute of Science and Technology, Moscow 121205, Russia;* orcid.org/0000-0003-2394-6512

David R. Goodlett − *University of Victoria − Genome BC Proteomics Centre, Victoria, British Columbia V8Z 7X8, Canada; University of Victoria, Victoria, British Columbia V8P 5C2, Canada; University of Gdansk, International Centre for Cancer Vaccine Science, 80-309 Gdansk, Poland*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.0c00961

Notes

The authors declare the following competing financial interest(s): C.H.B. is the Chief Scientific Officer of MRM Proteomics, Inc., the co-founder and Chief Technology Officer of Creative Molecules, Inc. and Chief Technology Officer of Molecular You. The other authors declare no competing financial interests.

■ ACKNOWLEDGMENTS

This work was supported by Genome Canada and Genome British Columbia grants (204PRO, 214PRO, 264PRO, and 282PQP). DRG thanks the International Centre for Cancer Vaccine Science project, which is carried out within the International Research Agendas program of the Foundation for Polish Science cofinanced by the European Union under the European Regional Development Fund (MAB/2017/03), for support. C.H.B. is also grateful for financial support from the Terry Fox Research Institute, for support from the Segal McGill Chair in Molecular Oncology at McGill University (Montreal, Quebec, Canada), and for support from the Warren Y. Soper Charitable Trust and the Alvin Segal Family Foundation to the Jewish General Hospital (Montreal, Quebec, Canada). The study was also supported by the MegaGrant of the Ministry of Science and Higher Education of the Russian Federation (Agreement with Skolkovo Institute of Science and Technology, No. 075-10-2019-083).

■ REFERENCES

(1) Lange, V.; Picotti, P.; Domon, B.; Aebersold, R. Selected reaction monitoring for quantitative proteomics: a tutorial. *Mol. Syst. Biol.* **2008**, *4*, 222.

(2) Kirkpatrick, D. S.; Gerber, S. A.; Gygi, S. P. The absolute quantification strategy: a general procedure for the quantification of proteins and post-translational modifications. *Methods* **2005**, *35* (3), 265−73.

(3) Schiess, R.; Wollscheid, B.; Aebersold, R. Targeted proteomic strategy for clinical biomarker discovery. *Mol. Oncol.* **2009**, *3* (1), 33−44.

(4) Shi, T.; Song, E.; Nie, S.; Rodland, K. D.; Liu, T.; Qian, W. J.; Smith, R. D. Advances in targeted proteomics and applications to biomedical research. *Proteomics* **2016**, *16* (15−16), 2160−82.

(5) Pappireddi, N.; Martin, L.; Wuhr, M. A Review on Quantitative Multiplexed Proteomics. *ChemBioChem* **2019**, *20* (10), 1210−1224.

(6) 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* (11), 1475−88.

(7) Gerber, S. A.; Rush, J.; Stemman, O.; Kirschner, M. W.; Gygi, S. P. Absolute quantification of proteins and phosphoproteins from cell lysates by tandem MS. *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (12), 6940−5.

(8) 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* (8), 1111−21.

(9) Aebersold, R.; Mann, M. Mass spectrometry-based proteomics. *Nature* **2003**, *422* (6928), 198−207.

(10) Keshishian, H.; Addona, T.; Burgess, M.; Mani, D. R.; Shi, X.; Kuhn, E.; Sabatine, M. S.; Gerszten, R. E.; Carr, S. A. Quantification of cardiovascular biomarkers in patient plasma by targeted mass spectrometry and stable isotope dilution. *Mol. Cell. Proteomics* **2009**, *8*, 2339−2349.

(11) Picotti, P.; Bodenmiller, B.; Aebersold, R. Proteomics meets the scientific method. *Nat. Methods* **2013**, *10* (1), 25−27.

(12) Percy, A. J.; Chambers, A. G.; Yang, J.; Hardie, D. B.; Borchers, C. H. Advances in multiplexed MRM-based protein biomarker

