Journal of proteome research

pubs.acs.org/jpr

Article

Mass Spectrometric Detected Cancer Proteins as Resources for Cancer Research

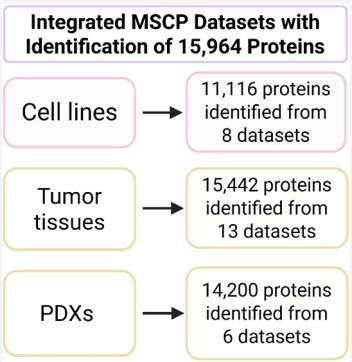
Yuanyu Huang,* Lijun Chen, Peter I-Fan Wu, Poorva Juneja, Li Chen, Yvonne A. Evrard, Liyuan Jiao, Yingwei Hu, Xu Zhang, James H. Doroshow, Ratna R. Thangudu, Alexander Pilozzi, and Hui Zhang

Cite This: *J. Proteome Res.* 2026, 25, 2901−2914

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Protein evidence derived from mass spectrometry (MS) across cancer cohorts and model systems is extensive but remains fragmented across individual studies and repositories, limiting rapid retrieval and evidence-based benchmarking of cancer-context protein detection. Here we present the Mass Spectrometric Detected Cancer Proteins (MSCP) resource, an integrated database assembled from 27 large-scale cancer proteomics sources spanning human tumor cohorts, cancer cell lines, and patient-derived xenograft (PDX) models. Protein identifications were harmonized to UniProtKB-Swiss-Prot (release 2025_01) and integrated under FDR-controlled identification outputs to generate a unified catalog of 15,964 MS-supported human proteins. Benchmarking against neXtProt PE1 identified 525 proteins newly supported by MS evidence in the integrated cancer context, including proteins previously associated with chromosome-level evidence inconsistencies. Functional interpretation of the newly identified set using GO and Reactome enrichment highlighted immune- and barrier-associated processes and chromatin- and genome-regulatory pathways, including DNA methylation and histone deacetylation. Orthogonal verification using synthetic unique peptides confirmed representative newly identified proteins by concordant precursor m/z and fragment-ion patterns. MSCP provides a provenance-aware, UniProtKB-aligned resource for cancer proteomics that supports both cohort- and model-specific querying and coverage-oriented evidence aggregation, enabling standardized comparisons to reference proteomes and facilitating downstream assay planning and translational studies.



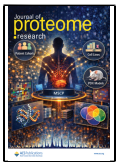
KEYWORDS: *mass spectrometry, cancer proteins, proteomics database, NeXtProt protein existence (PE) metrics, missing proteins, peptide identification*

INTRODUCTION

The primary goal of proteomics is to characterize proteins and advance methodologies that enable protein detection from different biological resources.[1−3] Historically, the field of proteomics has evolved significantly with the advent of high-throughput mass spectrometry-based proteomics, antibody development, and protein microarrays, enabling the rapid and accurate identification of proteins.[4−6] Advancements in mass spectrometry have transformed the field by providing greater sensitivity, resolution, and throughput.[7−9] Proteomics databases have played a pivotal role in this evolution by serving as repositories for data derived from diverse experimental and computational workflows.[10,11] Initiatives such as the UniProt Knowledgebase (UniProtKB), Proteomics Identifications archive database (PRIDE), and PeptideAtlas have set the groundwork for systematic protein cataloging, allowing researchers to explore complex proteomes and validate findings against established data sets.[12−26] These databases not only store data but also facilitate the annotation, integration, and cross-referencing of proteomic information, thereby accelerating discoveries in biology and medicine.[27−32] The UniProtKB is known for its comprehensive protein annotations, providing essential resources for researchers, while the Mass Spectrometry Interactive Virtual Environment Knowledge Base (MassIVE-KB) supports large-scale data integration and quantitation by computational analysis tools.[33] These platforms collectively enhance the accessibility and utility of proteomics data, enabling the detailed exploration of protein expression and functionality. For the documentation of inter- and intra-individual diversity of human proteins at proteomic levels, the neXtProt database has compiled a reviewed corpus of information available in UniProtKB that provides the groundwork.[34]

Mass spectrometric characterization of proteins is instrumental in achieving these objectives by offering a structured repository of protein data that integrates diverse applications of tissue and cancer research.[16,17,35,36] Despite the availability of major proteomics resources (UniProtKB, neXtProt, PRIDE/ProteomeXchange, PeptideAtlas, and MassIVE-KB), MS-

Received: December 18, 2025
Revised: March 17, 2026
Accepted: April 14, 2026
Published: May 1, 2026



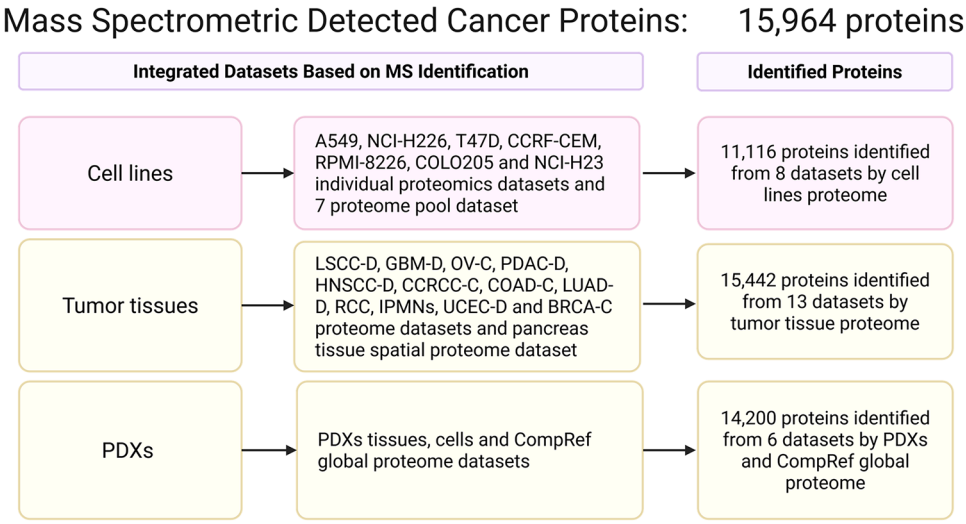


Figure 1. Overview of MSCP construction and coverage across cancer models. MSCP integrates MS-based protein identifications from major cancer-related sample types, including human tumor cohorts, cancer cell lines, and patient-derived xenograft (PDX) models, resulting in a unified catalog of 15,964 UniProtKB-mapped human proteins. The schematic summarizes the contributing cohorts and models with the integration outcome, highlighting MSCP as a cancer-context protein evidence resource derived from multiple independent proteomics sources.

confirmed cancer proteins remain dispersed across individual studies and repositories, often without harmonized identifiers, unified cohort provenance, or cancer-centric aggregation. As a result, researchers conducting cancer proteomics, proteogenomics, and biomarker discovery frequently must manually assemble and reconcile MS-evidence protein lists across cohorts, acquisition modes, and database versions. Here we describe the development of mass spectrometric detected cancer proteins (MSCP) based on mass spectrometry–based protein identifications obtained from cancer-derived cell lines, patient-derived xenograft models (PDXs), and human tumor samples with data sets and benchmarked against neXtProt PE1 to quantify coverage gains and newly detected proteins, with additional chromosome analyses to demonstrate utility. MSCP was developed to address this gap by integrating MS-confirmed protein evidence from large-scale cancer cohorts and model systems into a unified, traceable, and UniProtKB-aligned resource, enabling a standardized comparison to neXtProt PE metrics and supporting downstream annotation and biological context analyses. MSCP has facilitated the detection of 15,964 proteins by mass spectrometry in the human proteome, with 525 newly identified proteins. These advancements are supported by database establishment for mass spectrometry data mining, annotation, and validation by synthetic unique sequences, assessed by protein existence and organ-specific localization analysis.

METHODS

Data Sources and Retrieval

Proteomics data sets used to construct MSCP were collected from publicly accessible cancer proteomics resources spanning tumor cohorts, cancer cell lines, and patient-derived xenograft (PDX) models. For each of the 27 sources, we retrieved the deposited high-confidence identification outputs (processed peptide and protein identification tables) from the National Cancer Institute Proteomic Data Commons (PDC) and ProteomeXchange/PRIDE repositories (accession numbers are listed in the Data Accessibility section). When multiple result files were available per study in different batches or acquisition modes, we used the corresponding final-study-level identification outputs to represent that source.

Ethics Approval and Informed Consent

Cited cancer proteomics data sets included in MSCP were derived from previously published studies and public repositories in accordance with the applicable institutional approvals and repository/CPTAC data-use guidelines. Institutional review boards at the tissue source sites reviewed protocols and informed consent documentation in accordance with CPTAC guidelines.

Identifier Harmonization and Integration Strategy

To enable consistent cross-source comparison and to minimize version drift, all protein identifiers from contributing sources were harmonized to the UniProtKB-Swiss-Prot human reference proteome (release 2025_01). Protein accessions were mapped to current UniProtKB primary accessions by using UniProt-provided accession history and cross-references when needed. To reduce redundancy across heterogeneous inference workflows, MSCP was represented in a gene-centric UniProt entry framework, in which proteoforms/isoforms reported by different studies were consolidated under the corresponding UniProtKB entry where appropriate. After harmonization, MSCP was constructed by using a union-based integration scheme, where each UniProtKB entry was included if it met the identification criteria described below in at least one contributing source. For transparency and downstream subsetting, MSCP retains source-resolved evidence: each entry is annotated with the set of contributing sources, enabling users to generate model- or cohort-specific subsets in addition to the full integrated catalog.

Peptide/Protein Identification Criteria and FDR Control

MSCP integrates protein identifications that were generated under the FDR-controlled identification pipelines of the originating data sets. Specifically, for each contributing source, we retained proteins supported by peptide-spectrum match (PSM) and peptide/protein identifications controlled at ≤1% false discovery rate (FDR), as reported by the original study processing workflows and/or repository quality control. When a data set provided multiple filtered result tiers, we used the high-confidence identification tier consistent with study reporting. To standardize interpretability across heterogeneous studies, we define an "identified protein" in MSCP as a UniProtKB-Swiss-Prot entry supported by the source's high-confidence identification output. This definition is intentionally distinct from publication-specific "quantified proteins," which in some studies may reflect additional filters beyond identification confidence, such as quantitative completeness across samples, minimum nonmissingness, and cohort-specific reporting thresholds. Accordingly, per-source

https://doi.org/10.1021/acs.jproteome.5c01259
J. Proteome Res. 2026, 25, 2901−2914

protein totals in MSCP reflect identification evidence, while MSCP also provides source proof to support stricter, user-defined criteria such as recurrence across multiple sources or restriction to specific cohorts. The newly identified proteins were verified by the identified peptides from the protein unique sequences using synthetic peptides from Complete Omics.

RESULTS

Database Construction

MSCP integrates data from diverse sources, including NCI-7 cell lines and tumor proteomes (Figure 1). Different data acquisition modes, mass spectrometry, and quantification strategies from different cancer data sources contributed to various sample types and were collected for the database construction (Table S1). The comprehensive overview of the integration MSCP pipeline highlights the NCI-7 proteome, derived from cell lines such as A549 and CCRF-CEM, which contributed 11,116 identified proteins from eight data sources. Similarly, tumor proteome data from CPTAC cohorts (Table 1) and PDXs with CompRef data sources added 15,442 and 14,200 proteins, respectively.[16,17,37−47] These contributions collectively enabled the identification of 15,964 proteins. The balanced integration of diverse data sets and the application of stringent quality control measures ensure robust and reliable database construction. Importantly, MSCP preserves data set-level provenance at the protein-entry level. In Supporting Table S1, each protein is annotated with binary identification indicators (1/NA) across the 27 contributing sources, where each source column is explicitly labeled by cohort/model and acquisition mode (DDA/DIA), which enables users to generate model-specific or cohort-specific subsets in addition to the integrated view. To normalize identification confidence across heterogeneous sources, all protein identifications were mapped to UniProtKB-Swiss-Prot (release 2025_01) and integrated using consistent high-confidence inclusion criteria. Specifically, MSCP retains protein entries supported by peptide- and protein-level identifications controlled at ≤1% FDR as defined in the originating data set pipelines and repository QC. To improve comparability of protein inference across studies and acquisition strategies, evidence is represented in a gene-centric manner with consistent peptide evidence rules, and each entry retains source provenance, including acquisition mode (DDA/DIA), enabling users to apply additional stratified stringency, such as cohort-restricted or recurrence-based filters as needed.

Table 1. CPTAC Tumor Proteome Cohorts Included in MSCP[a]

cancer types	cancer ID	samples	refs
CPTAC − Pancreatic Ductal Adenocarcinoma Discovery Study	PDAC-D	140	16
CPTAC − Glioblastoma Discovery Study	GBM-D	99	17
CPTAC - Lung Squamous Cell Carcinoma Discovery Study	LSCC-D	108	18
CPTAC − Ovarian Cancer Confirmatory Study	OV−C	83	19
CPTAC − Head and Neck Squamous Cell Carcinoma Discovery Study	HNSCC-D	109	20
CPTAC − Clear Cell Renal Cell Carcinoma Discovery Cohort	CCRCC-D	110	21
CPTAC − Colon Adenocarcinoma Confirmatory Study	COAD-C	97	22
CPTAC − Lung Adenocarcinoma Discovery Study	LUAD-D	110	23
CPTAC − Uterine Corpus Endometrial Carcinoma Discovery Study	UCEC-D	95	24
CPTAC − Breast Cancer Prospective Cohort	BRCA-C	122	25
CPTAC − Renal Cell Carcinomas Combined Study	RCC	456	26
Intraductal Papillary Mucinous Neoplasms Study	IPMNs	155	48

[a]List of CPTAC tumor proteome cohorts incorporated into MSCP, including cancer type, cohort identifier, number of samples, and primary reference. These cohorts contribute tumor-derived MS evidence that complements cancer cell line and PDX sources in MSCP and support cohort-level interpretation of protein detection within MSCP.

Database Coverage and Distribution

To evaluate MSCP coverage, we mapped all MSCP protein entries (n = 15,964) supported by identified peptides to UniProtKB (release 2025_01) and quantified the extent of overlap with the reference human proteome. The data sets showcased in Figure 2 illustrate a systematic approach to expanding protein identification, contributing to enriching the database with insights into previously uncharacterized proteins. The contribution of each data set was exhibited by the barplot (Figure 2A), with notable inputs from sources like the Prospective Breast Cancer Proteome and PDX proteome, both exceeding 12,000 identified proteins. These contributions also covered different previously uncharacterized proteins, as the newly identified proteins in MSCP. The quantitative trends are highlighted in Figure 2B, where the addition of data sets resulted in a consistent increase in the protein identification. To quantify the contribution of each data set beyond the cumulative trends, we computed source-by-source incremental gains in both total identified proteins and newly detected proteins (defined relative to neXtProt PE1) as each data source was added to the merged MSCP catalog, and we report these values in Supporting Table S2. The approximately steady incremental increase arises because each additional cancer cohort/model contributes a highly overlapping core of well-established proteins and a small but persistent set of proteins outside neXtProt PE1 that become detectable across heterogeneous sample types and acquisition strategies (DDA/DIA). In addition, the apparent smoothness of the incremental curve is influenced by data set ordering; therefore, per-source incremental values are provided to make individual data set contributions directly interpretable independent of plotting order. For each contributing study, MSCP was generated from the final high-confidence protein identification outputs deposited in PDC/PRIDE with processed identification tables after mapping all entries to a consistent UniProtKB reference release (2025_01). Here, "identified proteins" refers to proteins supported by the study's standard identification criteria, whereas some original publications may report a smaller "quantified protein" subset after additional cohort-specific filtering, such as quantitative completeness across samples, which can lead to different headline counts for the same cohort.

As in UniProtKB, all of the proteoforms arising from a given protein-coding gene are usually described in a single entry, which is complemented with proteins existence evidence with categories from 1 to 5: evidence of protein with MS or non-MS methods (PE1), transcript expression (PE2), homology to other species (PE3), gene models (PE4), and uncertain genes

https://doi.org/10.1021/acs.jproteome.5c01259
J. Proteome Res. 2026, 25, 2901−2914

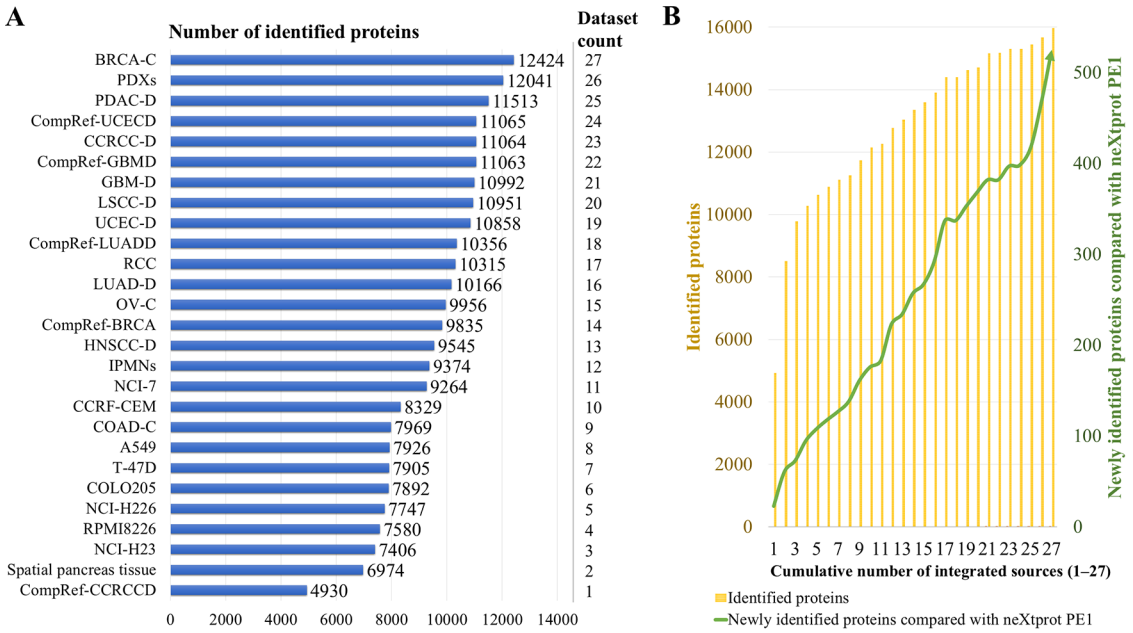


Figure 2. Protein identifications across individual sources and cumulative integration. (A) Number of proteins identified in each of the 27 data sources included in MSCP (see Supporting Table S1 for source definitions). (B) Cumulative union of identified proteins after sequential addition of data sources (x-axis = cumulative number of integrated sources, 1−27). Yellow bars indicate the cumulative total proteins in the merged MSCP catalog, and the green line indicates the incremental newly identified proteins added by each source relative to neXtProt PE1.

along with duplicates (PE5).[49] In the final neXtProt release (2023−09), there are 20,389 entries from the human proteome with different protein existence information depending on the MS-evidence experiment.

Notably, newly identified proteins compared to the neXtProt PE1 proteome increased steadily, with significant contributions from each of the proteomics data sets. The increasing trend underscored the importance of diverse data set integration and took advantage of both data-dependent acquisition (DDA) and data-independent acquisition (DIA) modes. The MSCP not only expanded proteome coverage through integration of diverse cancer data sets but also enabled in-depth protein existence analysis by linking cancer proteins to chromosomal distribution and evidence status. MSCP also supports both integrated and stratified use. For system-focused analyses, users can filter the Supporting Table S1 by selecting only the relevant source columns (e.g., specific CPTAC cohorts, PDX data sets, or NCI-7 cell line data sets) and optionally restricting to a single acquisition mode (DDA or DIA). For coverage-oriented applications, users can use the integrated view or require repeated detection across multiple independent sources to prioritize proteins with broader supporting evidence.

Protein Existence Analysis by Chromosome

Protein existence status indicates the type of evidence that supports the existence of the protein and the data reliability in the database, especially for the analysis by chromosome.[34] The distribution and coverage of proteins across chromosomes highlight the stability of characterization and discrepancies in neXtProt. A detailed visualization of protein existence on the chromosome was provided to exhibit chromosome-specific coverage with quantification results (Figure 3). The sunburst plot in Figure 3A highlights the relative abundance of proteins identified from MSCP and neXtProt across different chromosomes (Figure 3B), providing a summary of coverage completeness and unbiased protein identification by chromosome. By quantifying the extent of protein identification on each chromosome, we showcase areas of higher detection and coverage improvement (Figure 3B). The overlap and novelty of 525 proteins identified in MSCP were also compared to previously annotated chromosome-false proteins, which are gene entries later reassigned or removed due to incorrect genomic localization or evidence of inconsistencies (Figure 3C). This illustrated MSCP's ability to address the gaps and cover more uncharacterized proteins, including 37 chromosome-false proteins, which are gene entries later reassigned or removed due to incorrect genomic localization or evidence inconsistencies. The tree map further highlighted the locations of 525 proteins previously categorized as missing proteins, now identified, with 85% of them located at the membrane, nucleus, or extracellular region components, which belong to the categories of proteins as difficult to detect (Figure 3D). Together, the protein existence analysis by chromosome underscored the importance of integrating multisource data sets to improve proteomics annotations in existing databases.

Database Update with Newly Identified Proteins

Combining and categorizing the analysis acquisition modes by DDA and DIA, new proteins were identified from NCI-7 and tumor proteomics data sets, providing a comprehensive visualization of the database update. The heatmap in Figure 4A demonstrates the distribution and categorization of 51 of 525 newly identified proteins' shared identification by at least 9 different data sources across two different acquisition modes, highlighting the complementary DDA and DIA approaches in achieving broad coverage. Notably, the DIA acquisition mode contributed more newly identified proteins than DDA. Recently, the Human Proteome Project (HPP) has been aligning protein lists from UniProtKB to the human genome annotation in the encyclopedia of DNA elements project (GENCODE) database, which is also linked to UniProtKB, including the protein existence information.[49] Thus, the coverage of newly identified proteins was further explored, including those not presented in the GENCODE list of

https://doi.org/10.1021/acs.jproteome.5c01259
J. Proteome Res. 2026, 25, 2901−2914