quantitation toward clinical utility. *Biochim. Biophys. Acta, Proteins Proteomics* **2014**, *1844* (5), 917−26.
(13) Addona, T. A.; Abbatiello, S. E.; Schilling, B.; Skates, S. J.; Mani, D. R.; Bunk, D. M.; Spiegelman, C. H.; Zimmerman, L. J.; Ham, A. J.; Keshishian, H.; Hall, S. C.; Allen, S.; Blackman, R. K.; Borchers, C. H.; Buck, C.; Cardasis, H. L.; Cusack, M. P.; Dodder, N. G.; Gibson, B. W.; Held, J. M.; Hiltke, T.; Jackson, A.; Johansen, E. B.; Kinsinger, C. R.; Li, J.; Mesri, M.; Neubert, T. A.; Niles, R. K.; Pulsipher, T. C.; Ransohoff, D.; Rodriguez, H.; Rudnick, P. A.; Smith, D.; Tabb, D. L.; Tegeler, T. J.; Variyath, A. M.; Vega-Montoto, L. J.; Wahlander, A.; Waldemarson, S.; Wang, M.; Whiteaker, J. R.; Zhao, L.; Anderson, N. L.; Fisher, S. J.; Liebler, D. C.; Paulovich, A. G.; Regnier, F. E.; Tempst, P.; Carr, S. A. Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat. Biotechnol.* **2009**, *27* (7), 633−41.
(14) Tilburg, J.; Michaud, S. A.; Maracle, C. X.; Versteeg, H. H.; Borchers, C. H.; van Vlijmen, B. J. M.; Mohammed, Y. Plasma Protein Signatures of a Murine Venous Thrombosis Model and Slc44a2 Knockout Mice Using Quantitative-Targeted Proteomics. *Thromb. Haemostasis* **2020**, *120* (3), 423−436.
(15) Mohammed, Y.; van Vlijmen, B. J.; Yang, J.; Percy, A. J.; Palmblad, M.; Borchers, C. H.; Rosendaal, F. R. Multiplexed targeted proteomic assay to assess coagulation factor concentrations and thrombosis-associated cancer. *Blood Adv.* **2017**, *1* (15), 1080−1087.
(16) Burnum-Johnson, K. E.; Nie, S.; Casey, C. P.; Monroe, M. E.; Orton, D. J.; Ibrahim, Y. M.; Gritsenko, M. A.; Clauss, T. R.; Shukla, A. K.; Moore, R. J.; Purvine, S. O.; Shi, T.; Qian, W.; Liu, T.; Baker, E. S.; Smith, R. D. Simultaneous Proteomic Discovery and Targeted Monitoring using Liquid Chromatography, Ion Mobility Spectrometry, and Mass Spectrometry. *Mol. Cell Proteomics* **2016**, *15* (12), 3694−3705.
(17) Liebler, D. C.; Zimmerman, L. J. Targeted quantitation of proteins by mass spectrometry. *Biochemistry* **2013**, *52* (22), 3797−806.
(18) Mohammed, Y.; Domański, D.; Jackson, A. M.; Smith, D. S.; Deelder, A. M.; Palmblad, M.; Borchers, C. H. PeptidePicker: a scientific workflow with web interface for selecting appropriate peptides for targeted proteomics experiments. *J. Proteomics* **2014**, *106*, 151−61.
(19) Perez-Riverol, Y.; Alpi, E.; Wang, R.; Hermjakob, H.; Vizcaino, J. A. Making proteomics data accessible and reusable: current state of proteomics databases and repositories. *Proteomics* **2015**, *15* (5−6), 930−49.
(20) Vizcaino, J. A.; Foster, J. M.; Martens, L. Proteomics data repositories: providing a safe haven for your data and acting as a springboard for further research. *J. Proteomics* **2010**, *73* (11), 2136−46.
(21) Riffle, M.; Eng, J. K. Proteomics data repositories. *Proteomics* **2009**, *9* (20), 4653−63.
(22) Deutsch, E. W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J. J.; Kundu, D. J.; Garcia-Seisdedos, D.; Jarnuczak, A. F.; Hewapathirana, S.; Pullman, B. S.; Wertz, J.; Sun, Z.; Kawano, S.; Okuda, S.; Watanabe, Y.; Hermjakob, H.; MacLean, B.; MacCoss, M. J.; Zhu, Y.; Ishihama, Y.; Vizcaino, J. A. The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* **2020**, *48* (D1), D1145−D1152.
(23) Farrah, T.; Deutsch, E. W.; Kreisberg, R.; Sun, Z.; Campbell, D. S.; Mendoza, L.; Kusebauch, U.; Brusniak, M. Y.; Hüttenhain, R.; Schiess, R.; Selevsek, N.; Aebersold, R.; Moritz, R. L. PASSEL: the PeptideAtlas SRM experiment library. *Proteomics* **2012**, *12* (8), 1170−5.
(24) Clinical Proteomic Tumor Analysis Consortium (CPTAC). 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.; Meyer, M. R.; Mesri, M.; Rodriguez, H.; Paulovich, A. CPTAC Assay Portal: a repository of targeted proteomic assays. *Nat. Methods* **2014**, *11* (7), 703−4.