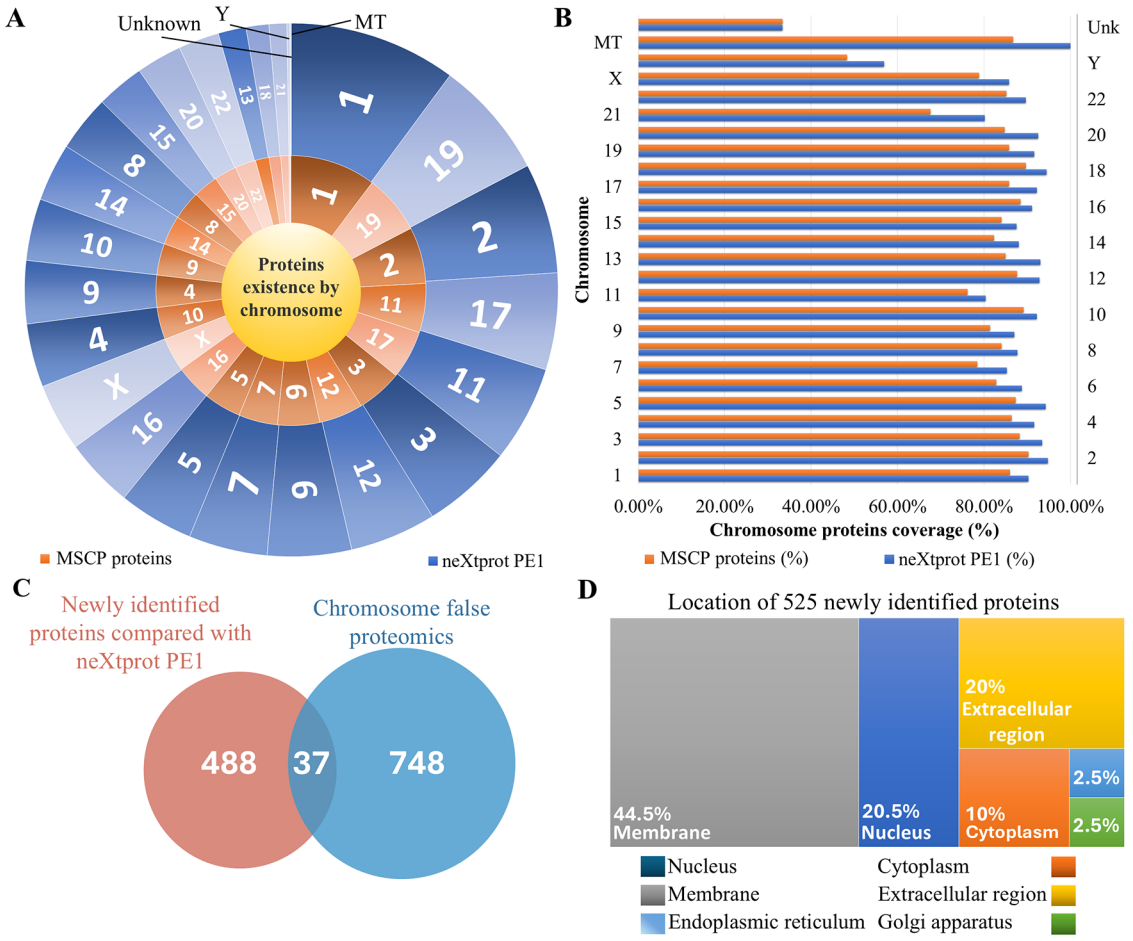


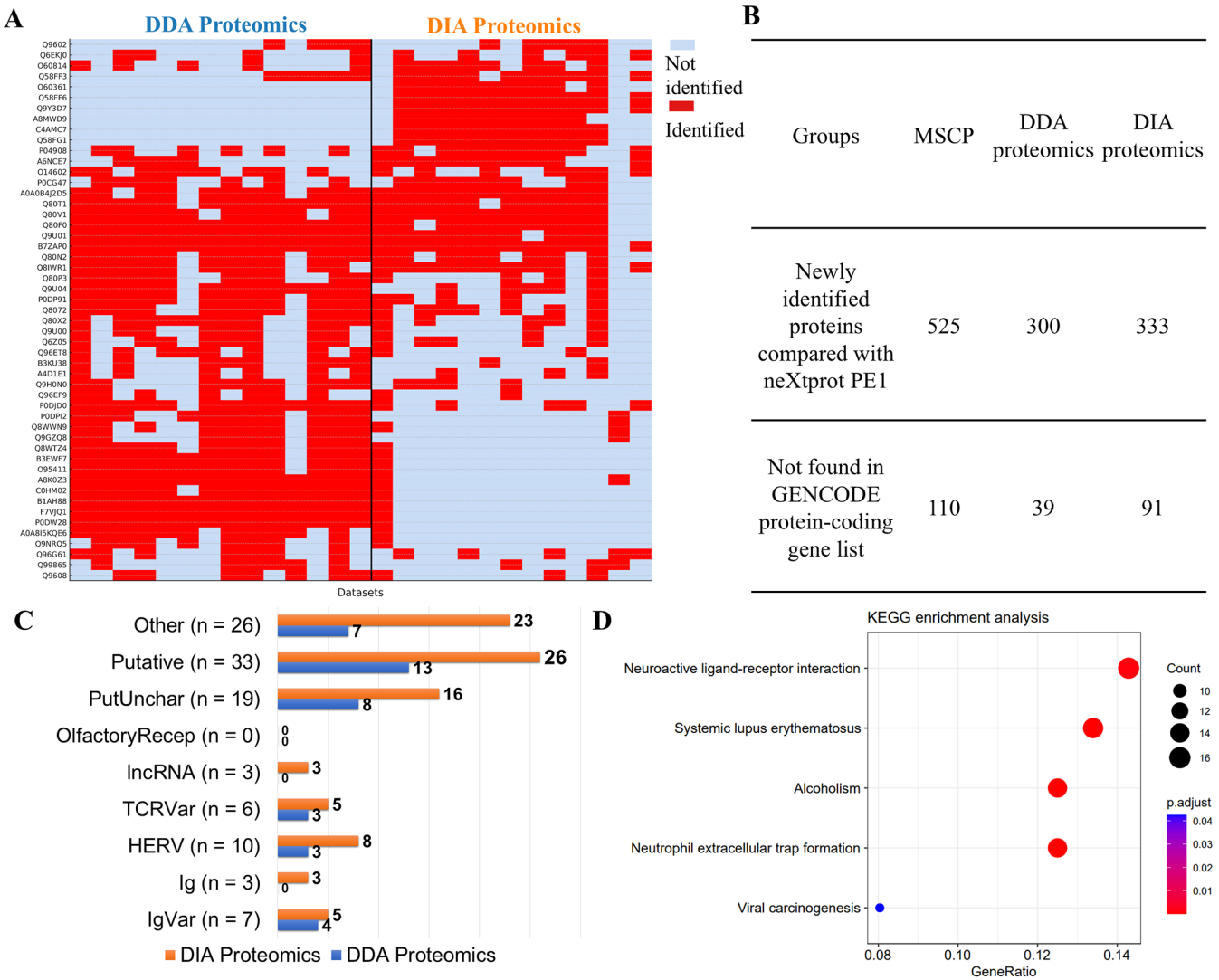
Figure 3. Protein existence benchmarking and chromosome-level coverage of MSCP. (A) Sunburst visualization of protein distributions across human chromosomes for MSCP and the neXtProt PE1 reference set (PE1: protein-level evidence supported by experimental observation, including MS and non-MS evidence). (B) Chromosome-by-chromosome comparison of the number of proteins detected in MSCP versus neXtProt PE1, highlighting chromosome-level coverage and potential gaps addressed by MSCP integration. (C) Overlap of 525 newly identified MSCP proteins (defined relative to neXtProt PE1) with proteins previously annotated as chromosome-false entries, illustrating recovery of proteins with previously inconsistent genomic/evidence annotations. (D) Cellular component distribution of newly identified proteins by tree map, emphasizing enrichment in challenging classes such as membrane-associated, nuclear, and extracellular proteins.

protein-coding genes (based on GENCODEv46, Figure 4B,C). The stacked barplot reveals patterns in protein categories covered by subproteomics in MSCP, including 39 by DDA mode identification and 91 by DIA mode identification. In total, 110 proteins were covered in the Ensembl-GENCODE not-found list of protein-coding genes (1254 UniProtKB-Swiss-Prot entries[49]). The contributions of various proteomics sources and acquisition modes demonstrated the synergy between different approaches in uncovering novel proteins and enhancing database completeness. Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis in Figure 4D emphasizes the enrichment of membrane-related proteins with the neuroactive ligand−receptor interaction pathway covered by MSCP. The same conclusion was also drawn by the protein existence analysis on chromosome-false proteomics, which indicated the enhanced MS detection of previously challenging protein classes, such as the proteins located at membranes, reinforcing the robustness of the data set.

The contribution of newly identified proteins in the database was compared to the human organ and tissue proteome data from The Human Protein Atlas. The top organs from humans with different specific proteins were covered by the database and contributed most to the testis (Figure 5A). The most enriched protein interaction networks were related to previously missing proteins and RNA-binding proteins (Figure 5B), which benefited from the low bias to chromosomes and the high coverage of the MSCP.

Combined with all of the newly identified proteins in MSCP, more proteins with MS-evidence proved the existence of cancer proteins by the proteomics tools. All of the categorized tumor proteomics data sources have contributed to the identification and verification of them. To clarify the positioning of MSCP relative to established community resources, we summarize below how MSCP complements existing databases and what unique utility it provides for cancer proteomics (Supporting Table S3). MSCP focuses on mass spectrometry (MS)-confirmed protein detection in cancer contexts, integrating evidence across tumor cohorts, cancer cell lines, and PDX models into a single UniProtKB-mapped catalog with explicit source provenance. In contrast, CPTAC/PDC primarily organizes and disseminates cohort- and study-specific proteogenomic data sets (raw and processed), enabling deep within-cohort analyses but not inherently providing a unified, cross-cohort cancer protein presence compendium. neXtProt provides a curated reference framework centered on protein existence (PE) evidence and rich annotation for the

https://doi.org/10.1021/acs.jproteome.5c01259
J. Proteome Res. 2026, 25, 2901−2914



Groups	MSCP	DDA proteomics	DIA proteomics
Newly identified proteins compared with neXtprot PE1	525	300	333
Not found in GENCODE protein-coding gene list	110	39	91

Figure 4. Characterization of newly identified proteins and acquisition-mode contributions. (A) Heatmap showing detection patterns of a representative subset of newly identified proteins across contributing sources, stratified by acquisition mode (DDA vs DIA); proteins shown are shared across multiple sources to illustrate cross-study recurrence. (B) Summary of newly identified proteins by acquisition mode, indicating the relative contributions of DDA and DIA data sets. (C) Classification of newly identified proteins with respect to the Ensembl/GENCODE protein-coding gene list (GENCODEv46), including proteins absent from the protein-coding gene list and their distribution across acquisition modes. (D) KEGG pathway enrichment analysis of newly identified proteins, highlighting functional themes associated with proteins newly supported by MS evidence in the integrated MSCP cancer context.

human proteome, but it is not designed as a cancer-model aggregation of MS-detected proteins with per-cohort provenance. The Human Protein Atlas (HPA) emphasizes tissue and organ expression and localization evidence, offering complementary biological interpretation rather than a cancer MS-identification integration layer.

The key contribution of MSCP is therefore 3-fold. First, it offers a cancer-context MS-evidence union catalog (15,964 proteins in total) that supports rapid retrieval of proteins observed by MS in cancer-related biological systems. Second, by mapping entries to a consistent UniProtKB reference release and benchmarking against neXtProt PE1, MSCP enables standardized quantification of incremental coverage gains and identification of proteins newly supported by MS evidence in an integrated cancer context. Third, MSCP retains data set-level provenance (source cohort and acquisition mode), enabling users to conduct either system-focused queries or coverage-oriented queries. Together, these features provide an interpretable, provenance-aware integration layer that complements CPTAC/PDC archival functions, neXtProt PE-centric reference annotation, and HPA biological context, thereby supporting cancer proteomics study design, evidence triage, and downstream assay planning in a transparent manner.

A total of 443 of the newly identified proteins were further verified by using synthetic peptides with their uniquely identified sequences that share none of the common sequences of other proteins in the human proteome. Such as ATP-binding cassette subfamily A member 10 (ABCAA_HUMAN), its unique sequence EVEQEVK was synthesized and found in both MS1 and MS2 data from the CPTAC-Breast cancer prospective cohort by Fusion DDA mode, PDC000120. The precursors in MS1 were matched, while in MS2, fragments b and y ions uniquely existed individually, but we still could match the most highly abundant fragments and then verified the existence of ABCAA_HUMAN (Figure 6A,B). Due to the

https://doi.org/10.1021/acs.jproteome.5c01259
J. Proteome Res. 2026, 25, 2901−2914

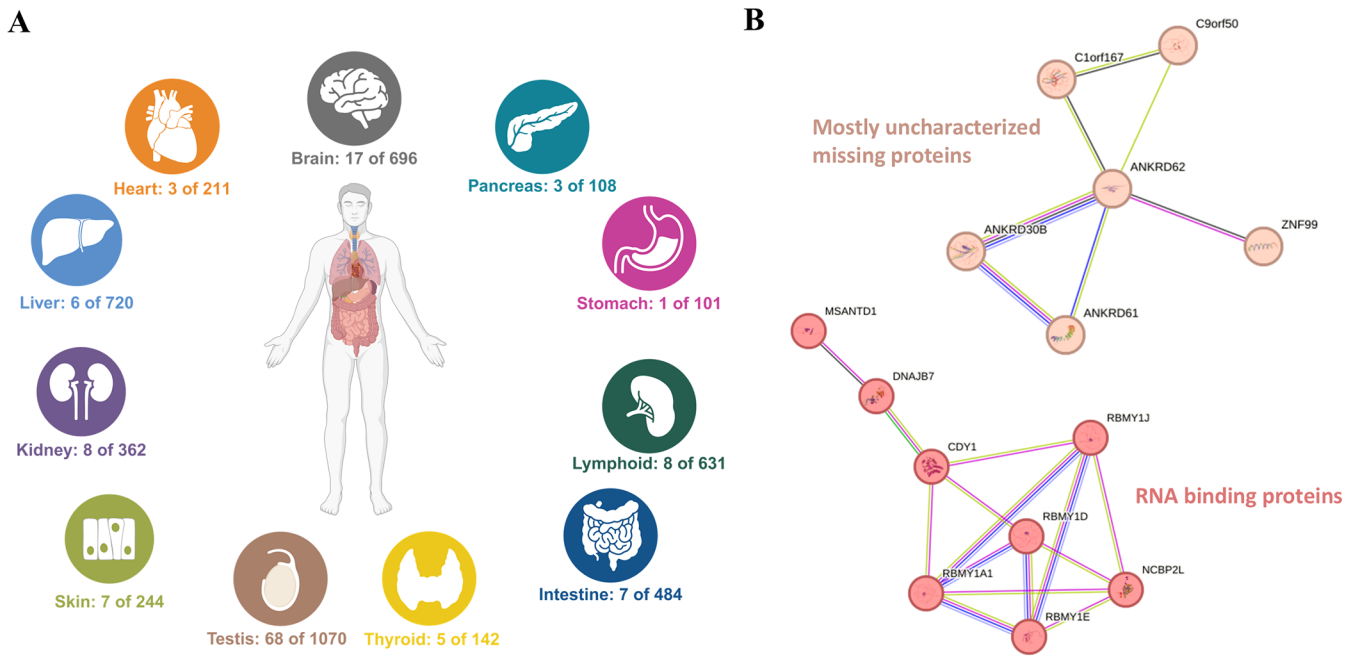


Figure 5. Biological contextualization of newly identified proteins using Human Protein Atlas annotations. (A) Contribution of newly identified MSCP proteins to organ/tissue-specific protein sets curated by the Human Protein Atlas, summarizing which organs receive the incremental coverage from MSCP newly identified proteins. (B) Enrichment of protein−protein interaction modules among newly identified proteins, highlighting functional groupings supported by the expanded MS evidence captured in MSCP.

TMT10 (tandem mass tag) quantification labeling strategy from Prospective Breast BI Proteome, both precursors with the isotopic peaks in the MS1 spectrum and y fragments in the MS2 spectrum showed an m/z increase of 229.163 ($+H_{20}C_8{}^{13}C_4{}^{15}NNO_2$). After the correlation by this shift, two precursor peaks and 5 y fragments peaks could be matched to the sequence EVEQEVK. As an ATP-Binding cassette (ABC) transporter protein, it uses ATP hydrolysis to move substrates across lipid membranes.[50] This multipass membrane protein contained MS-evidence information in MSCP with 2 cancer data sources (Supporting Table S4). Similarly, the uncharacterized protein C12orf56 (CL056_HUMAN) with the unique sequence YVVLSDR was identified in the data from CPTAC-Lung adenocarcinoma by QE HF-X DDA mode with both MS1 and MS2 spectra matching with the synthetic sequence (PDC000219, Figure 6C,D). Notably, the same TMT10-quantification labeling strategy generated different matching principles for sequences ending with lysine and arginine. Compared to the 2 TMT labeling sites on the sequence EVEQEVK, the sequence YVVLSDR was only labeled with a single TMT tag at the N-terminal. Thus, the m/z shift in the MS1 spectrum was 114.582 (z = +2) while there was no shift for fragments y ions in the MS2 spectrum. In total, three precursor peaks and 5 y fragments peaks could be matched to the sequence YVVLSDR. The MS-evidence of CL056_HUMAN was collected in 3 cancer data sources of MSCP, which has the potential to contribute to the limited research in mostly uncharacterized missing proteins.[51] Different from TMT-labeled DDA data, the label-free quantification (LFQ) method with DIA mode generated spectra that directly matched with the synthetic sequence without any m/z shift in both MS1 and MS2 spectra. In the NCI-7 proteome results, individual samples and seven-pool samples were all identified by the LFQ method with the DIA mode. From these results, a putative tubulin-like protein α-4B (TBA4B_HUMAN) with the unique sequence QDPSQPLSR was identified by RPMI8226 cell lysate with the TIMS-TOF DIA mode. Four precursor peaks and seven fragment peaks were matched to the sequence QDPSQPLSR (Figure 6E,F). The MS-evidence of TBA4B_HUMAN was collected in two NCI-7 data sources and 5 cancer data sources of MSCP. Not limited to the cancer proteins database, TBA4B_HUMAN was also recently identified by proteomics method in human lens epithelium cells,[52] indicating the database coverage contributed by the diversity of sample types.

Combined with all of the newly identified proteins in MSCP, more proteins with MS-evidence proved the cancer proteins' existence by the proteomics tools. In total, 23 newly identified proteins were verified with their synthetic unique peptide sequences in both MS1 spectra and MS2 fragments (Supporting Table S4). All of the categorized tumor proteomics data sources have contributed to the identification and verification of them.

Beyond expanding proteome coverage, an important question is what biological themes are represented among the 525 newly identified proteins (relative to neXtProt PE1). Functional enrichment analysis (Supporting Table S5) indicates that these proteins are not randomly distributed across cellular processes but instead cluster into pathways and compartments that are both biologically coherent and technically consistent with classes that are challenging to detect in single-cohort proteomics studies. GO biological process enrichment highlights immune and host−environment interface biology, including innate immune response in mucosa, mucosal immune response, and the antibacterial humoral response. These enrichments suggest that a subset of newly identified proteins corresponds to specialized immune effector and barrier-associated programs that may be intermittently sampled depending on tissue composition, the microenvironmental state, and cohort selection. In cancer

https://doi.org/10.1021/acs.jproteome.5c01259
J. Proteome Res. 2026, 25, 2901−2914

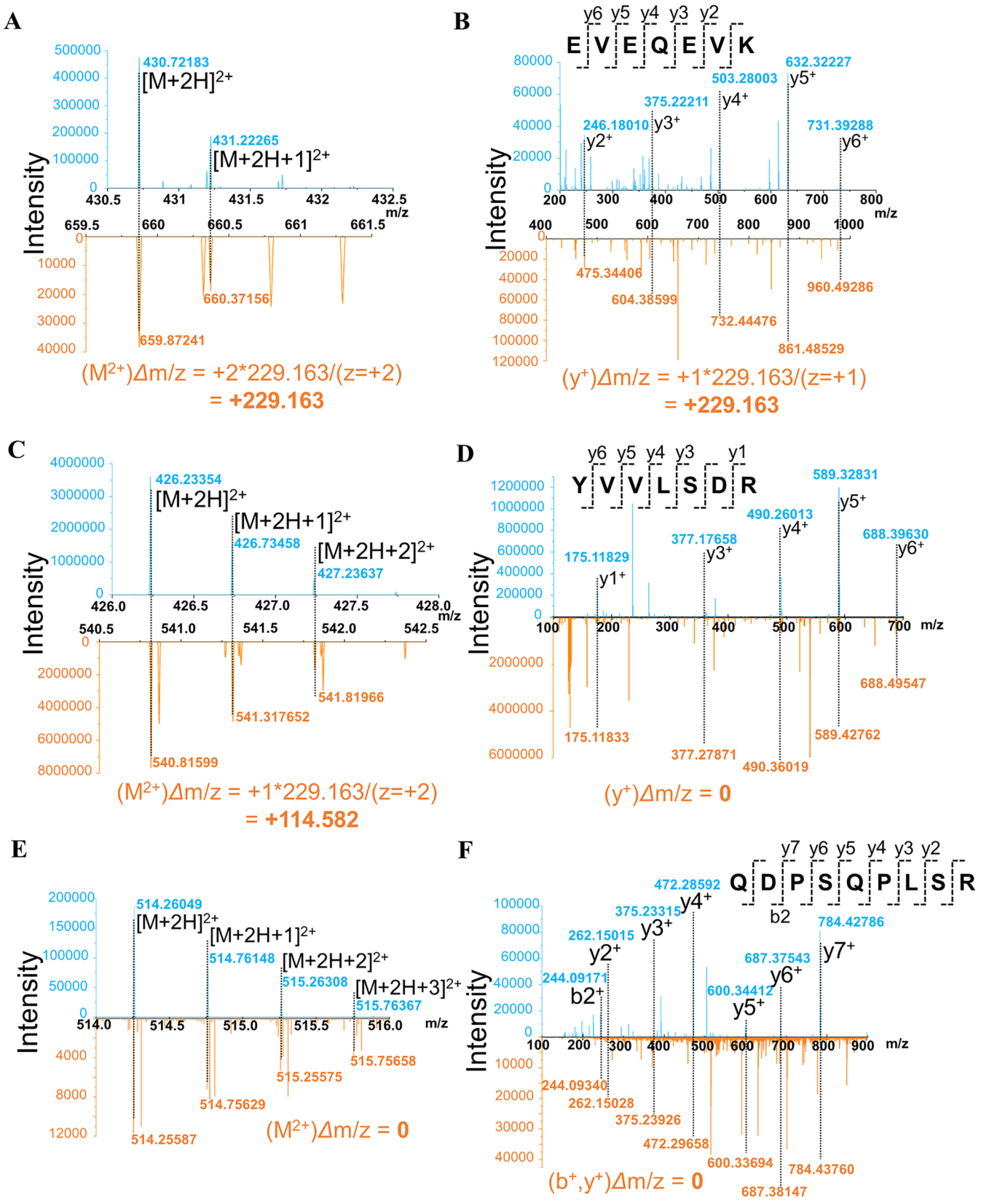


Figure 6. Newly identified proteins verified by synthetic peptides (A) MS1 and (B) MS2 spectra from the unique sequence EVEQEVK related to ABCAA_HUMAN. (C) MS1 and (D) MS2 spectra from the unique sequence YVVLSDR related to CL056_HUMAN. (E) MS1 and (F) MS2 spectra from the unique sequence QDPSQPLSR related to TBA4B_HUMAN. Spectra in blue lines and labels were generated by the synthetic peptides' DDA identification results. Spectra in orange lines and labels were generated by MSCP identification results.

https://doi.org/10.1021/acs.jproteome.5c01259
J. Proteome Res. 2026, 25, 2901−2914

contexts, such proteins can be particularly relevant because tumor−immune interactions and mucosal barrier biology vary markedly by anatomical site, tumor subtype, and stromal/immune infiltration, which reinforces the value of integrating diverse cohorts and model systems when seeking MS evidence for low-frequency or context-dependent proteins.