(25) Sharma, V.; Eckels, J.; Taylor, G. K.; Shulman, N. J.; Stergachis, A. B.; Joyner, S. A.; Yan, P.; Whiteaker, J. R.; Halusa, G. N.; Schilling, B.; Gibson, B. W.; Colangelo, C. M.; Paulovich, A. G.; Carr, S. A.; Jaffe, J. D.; MacCoss, M. J.; MacLean, B. Panorama: a targeted proteomics knowledge base. *J. Proteome Res.* **2014**, *13* (9), 4205−10.
(26) Kusebauch, U.; Campbell, D. S.; Deutsch, E. W.; Chu, C. S.; Spicer, D. A.; Brusniak, M. Y.; Slagel, J.; Sun, Z.; Stevens, J.; Grimes, B.; Shteynberg, D.; Hoopmann, M. R.; Blattmann, P. A.V. R.; Rinner, O.; Picotti, P.; Carapito, C.; Huang, C. Y.; Kapousouz, M.; Lam, H.; Tran, T.; Demir, E.; Aitchison, J. D.; Sander, C.; Hood, L.; Aebersold, R.; Moritz, R. L. Human SRMAtlas: A Resource of Targeted Assays to Quantify the Complete Human Proteome. *Cell* **2016**, *166* (3), 766−778.
(27) Schubert, O. T.; Mouritsen, J.; Ludwig, C.; Rost, H. L.; Rosenberger, G.; Arthur, P. K.; Claassen, M.; Campbell, D. S.; Sun, Z.; Farrah, T.; Gengenbacher, M.; Maiolica, A.; Kaufmann, S. H. E.; Moritz, R. L.; Aebersold, R. The Mtb proteome library: a resource of assays to quantify the complete proteome of Mycobacterium tuberculosis. *Cell Host Microbe* **2013**, *13* (5), 602−612.
(28) Picotti, P.; Clement-Ziza, M.; Lam, H.; Campbell, D. S.; Schmidt, A.; Deutsch, E. W.; Rost, H.; Sun, Z.; Rinner, O.; Reiter, L.; Shen, Q.; Michaelson, J. J.; Frei, A.; Alberti, S.; Kusebauch, U.; Wollscheid, B.; Moritz, R. L.; Beyer, A.; Aebersold, R. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **2013**, *494* (7436), 266−70.
(29) Mohammed, Y.; Bhowmick, P.; Smith, D. S.; Domanski, D.; Jackson, A. M.; Michaud, S. A.; Malchow, S.; Percy, A. J.; Chambers, A. G.; Palmer, A.; Zhang, S.; Sickmann, A.; Borchers, C. H. PeptideTracker: A knowledge base for collecting and storing information on protein concentrations in biological tissues. *Proteomics* **2017**, DOI: 10.1002/pmic.201600210.
(30) Bhowmick, P.; Mohammed, Y.; Borchers, C. H. MRMAssayDB: an integrated resource for validated targeted proteomics assays. *Bioinformatics* **2018**, *34* (20), 3566−3571.
(31) 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.; Meyer, M. R.; Mesri, M.; Boja, E.; Carr, S. A.; Chan, D. W.; Chen, X.; Chen, J.; Davies, S. R.; Ellis, M. J.; Fenyö, D.; Hiltke, T.; Ketchum, K. A.; Kinsinger, C.; Kuhn, E.; Liebler, D. C.; Liu, T.; Loss, M.; MacCoss, M. J.; Qian, W. J.; Rivers, R.; Rodland, K. D.; Ruggles, K. V.; Scott, M. G.; Smith, R. D.; Thomas, S.; Townsend, R. R.; Whiteley, G.; Wu, C.; Zhang, H.; Zhang, Z.; Rodriguez, H.; Paulovich, A. G. Using the CPTAC Assay Portal to Identify and Implement Highly Characterized Targeted Proteomics Assays. *Methods Mol. Biol.* **2016**, *1410*, 223−36.
(32) 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* (7), 966−8.
(33) UniProt_Consortium. Ongoing and future developments at the Universal Protein Resource. *Nucleic Acids Res.* **2011**, *39*, D214.
(34) Rodchenkov, I.; Babur, O.; Luna, A.; Aksoy, B. A.; Wong, J. V.; Fong, D.; Franz, M.; Siper, M. C.; Cheung, M.; Wrana, M.; Mistry, H.; Mosier, L.; Dlin, J.; Wen, Q.; O'Callaghan, C.; Li, W.; Elder, G.; Smith, P. T.; Dallago, C.; Cerami, E.; Gross, B.; Dogrusoz, U.; Demir, E.; Bader, G. D.; Sander, C. Pathway Commons 2019 Update: integration, analysis and exploration of pathway data. *Nucleic Acids Res.* **2020**, *48* (D1), D489−D497.
(35) Cerami, E. G.; Gross, B. E.; Demir, E.; Rodchenkov, I.; Babur, O.; Anwar, N.; Schultz, N.; Bader, G. D.; Sander, C. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **2011**, *39*, D685.
(36) Kanehisa, M.; Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **2000**, *28* (1), 27−30.
(37) Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N. T.; Morris, J. H.; Bork, P.; Jensen, L. J.; Mering, C. V. STRING v11: protein-protein association networks with increased coverage, supporting functional