A second prominent theme is chromatin and nucleosome biology. GO molecular function enrichment strongly prioritizes structural constituents of chromatin, and GO cellular component terms include nucleosome and protein−DNA complex. Concordantly, Reactome analysis shows enrichment for chromatin- and genome-regulatory pathways, including DNA methylation, HDACs deacetylate histones, nucleosome assembly, and additional DNA maintenance programs such as the packaging of telomere ends and DNA damage/telomere stress−induced senescence. These results suggest that MSCP integration captures additional evidence for histone/chromatin-associated proteins and related regulatory factors that are often influenced by extraction chemistry, fractionation depth, and instrument acquisition strategy. Importantly, these pathways are central to cancer biology through epigenetic remodeling, genome stability, and stress-response programs, providing a mechanistic context for why these proteins represent meaningful additions to MS-supported entries in cancer-derived samples. Subcellular localization enrichment is also consistent with this functional profile. In addition to nuclear-associated categories (nuclear lumen and nucleoplasm), we observe enrichment of Golgi-related components, implying that a subset of newly identified proteins may represent compartmentalized trafficking/processing machinery that becomes detectable with improved depth and diversity of sample types. Together with our localization analysis of the newly identified set (Figure 3D) and pathway-level enrichment, these findings support the view that many newly identified proteins belong to biologically specialized or technically under-sampled compartments (chromatin-associated nuclear complexes, secretory/trafficking compartments, and immune/barrier programs). This provides a functional explanation for why newly identified proteins continue to accrue when heterogeneous cancer cohorts are integrated with complementary acquisition strategies (DDA/DIA).

Collectively, the enrichment results extend MSCP beyond a coverage-oriented catalog by providing a functional and mechanistic context for the newly MS-supported proteins. This layered interpretation—combining protein existence benchmarking, localization patterns, and pathway enrichment—supports MSCP as a resource for prioritizing proteins for downstream follow-up while retaining data set provenance to enable system-focused interpretation.

Current Limitations

Challenges persist in protein detection and classification, particularly with regard to missing proteins. Technological limitations in mass spectrometry for resolving the low-abundance protein identification remain a significant barrier. Additionally, data integration across diverse platforms and data sets poses computational challenges, including harmonization of data formats, scalability, and processing speed. Ambiguities in protein identification caused by post-translational modifications and sequence variations further complicate data interpretation and validation. Achieving comprehensive protein pathway mapping and functional annotation also requires more robust tools and algorithms to interpret large-scale data sets effectively. Because proteome detection depends on cohort composition, tissue context, and instrument and acquisition strategy, interpretation is often strongest when the analysis is aligned to a defined biological system. MSCP addresses this by providing source-resolved evidence, allowing users to focus on one cohort/model when system specificity is required, while also enabling an integrated view when broad protein coverage is advantageous for discovery and method development. Consequently, the database reflects heterogeneity in sample composition and study design, which can influence detectability and apparent protein presence. Differences in tumor purity, stromal/immune infiltration, anatomical site, clinical covariates, and model-specific biology may introduce cohort- or model-dependent biases. Accordingly, detection in MSCP should be interpreted as evidence of MS observation in at least one cancer-related context, and system-specific conclusions are best supported by restricting analyses to the relevant cohort and model subsets using the source-resolved evidence. MSCP also aggregates results generated by using heterogeneous technical platforms and analytical workflows, including variations in instrument type, acquisition mode (DDA vs DIA), fractionation depth, labeling strategy, sample preparation, and protein inference procedures. These differences can yield systematic shifts in sensitivity and reporting, leading to potential false negatives, in which proteins are not observed in each data set due to limited depth, ion suppression, or study-specific reporting thresholds, and occasional false positives such as ambiguity from shared peptides, homologous sequences, PTM-dependent assignments, or database-version effects, even under FDR-controlled pipelines. While MSCP relies on high-confidence, FDR-controlled identifications from the originating sources and includes orthogonal synthetic-peptide verification for representative newly identified proteins, users should account for platform variability when comparing sources and consider recurrence across independent data sets/acquisition modes as an additional confidence criterion for prioritization.

Outlook

Planned expansions include the integration of advanced technologies and the addressing of research gaps in protein functions and pathways, especially in cancer-related studies. Emerging advancements in AI-driven data analysis and machine learning models offer promising tools for processing large data sets, improving protein classification accuracy, and identifying novel patterns. Single-cell proteomics technologies are also being integrated to capture protein expression at an unprecedented resolution, enabling deeper insights into cellular heterogeneity and microenvironments. While increased data set aggregation is expected to yield more protein identifications and improved overlap with multisource collection databases, MSCP focuses on making these identifications interpretable and traceable. Accordingly, we quantify nonredundant incremental contributions (total and newly identified proteins relative to neXtProt PE1) and preserve source-level provenance, enabling users to distinguish broad coverage effects from cohort-specific evidence. In addition, MSCP is intended as an evolving resource that incorporates newly published cancer proteomics data sets in updates, which is expected to further increase coverage and add additional MS-supported proteins while maintaining transparent provenance and comparable benchmarking across updates. Future releases will add more derived fields, such as

https://doi.org/10.1021/acs.jproteome.5c01259
J. Proteome Res. 2026, 25, 2901−2914

model type, new cancer cohorts, acquisition mode, and pregenerated cohort/model-specific exports to streamline stratified querying. We also plan to implement confidence tiers based on recurrence across independent sources and concordance across acquisition modes to further support the context-appropriate prioritization of MS evidence. Expansions include integrating advanced technologies and addressing research gaps in protein functions and pathways, especially in cancer-related studies.

CONCLUSIONS

In this study, we developed MSCP, a cancer-context MS-evidence resource that integrates proteomics identifications from tumors, cancer cell lines, and PDX models into a unified UniProtKB-mapped catalog. The current release contains 15,964 MS-supported UniProtKB-Swiss-Prot entries and adds 525 proteins newly supported by MS evidence relative to the neXtProt PE1 benchmark, demonstrating the value of integrating heterogeneous cohorts and complementary acquisition strategies. Functional enrichment analyses of the newly identified proteins extend MSCP beyond coverage metrics by highlighting coherent biological themes, including immune-associated processes and chromatin- and genome-regulatory pathways, while synthetic-peptide verification provides orthogonal evidence for representative entries. By retaining data set-level provenance, MSCP enables both system-focused queries and integrated evidence aggregation, supporting cancer proteomics study design, evidence triage, and downstream targeted or DIA-enabled assay development. Future updates incorporating newly published data sets and additional annotation layers will further expand MSCP coverage while maintaining transparent benchmarking and reproducible evidence tracking.

ASSOCIATED CONTENT

Data Availability Statement

The database and related tools are openly accessible by the National Cancer Institute Proteomic Data Commons and ProteomeXChange Consortium via the PRIDE repository (PDC000219, PDC000154, PDC000204, PDC000203, PDC000110, PDC000270, PDC000221, PDC000200, PDC000129, PDC000116, PDC000234, PDC000125, PDC000124, PDC000120, PDC000242, PDC000613, PDC000630, PXD060378, and PXD071155).

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jproteome.5c01259.

15,964 UniProtKB-Swiss-Prot entries identified by MSCP from 27 data sources (XLSX)

Cumulative contribution of integrated data sources to MSCP coverage (XLSX)

Comparison of MSCP with representative proteomics resources (XLSX)

UniProtKB-Swiss-Prot entries newly identified by MSCP and verified by synthetic unique peptides (XLSX)

GO and Reactome enrichment results for the 525 newly identified proteins in MSCP (XLSX)

AUTHOR INFORMATION

Corresponding Author

Yuanyu Huang − *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, United States;* orcid.org/0009-0002-5930-8424; Email: yhuan295@jh.edu

Authors

Lijun Chen − *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, United States*

Peter I-Fan Wu − *Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, Maryland 21702, United States*

Poorva Juneja − *Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, Maryland 21702, United States*

Li Chen − *Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, Maryland 21702, United States*

Yvonne A. Evrard − *Frederick National Laboratory for Cancer Research, Leidos Biomedical Research, Inc., Frederick, Maryland 21702, United States;* orcid.org/0000-0002-3475-4850

Liyuan Jiao − *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, United States*

Yingwei Hu − *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, United States;* orcid.org/0000-0002-4629-0985

Xu Zhang − *National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States*

James H. Doroshow − *National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, United States*

Ratna R. Thangudu − *Inner City Fund (ICF) International, Inc., Rockville, Maryland 20850, United States*

Alexander Pilozzi − *Inner City Fund (ICF) International, Inc., Rockville, Maryland 20850, United States*

Hui Zhang − *Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, United States;* orcid.org/0000-0001-8726-7098

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jproteome.5c01259

Author Contributions

The manuscript was written through the contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by National Institutes of Health, National Cancer Institute, the Clinical Proteomic Tumor Analysis Consortium (CPTAC, U24CA271079), and Pancreatic Cancer Detection Consortium (PCDC, U01CA274514).

ABBREVIATIONS

MSCP, mass spectrometric detected cancer proteins; UniProtKB, The Universal Protein Resource Knowledgebase;

https://doi.org/10.1021/acs.jproteome.5c01259
J. Proteome Res. 2026, 25, 2901−2914

PRIDE, Proteomics Identifications archive database; MassIVE-KB, Mass Spectrometry Interactive Virtual Environment Knowledge Base; PE, protein existence; CPTAC, Clinical Proteomic Tumor Analysis Consortium; CompRef, Comprehensive Reference; PDXs, Patient-Derived Xenografts; DDA, data-dependent acquisition; DIA, data-independent acquisition; LSCC, Lung Squamous Cell Carcinoma; GBM, glioblastoma; OV, ovarian cancer; PDAC, Pancreatic Ductal Adenocarcinoma; HNSCC, head and neck squamous cell carcinoma; CCRCC, clear cell renal cell carcinomas; COAD, colon adenocarcinoma; LUAD, lung adenocarcinoma; UCEC, uterine corpus endometrial carcinoma; BRCA, breast invasive carcinoma; RCC, renal cell carcinomas; IPMNs, intraductal papillary mucinous neoplasms; KEGG, Kyoto Encyclopedia of Genes and Genomes

REFERENCES

(1) Pandey, A.; Mann, M. Proteomics to study genes and genomes. *Nature* **2000**, *405*, 837−846.

(2) Cravatt, B. F.; Simon, G. M.; Yates III, J. R. The biological impact of mass-spectrometry-based proteomics. *Nature* **2007**, *450*, 991.

(3) Aebersold, R.; Mann, M. Mass-spectrometric exploration of proteome structure and function. *Nature* **2016**, *537*, 347.

(4) Guo, T.; Steen, J. A.; Mann, M. Mass-spectrometry-based proteomics: from single cells to clinical applications. *Nature* **2025**, *638*, 901.

(5) Carter, P. J.; Rajpal, A. Designing antibodies as therapeutics. *Cell* **2022**, *185*, 2789.

(6) Phizicky, E.; Bastiaens, P. I. H.; Zhu, H.; Snyder, M.; Fields, S. Protein analysis on a proteomic scale. *Nature* **2003**, *422*, 208−215.

(7) Phlairaharn, T.; Grégoire, S.; Woltereck, L. R.; Petrosius, V.; Furtwängler, B.; Searle, B. C.; Schoof, E. M. High Sensitivity Limited Material Proteomics Empowered by Data-Independent Acquisition on Linear Ion Traps. *J. Proteome Res.* **2022**, *21*, 2815.

(8) Michalski, A.; Cox, J.; Mann, M. More than 100,000 Detectable Peptide Species Elute in Single Shotgun Proteomics Runs but the Majority is Inaccessible to Data-Dependent LC−MS/MS. *J. Proteome Res.* **2011**, *10*, 1785.

(9) Müller, F.; Birklbauer, M. J.; Bubis, J.; Stejskal, K.; Dorfer, V.; Mechtler, K. Breaking barriers in crosslinking mass spectrometry with enhanced throughput and sensitivity using Orbitrap Astral. *Nat. Commun.* **2025**, *16*, No. 9877, DOI: 10.1038/s41467-025-64844-7.

(10) Deutsch, E. W.; Bandeira, N.; Sharma, V.; Perez-Riverol, Y.; Carver, J. J.; Kundu, D. J.; García-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.; Vizcaíno, J. A. The ProteomeXchange consortium in 2020: enabling 'big data' approaches in proteomics. *Nucleic Acids Res.* **2020**, *48*, D1145 DOI: 10.1093/nar/gkz984.

(11) Deutsch, E. W.; Bandeira, N.; Perez-Riverol, Y.; Sharma, V.; Carver, J. J.; Mendoza, L.; Kundu, D. J.; Wang, S.; Bandla, C.; Kamatchinathan, S.; Hewapathirana, S.; Pullman, B. S.; Wertz, J.; Sun, Z.; Kawano, S.; Okuda, S.; Watanabe, Y.; Maclean, B.; Maccoss, M. J.; Zhu, Y.; Ishihama, Y.; Vizcaíno, J. A. The ProteomeXchange consortium at 10 years: 2023 update. *Nucleic Acids Res.* **2023**, *51*, D1539.

(12) Bateman, A.; Martin, M.; Orchard, S.; Magrane, M.; Adesina, A.; Ahmad, S.; Bowler-Barnett, E. H.; Bye-A-Jee, H.; Carpentier, D.; Denny, P.; Fan, J.; Garmiri, P.; Gonzales, L. J. D. C.; Hussein, A.; Ignatchenko, A.; Insana, G.; Ishtiaq, R.; Joshi, V.; Jyothi, D.; Kandasaamy, S.; Lock, A.; Luciani, A.; Luo, J.; Lussi, Y.; Marin, J. S. M.; Raposo, P.; Rice, D. L.; Santos, R.; Speretta, E.; Stephenson, J.; Totoo, P.; Tyagi, N.; Urakova, N.; Vasudev, P.; Warner, K.; Wijerathne, S.; Yu, C. W.; Zaru, R.; Bridge, A. J.; Aimo, L.; Argoud-Puy, G.; Auchincloss, A. H.; Axelsen, K. B.; Bansal, P.; Baratin, D.; Batista Neto, T. M.; Blatter, M.; Bolleman, J. T.; Boutet, E.; Breuza, L.; Gil, B. C.; Casals-Casas, C.; Echioukh, K. C.; Coudert, E.; Cuche, B.; De Castro, E.; Estreicher, A.; Famiglietti, M. L.; Feuermann, M.; Gasteiger, E.; Gaudet, P.; Gehant, S.; Gerritsen, V.; Gos, A.; Gruaz, N.; Hulo, C.; Hyka-Nouspikel, N.; Jungo, F.; Kerhornou, A.; Mercier, P. L.; Lieberherr, D.; Masson, P.; Morgat, A.; Paesano, S.; Pedruzzi, I.; Pilbout, S.; Pourcel, L.; Poux, S.; Pozzato, M.; Pruess, M.; Redaschi, N.; Rivoire, C.; Sigrist, C. J. A.; Sonesson, K.; Sundaram, S.; Sveshnikova, A.; Wu, C. H.; Arighi, C. N.; Chen, C.; Chen, Y.; Huang, H.; Laiho, K.; Lehvaslaiho, M.; Mcgarvey, P.; Natale, D. A.; Ross, K.; Vinayaka, C. R.; Wang, Y.; Zhang, J. UniProt: the Universal Protein Knowledgebase in 2025. *Nucleic Acids Res.* **2025**, *53*, D609.

(13) Deutsch, E. The ProteomeXchange consortium in 2026: making proteomics data FAIR. *Nucleic Acids Res.* **2026**, *54*, 549.

(14) Desiere, F.; Deutsch, E. W.; King, N. L.; Nesvizhskii, A. I.; Mallick, P.; Eng, J.; Chen, S.; Eddes, J.; Loevenich, S. N.; Aebersold, R. The PeptideAtlas project. *Nucleic Acids Res.* **2006**, *34*, D655.

(15) Clark, D. J.; Hu, Y.; Bocik, W.; Chen, L.; Schnaubelt, M.; Roberts, R.; Shah, P.; Whiteley, G.; Zhang, H. Evaluation of NCI-7 Cell Line Panel as a Reference Material for Clinical Proteomics. *J. Proteome Res.* **2018**, *17*, 2205.

(16) Cao, L.; Huang, C.; Cui Zhou, D.; Hu, Y.; Lih, T. M.; Savage, S. R.; Krug, K.; Clark, D. J.; Schnaubelt, M.; Chen, L.; Da Veiga Leprevost, F.; Eguez, R. V.; Yang, W.; Pan, J.; Wen, B.; Dou, Y.; Jiang, W.; Liao, Y.; Shi, Z.; Terekhanova, N. V.; Cao, S.; Lu, R. J.; Li, Y.; Liu, R.; Zhu, H.; Ronning, P.; Wu, Y.; Wyczalkowski, M. A.; Easwaran, H.; Danilova, L.; Mer, A. S.; Yoo, S.; Wang, J. M.; Liu, W.; Haibe-Kains, B.; Thiagarajan, M.; Jewell, S. D.; Hostetter, G.; Newton, C. J.; Li, Q. K.; Roehrl, M. H.; Fenyö, D.; Wang, P.; Nesvizhskii, A. I.; Mani, D. R.; Omenn, G. S.; Boja, E. S.; Mesri, M.; Robles, A. I.; Rodriguez, H.; Bathe, O. F.; Chan, D. W.; Hruban, R. H.; Ding, L.; Zhang, B.; Zhang, H.; Amin, M.; An, E.; Ayad, C.; Bauer, T.; Birger, C.; Birrer, M. J.; Boca, S. M.; Bocik, W.; Borucki, M.; Cai, S.; Carr, S. A.; Cerda, S.; Chen, H.; Chen, S.; Chesla, D.; Chinnaiyan, A. M.; Colaprico, A.; Cottingham, S.; Derejska, M.; Dhanasekaran, S. M.; Domagalski, M. J.; Druker, B. J.; Duffy, E.; Dyer, M. A.; Edwards, N. J.; Ellis, M. J.; Eschbacher, J.; Francis, A.; Francis, J.; Gabriel, S.; Gabrovski, N.; Gardner, J.; Getz, G.; Gillette, M. A.; Goldthwaite, C. A.; Grady, P.; Guo, S.; Hariharan, P.; Hiltke, T.; Hindenach, B.; Hoadley, K. A.; Huang, J.; Jones, C. D.; Ketchum, K. A.; Kinsinger, C. R.; Koziak, J. M.; Kusnierz, K.; Liu, T.; Long, J.; Mallery, D.; Mareedu, S.; Matteotti, R.; Maunganidze, N.; Mcgarvey, P. B.; Minoo, P.; Paklina, O. V.; Paulovich, A. G.; Payne, S. H.; Potapova, O.; Pruetz, B.; Qi, L.; Roche, N.; Rodland, K. D.; Rohrer, D. C.; Schadt, E. E.; Shabunin, A. V.; Shelton, T.; Shutack, Y.; Singh, S.; Smith, M.; Smith, R. D.; Sokoll, L. J.; Suh, J.; Thangudu, R. R.; Tsang, S. X.; Um, K. S.; Valley, D. R.; Vatanian, N.; Wang, W.; Wilson, G. D.; Wiznerowicz, M.; Zhang, Z.; Zhao, G. Proteogenomic characterization of pancreatic ductal adenocarcinoma. *Cell* **2021**, *184*, 5031.

(17) Wang, L.-B.; Karpova, A.; Gritsenko, M. A.; Kyle, J. E.; Cao, S.; Li, Y.; Rykunov, D.; Colaprico, A.; Rothstein, J. H.; Hong, R.; Stathias, V.; Cornwell, M.; Petralia, F.; Wu, Y.; Reva, B.; Krug, K.; Pugliese, P.; Kawaler, E.; Olsen, L. K.; Liang, W.; Song, X.; Dou, Y.; Wendl, M. C.; Caravan, W.; Liu, W.; Cui Zhou, D.; Ji, J.; Tsai, C.; Petyuk, V. A.; Moon, J.; Ma, W.; Chu, R. K.; Weitz, K. K.; Moore, R. J.; Monroe, M. E.; Zhao, R.; Yang, X.; Yoo, S.; Krek, A.; Demopoulos, A.; Zhu, H.; Wyczalkowski, M. A.; Mcmichael, J. F.; Henderson, B. L.; Lindgren, C. M.; Boekweg, H.; Lu, S.; Baral, J.; Yao, L.; Stratton, K. G.; Bramer, L. M.; Zink, E.; Couvillion, S. P.; Bloodsworth, K. J.; Satpathy, S.; Sieh, W.; Boca, S. M.; Schürer, S.; Chen, F.; Wiznerowicz, M.; Ketchum, K. A.; Boja, E. S.; Kinsinger, C. R.; Robles, A. I.; Hiltke, T.; Thiagarajan, M.; Nesvizhskii, A. I.; Zhang, B.; Mani, D. R.; Ceccarelli, M.; Chen, X. S.; Cottingham, S. L.; Li, Q. K.; Kim, A. H.; Fenyö, D.; Ruggles, K. V.; Rodriguez, H.; Mesri, M.; Payne, S. H.; Resnick, A. C.; Wang, P.; Smith, R. D.; Iavarone, A.; Chheda, M. G.; Barnholtz-Sloan, J. S.; Rodland, K. D.; Liu, T.; Ding, L.; Agarwal, A.; Amin, M.; An, E.; Anderson, M. L.; Andrews, D. W.; Bauer, T.; Birger, C.; Birrer, M. J.; Blumenberg, L.; Bocik, W. E.; Borate, U.; Borucki, M.; Burke, M. C.; Cai, S.; Calinawan, A. P.; Carr,

https://doi.org/10.1021/acs.jproteome.5c01259
J. Proteome Res. 2026, 25, 2901−2914