https://dx.doi.org/10.1021/acs.jproteome.0c00961
J. Proteome Res. 2021, 20, 2105−2115

discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47* (D1), D607−D613.
(38) Pinero, J.; Bravo, A.; Queralt-Rosinach, N.; Gutierrez-Sacristan, A.; Deu-Pons, J.; Centeno, E.; Garcia-Garcia, J.; Sanz, F.; Furlong, L. I. DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants. *Nucleic Acids Res.* **2017**, *45* (D1), D833−D839.
(39) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46* (D1), D1074−D1082.
(40) Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; Feng, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. The Protein Data Bank. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2002**, *58* (6), 899−907.
(41) Hoksza, D.; Gawron, P.; Ostaszewski, M.; Schneider, R. MolArt: a molecular structure annotation and visualization tool. *Bioinformatics* **2018**, *34* (23), 4127−4128.
(42) UniProt Consortium. Watkins, X.; Garcia, L. J.; Pundir, S.; Martin, M. J. ProtVista: visualization of protein sequence annotations. *Bioinformatics* **2017**, *33* (13), 2040−2041.
(43) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* **2019**, *47* (D1), D1102−D1109.
(44) Wang, Y.; Xiao, J.; Suzek, T. O.; Zhang, J.; Wang, J.; Bryant, S. H. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res.* **2009**, *37*, W623.
(45) Franz, M.; Lopes, C. T.; Huck, G.; Dong, Y.; Sumer, O.; Bader, G. D. Cytoscape.js: a graph theory library for visualisation and analysis. *Bioinformatics* **2016**, *32* (2), 309−311.
(46) Binns, D.; Dimmer, E.; Huntley, R.; Barrell, D.; O'Donovan, C.; Apweiler, R. QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **2009**, *25* (22), 3045−6.
(47) Sehnal, D.; Deshpande, M.; Varekova, R. S.; Mir, S.; Berka, K.; Midlik, A.; Pravda, L.; Velankar, S.; Koca, J. LiteMol suite: interactive web-based visualization of large-scale macromolecular structure data. *Nat. Methods* **2017**, *14* (12), 1121−1122.
(48) Nehrt, N. L.; Peterson, T. A.; Park, D.; Kann, M. G. Domain landscapes of somatic mutations in cancer. *BMC Genomics* **2012**, *13* (Suppl 4), S9.
(49) Lam, S. D.; Dawson, N. L.; Das, S.; Sillitoe, I.; Ashford, P.; Lee, D.; Lehtinen, S.; Orengo, C. A.; Lees, J. G. Gene3D: expanding the utility of domain assignments. *Nucleic Acids Res.* **2016**, *44* (D1), D404−9.
(50) Gene Ontology Consortium. Creating the gene ontology resource: design and implementation. *Genome Res.* **2011**, *11* (8), 1425−1433.
(51) Mohammed, Y.; Kootte, R. S.; Kopatz, W. F.; Borchers, C. H.; Buller, H. R.; Versteeg, H. H.; Nieuwdorp, M.; van Mens, T. E. The intestinal microbiome potentially affects thrombin generation in human subjects. *J. Thromb. Haemostasis* **2020**, *18* (3), 642−650.
(52) Mohammed, Y.; Borchers, C. H. An extensive library of surrogate peptides for all human proteins. *J. Proteomics* **2015**, *129*, 93−97.

https://dx.doi.org/10.1021/acs.jproteome.0c00961
J. Proteome Res. 2021, 20, 2105−2115