S. A.; Cerda, S.; Chan, D. W.; Charamut, A.; Chen, L. S.; Chesla, D.; Chinnaiyan, A. M.; Chowdhury, S.; Cieślik, M. P.; Clark, D. J.; Culpepper, H.; Czernicki, T.; D'angelo, F.; Day, J.; De Young, S.; Demir, E.; Dhanasekaran, S. M.; Dhir, R.; Domagalski, M. J.; Druker, B.; Duffy, E.; Dyer, M.; Edwards, N. J.; Edwards, R.; Elburn, K.; Ellis, M. J.; Eschbacher, J.; Francis, A.; Gabriel, S.; Gabrovski, N.; Garofano, L.; Getz, G.; Gillette, M. A.; Godwin, A. K.; Golbin, D.; Hanhan, Z.; Hannick, L. I.; Hariharan, P.; Hindenach, B.; Hoadley, K. A.; Hostetter, G.; Huang, C.; Jaehnig, E.; Jewell, S. D.; Ji, N.; Jones, C. D.; Karz, A.; Kaspera, W.; Kim, L.; Kothadia, R. B.; Kumar-Sinha, C.; Lei, J.; Leprevost, F. D.; Li, K.; Liao, Y.; Lilly, J.; Liu, H.; Lubínski, J.; Madan, R.; Maggio, W.; Malc, E.; Malovannaya, A.; Mareedu, S.; Markey, S. P.; Marrero-Oliveras, A.; Martinez, N.; Maunganidze, N.; Mcdermott, J. E.; Mcgarvey, P. B.; Mcgee, J.; Mieczkowski, P.; Migliozzi, S.; Modugno, F.; Montgomery, R.; Newton, C. J.; Omenn, G. S.; Ozbek, U.; Paklina, O. V.; Paulovich, A. G.; Perou, A. M.; Pico, A. R.; Piehowski, P. D.; Placantonakis, D. G.; Polonskaya, L.; Potapova, O.; Pruetz, B.; Qi, L.; Ramkissoon, S.; Resnick, A.; Richey, S.; Riggins, G.; Robinson, K.; Roche, N.; Rohrer, D. C.; Rood, B. R.; Rossell, L.; Savage, S. R.; Schadt, E. E.; Shi, Y.; Shi, Z.; Shutack, Y.; Singh, S.; Skelly, T.; Sokoll, L. J.; Stawicki, J.; Stein, S. E.; Suh, J.; Szopa, W.; Tabor, D.; Tan, D.; Tansil, D.; Thangudu, R. R.; Tognon, C.; Traer, E.; Tsang, S.; Tyner, J.; Um, K. S.; Valley, D. R.; Vasaikar, S.; Vatanian, N.; Velvulou, U.; Vernon, M.; Wan, W.; Wang, J.; Webster, A.; Wen, B.; Whiteaker, J. R.; Wilson, G. D.; Zakhartsev, Y.; Zelt, R.; Zhang, H.; Zhang, L.; Zhang, Z.; Zhao, G.; Zhu, J. Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell* **2021**, *39*, 509.

(18) Satpathy, S.; Krug, K.; Jean Beltran, P. M.; Savage, S. R.; Petralia, F.; Kumar-Sinha, C.; Dou, Y.; Reva, B.; Kane, M. H.; Avanessian, S. C.; Vasaikar, S. V.; Krek, A.; Lei, J. T.; Jaehnig, E. J.; Omelchenko, T.; Geffen, Y.; Bergstrom, E. J.; Stathias, V.; Christianson, K. E.; Heiman, D. I.; Cieslik, M. P.; Cao, S.; Song, X.; Ji, J.; Liu, W.; Li, K.; Wen, B.; Li, Y.; Gümüş, Z. H.; Selvan, M. E.; Soundararajan, R.; Visal, T. H.; Raso, M. G.; Parra, E. R.; Babur, Ö.; Vats, P.; Anand, S.; Schraink, T.; Cornwell, M.; Rodrigues, F. M.; Zhu, H.; Mo, C.; Zhang, Y.; Da Veiga Leprevost, F.; Huang, C.; Chinnaiyan, A. M.; Wyczalkowski, M. A.; Omenn, G. S.; Newton, C. J.; Schurer, S.; Ruggles, K. V.; Fenyö, D.; Jewell, S. D.; Thiagarajan, M.; Mesri, M.; Rodriguez, H.; Mani, S. A.; Udeshi, N. D.; Getz, G.; Suh, J.; Li, Q. K.; Hostetter, G.; Paik, P. K.; Dhanasekaran, S. M.; Govindan, R.; Ding, L.; Robles, A. I.; Clauser, K. R.; Nesvizhskii, A. I.; Wang, P.; Carr, S. A.; Zhang, B.; Mani, D. R.; Gillette, M. A.; Green, A.; Molinolo, A.; Francis, A.; Paulovich, A. G.; Karnuta, A.; Colaprico, A.; Hindenach, B.; Pruetz, B. L.; Kubisa, B.; Druker, B. J.; Huynh, C.; Goldthwaite, C. A.; Birger, C.; Kinsinger, C. R.; Jones, C. D.; Rohrer, D.; Valley, D. R.; Chan, D. W.; Chesla, D.; Hansel, D.; Ponomareva, E. V.; Duffy, E.; Burks, E.; Schadt, E. E.; Fedorov, E. S.; An, E.; Ding, F.; Wilson, G. D.; Batra, H.; Zhang, H.; Maas, J. E.; Eschbacher, J.; Ketchum, K. A.; Rodland, K. D.; Hoadley, K. A.; Suzuki, K.; Um, K. S.; Qi, L.; Bernard, L.; Wiznerowicz, M.; Wojtyś, M.; Domagalski, M. J.; Ellis, M. J.; Dyer, M. A.; Borucki, M.; Anurag, M.; Birrer, M. J.; Xu, M.; Krotevich, M.; Roche, N.; Edwards, N. J.; Vatanian, N.; Mucci, N. R.; Maunganidze, N.; Gabrovski, N.; Potapova, O.; Fadare, O.; Grady, P.; Mcgarvey, P. B.; Hariharan, P.; Thangudu, R. R.; Montgomery, R.; Pandurengan, R.; Smith, R. D.; Welsh, R. J.; Mareedu, S.; Payne, S. H.; Cottingham, S.; Singh, S.; Tsang, S. X.; Cai, S.; Gabriel, S.; Liu, T.; Hiltke, T.; Vashist, T.; Bauer, T.; Sovenko, V.; Tourtellotte, W. G.; Ma, W.; Bocik, W.; Hasan, W.; Jing, X.; Tang, X.; Liao, Y.; Yvonne; Shutack; Zhang, Z.; Hanhan, Z. A proteogenomic portrait of lung squamous cell carcinoma. *Cell* **2021**, *184*, 4348.

(19) Mcdermott, J. E.; Arshad, O. A.; Petyuk, V. A.; Fu, Y.; Gritsenko, M. A.; Clauss, T. R.; Moore, R. J.; Schepmoes, A. A.; Zhao, R.; Monroe, M. E.; Schnaubelt, M.; Tsai, C.; Payne, S. H.; Huang, C.; Wang, L.; Foltz, S.; Wyczalkowski, M.; Wu, Y.; Song, E.; Brewer, M. A.; Thiagarajan, M.; Kinsinger, C. R.; Robles, A. I.; Boja, E. S.; Rodriguez, H.; Chan, D. W.; Zhang, B.; Zhang, Z.; Ding, L.; Smith, R. D.; Liu, T.; Rodland, K. D. Proteogenomic Characterization of Ovarian HGSC Implicates Mitotic Kinases, Replication Stress in Observed Chromosomal Instability. *Cell Rep. Med.* **2020**, *1*, No. 100075, DOI: 10.1016/j.xcrm.2020.100075.

(20) Huang, C.; Chen, L.; Savage, S. R.; Eguez, R. V.; Dou, Y.; Li, Y.; Da Veiga Leprevost, F.; Jaehnig, E. J.; Lei, J. T.; Wen, B.; Schnaubelt, M.; Krug, K.; Song, X.; Cieślik, M.; Chang, H.; Wyczalkowski, M. A.; Li, K.; Colaprico, A.; Li, Q. K.; Clark, D. J.; Hu, Y.; Cao, L.; Pan, J.; Wang, Y.; Cho, K.; Shi, Z.; Liao, Y.; Jiang, W.; Anurag, M.; Ji, J.; Yoo, S.; Zhou, D. C.; Liang, W.; Wendl, M.; Vats, P.; Carr, S. A.; Mani, D. R.; Zhang, Z.; Qian, J.; Chen, X. S.; Pico, A. R.; Wang, P.; Chinnaiyan, A. M.; Ketchum, K. A.; Kinsinger, C. R.; Robles, A. I.; An, E.; Hiltke, T.; Mesri, M.; Thiagarajan, M.; Weaver, A. M.; Sikora, A. G.; Lubiński, J.; Wierzbicka, M.; Wiznerowicz, M.; Satpathy, S.; Gillette, M. A.; Miles, G.; Ellis, M. J.; Omenn, G. S.; Rodriguez, H.; Boja, E. S.; Dhanasekaran, S. M.; Ding, L.; Nesvizhskii, A. I.; El-Naggar, A. K.; Chan, D. W.; Zhang, H.; Zhang, B.; Agarwal, A.; Anderson, M. L.; Avanessian, S. C.; Avtonomov, D.; Bathe, O. F.; Birger, C.; Birrer, M. J.; Blumenberg, L.; Bocik, W. E.; Borate, U.; Borucki, M.; Burke, M. C.; Cai, S.; Calinawan, A. P.; Cerda, S.; Charamut, A.; Chen, L. S.; Chowdhury, S.; Clauser, K. R.; Culpepper, H.; Czernicki, T.; D'angelo, F.; Day, J.; De Young, S.; Demir, E.; Ding, F.; Domagalski, M. J.; Dort, J. C.; Druker, B.; Duffy, E.; Dyer, M.; Edwards, N. J.; Elburn, K.; Ermakova, T. S.; Fenyo, D.; Ferrarotto, R.; Francis, A.; Gabriel, S.; Garofano, L.; Geffen, Y.; Getz, G.; Goldthwaite, C. A.; Hannick, L. I.; Hariharan, P.; Hayes, D. N.; Heiman, D.; Hindenach, B.; Hoadley, K. A.; Hostetter, G.; Hyrcza, M.; Jewell, S. D.; Jones, C. D.; Kane, M. H.; Karz, A.; Kothadia, R. B.; Krek, A.; Kumar-Sinha, C.; Liu, T.; Liu, H.; Ma, W.; Malc, E.; Malovannaya, A.; Mareedu, S.; Markey, S. P.; Marrero-Oliveras, A.; Maunganidze, N.; Mcdermott, J. E.; Mcgarvey, P. B.; Mcgee, J.; Mieczkowski, P.; Migliozzi, S.; Montgomery, R.; Newton, C. J.; Ozbek, U.; Paulovich, A. G.; Payne, S. H.; Pazardzhikliev, D. D.; Perou, A. M.; Petralia, F.; Petrenko, L.; Piehowski, P. D.; Placantonakis, D.; Polonskaya, L.; Ponomareva, E. V.; Potapova, O.; Qi, L.; Qu, N.; Ramkissoon, S.; Reva, B.; Richey, S.; Robinson, K.; Roche, N.; Rodland, K.; Rohrer, D. C.; Rykunov, D.; Schadt, E. E.; Shi, Y.; Shutack, Y.; Singh, S.; Skelly, T.; Smith, R.; Sokoll, L. J.; Stawicki, J.; Stein, S. E.; Suh, J.; Szopa, W.; Tabor, D.; Tan, D.; Tansil, D.; Teo, G. C.; Thangudu, R. R.; Tognon, C.; Traer, E.; Tsang, S.; Tyner, J.; Um, K. S.; Valley, D. R.; Vasilev, L. V.; Vatanian, N.; Velvulou, U.; Vernon, M.; Westbrook, T. F.; Whiteaker, J. R.; Wu, Y.; Xu, M.; Yao, L.; Yi, X.; Yu, F.; Zaalishvili, K.; Zakhartsev, Y.; Zelt, R.; Zhao, G.; Zhu, J. Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. *Cancer Cell* **2021**, *39*, 361.

(21) Clark, D. J.; Dhanasekaran, S. M.; Petralia, F.; Pan, J.; Song, X.; Hu, Y.; Da Veiga Leprevost, F.; Reva, B.; Lih, T. M.; Chang, H.; Ma, W.; Huang, C.; Ricketts, C. J.; Chen, L.; Krek, A.; Li, Y.; Rykunov, D.; Li, Q. K.; Chen, L. S.; Ozbek, U.; Vasaikar, S.; Wu, Y.; Yoo, S.; Chowdhury, S.; Wyczalkowski, M. A.; Ji, J.; Schnaubelt, M.; Kong, A.; Sethuraman, S.; Avtonomov, D. M.; Ao, M.; Colaprico, A.; Cao, S.; Cho, K.; Kalayci, S.; Ma, S.; Liu, W.; Ruggles, K.; Calinawan, A.; Gümüş, Z. H.; Geiszler, D.; Kawaler, E.; Teo, G. C.; Wen, B.; Zhang, Y.; Keegan, S.; Li, K.; Chen, F.; Edwards, N.; Pierorazio, P. M.; Chen, X. S.; Pavlovich, C. P.; Hakimi, A. A.; Brominski, G.; Hsieh, J. J.; Antczak, A.; Omelchenko, T.; Lubinski, J.; Wiznerowicz, M.; Linehan, W. M.; Kinsinger, C. R.; Thiagarajan, M.; Boja, E. S.; Mesri, M.; Hiltke, T.; Robles, A. I.; Rodriguez, H.; Qian, J.; Fenyö, D.; Zhang, B.; Ding, L.; Schadt, E.; Chinnaiyan, A. M.; Zhang, Z.; Omenn, G. S.; Cieslik, M.; Chan, D. W.; Nesvizhskii, A. I.; Wang, P.; Zhang, H.; Hashimi, A. S.; Pico, A. R.; Karpova, A.; Charamut, A.; Paulovich, A. G.; Perou, A. M.; Malovannaya, A.; Marrero-Oliveras, A.; Agarwal, A.; Hindenach, B.; Pruetz, B.; Kim, B.; Druker, B. J.; Newton, C. J.; Birger, C.; Jones, C. D.; Tognon, C.; Mani, D. R.; Valley, D. R.; Rohrer, D. C.; Zhou, D. C.; Tansil, D.; Chesla, D.; Heiman, D.; Wheeler, D.; Tan, D.; Chan, D.; Demir, E.; Malc, E.; Modugno, F.; Getz, G.; Hostetter, G.; Wilson, G. D.; Hart, G. W.; Zhu, H.; Liu, H.; Culpepper, H.; Sun, H.; Zhou, H.; Day, J.; Suh, J.; Huang, J.; Mcdermott, J.; Whiteaker, J. R.; Tyner, J. W.; Eschbacher, J.; Chen, J.; Mcgee, J.; Zhu, J.; Ketchum, K. A.; Rodland, K. D.; Clauser, K.;

https://doi.org/10.1021/acs.jproteome.5c01259
J. Proteome Res. 2026, 25, 2901−2914

Robinson, K.; Krug, K.; Hoadley, K. A.; Um, K. S.; Elburn, K.; Holloway, K.; Wang, L.; Blumenberg, L.; Hannick, L.; Qi, L.; Sokoll, L. J.; Cornwell, M.; Loriaux, M.; Domagalski, M. J.; Gritsenko, M. A.; Anderson, M.; Monroe, M. E.; Ellis, M. J.; Dyer, M.; Anurag, M.; Burke, M. C.; Borucki, M.; Gillette, M. A.; Birrer, M. J.; Lewis, M.; Ittmann, M. M.; Smith, M.; Vernon, M.; Chaikin, M.; Chheda, M. G.; Khan, M.; Roche, N.; Edwards, N. J.; Vatanian, N.; Tignor, N.; Beckmann, N.; Grady, P.; Castro, P.; Piehowski, P.; Mcgarvey, P. B.; Mieczkowski, P.; Hariharan, P.; Gao, Q.; Dhir, R.; Kothadia, R. B.; Thangudu, R. R.; Montgomery, R.; Jayasinghe, R. G.; Smith, R. D.; Edwards, R.; Zelt, R.; Bremner, R.; Liu, R.; Hong, R.; Mareedu, S.; Payne, S. H.; Cottingham, S.; Markey, S. P.; Jewell, S. D.; Patel, S.; Satpathy, S.; Richey, S.; Davies, S. R.; Cai, S.; Boca, S. M.; Patil, S.; Sengupta, S.; Carter, S.; Gabriel, S.; Thomas, S. N.; De Young, S.; Stein, S. E.; Carr, S. A.; Foltz, S. M.; Hilsenbeck, S.; Krubit, T.; Liu, T.; Skelly, T.; Westbrook, T.; Borate, U.; Velvulou, U.; Petyuk, V. A.; Bocik, W. E.; Chen, X.; Shi, Y.; Geffen, Y.; Lu, Y.; Wang, Y.; Maruvka, Y.; Li, Z.; Shi, Z.; Tu, Z. Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell* **2019**, *179*, 964.

(22) Vasaikar, S.; Huang, C.; Wang, X.; Petyuk, V. A.; Savage, S. R.; Wen, B.; Dou, Y.; Zhang, Y.; Shi, Z.; Arshad, O. A.; Gritsenko, M. A.; Zimmerman, L. J.; Mcdermott, J. E.; Clauss, T. R.; Moore, R. J.; Zhao, R.; Monroe, M. E.; Wang, Y.; Chambers, M. C.; Slebos, R. J. C.; Lau, K. S.; Mo, Q.; Ding, L.; Ellis, M.; Thiagarajan, M.; Kinsinger, C. R.; Rodriguez, H.; Smith, R. D.; Rodland, K. D.; Liebler, D. C.; Liu, T.; Zhang, B.; Pandey, A.; Paulovich, A.; Hoofnagle, A.; Mani, D. R.; Chan, D. W.; Ransohoff, D. F.; Fenyo, D.; Tabb, D. L.; Levine, D. A.; Boja, E. S.; Kuhn, E.; White, F. M.; Whiteley, G. A.; Zhu, H.; Zhang, H.; Shih, I.; Bavarva, J.; Whiteaker, J.; Ketchum, K. A.; Clauser, K. R.; Ruggles, K.; Elburn, K.; Hannick, L.; Watson, M.; Oberti, M.; Mesri, M.; Sanders, M. E.; Borucki, M.; Gillette, M. A.; Snyder, M.; Edwards, N. J.; Vatanian, N.; Rudnick, P. A.; Mcgarvey, P. B.; Mertins, P.; Townsend, R. R.; Thangudu, R. R.; Rivers, R. C.; Payne, S. H.; Davies, S. R.; Cai, S.; Stein, S. E.; Carr, S. A.; Skates, S. J.; Madhavan, S.; Hiltke, T.; Chen, X.; Zhao, Y.; Wang, Y.; Zhang, Z. Proteogenomic Analysis of Human Colon Cancer Reveals New Therapeutic Opportunities. *Cell* **2019**, *177*, 1035.

(23) Chen, Y.-J.; Roumeliotis, T. I.; Chang, Y.; Chen, C.; Han, C.; Lin, M.; Chen, H.; Chang, G.; Chang, Y.; Wu, C.; Lin, M.; Hsieh, M.; Wang, Y.; Chen, Y.; Jonassen, I.; Ghavidel, F. Z.; Lin, Z.; Lin, K.; Chen, C.; Sheu, P.; Hung, C.; Huang, K.; Yang, H.; Lin, P.; Yen, T.; Lin, Y.; Wang, J.; Raghav, L.; Lin, C.; Chen, Y.; Wu, P.; Lai, C.; Weng, S.; Su, K.; Chang, W.; Tsai, P.; Robles, A. I.; Rodriguez, H.; Hsiao, Y.; Chang, W.; Sung, T.; Chen, J.; Yu, S.; Choudhary, J. S.; Chen, H.; Yang, P.; Chen, Y. Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression. *Cell* **2020**, *182*, 226.

(24) Dou, Y.; Kawaler, E. A.; Cui Zhou, D.; Gritsenko, M. A.; Huang, C.; Blumenberg, L.; Karpova, A.; Petyuk, V. A.; Savage, S. R.; Satpathy, S.; Liu, W.; Wu, Y.; Tsai, C.; Wen, B.; Li, Z.; Cao, S.; Moon, J.; Shi, Z.; Cornwell, M.; Wyczalkowski, M. A.; Chu, R. K.; Vasaikar, S.; Zhou, H.; Gao, Q.; Moore, R. J.; Li, K.; Sethuraman, S.; Monroe, M. E.; Zhao, R.; Heiman, D.; Krug, K.; Clauser, K.; Kothadia, R.; Maruvka, Y.; Pico, A. R.; Oliphant, A. E.; Hoskins, E. L.; Pugh, S. L.; Beecroft, S. J. I.; Adams, D. W.; Jarman, J. C.; Kong, A.; Chang, H.; Reva, B.; Liao, Y.; Rykunov, D.; Colaprico, A.; Chen, X. S.; Czekanski, A.; Jedryka, M.; Matkowski, R.; Wiznerowicz, M.; Hiltke, T.; Boja, E.; Kinsinger, C. R.; Mesri, M.; Robles, A. I.; Rodriguez, H.; Mutch, D.; Fuh, K.; Ellis, M. J.; DeLair, D.; Thiagarajan, M.; Mani, D. R.; Getz, G.; Noble, M.; Nesvizhskii, A. I.; Wang, P.; Anderson, M. L.; Levine, D. A.; Smith, R. D.; Payne, S. H.; Ruggles, K. V.; Rodland, K. D.; Ding, L.; Zhang, B.; Liu, T.; Fenyo, D.; Clinical Proteomic Tumor Analysis Consortium. Proteogenomic Characterization of Endometrial Carcinoma. *Cell* **2020**, *180*, 729−748.

(25) Krug, K.; Jaehnig, E. J.; Satpathy, S.; Blumenberg, L.; Karpova, A.; Anurag, M.; Miles, G.; Mertins, P.; Geffen, Y.; Tang, L. C.; Heiman, D. I.; Cao, S.; Maruvka, Y. E.; Lei, J. T.; Huang, C.; Kothadia, R. B.; Colaprico, A.; Birger, C.; Wang, J.; Dou, Y.; Wen, B.; Shi, Z.; Liao, Y.; Wiznerowicz, M.; Wyczalkowski, M. A.; Chen, X. S.; Kennedy, J. J.; Paulovich, A. G.; Thiagarajan, M.; Kinsinger, C. R.; Hiltke, T.; Boja, E. S.; Mesri, M.; Robles, A. I.; Rodriguez, H.; Westbrook, T. F.; Ding, L.; Getz, G.; Clauser, K. R.; Fenyö, D.; Ruggles, K. V.; Zhang, B.; Mani, D. R.; Carr, S. A.; Ellis, M. J.; Gillette, M. A.; Avanessian, S. C.; Cai, S.; Chan, D.; Chen, X.; Edwards, N. J.; Hoofnagle, A. N.; Kane, M. H.; Ketchum, K. A.; Kuhn, E.; Levine, D. A.; Li, S.; Liebler, D. C.; Liu, T.; Luo, J.; Madhavan, S.; Maher, C.; Mcdermott, J. E.; Mcgarvey, P. B.; Oberti, M.; Pandey, A.; Payne, S. H.; Ransohoff, D. F.; Rivers, R. C.; Rodland, K. D.; Rudnick, P.; Sanders, M. E.; Shaw, K. M.; Shih, I.; Slebos, R. J. C.; Smith, R. D.; Snyder, M.; Stein, S. E.; Tabb, D. L.; Thangudu, R. R.; Thomas, S.; Wang, Y.; White, F. M.; Whiteaker, J. R.; Whiteley, G. A.; Zhang, H.; Zhang, Z.; Zhao, Y.; Zhu, H.; Zimmerman, L. J. Proteogenomic Landscape of Breast Cancer Tumorigenesis and Targeted Therapy. *Cell* **2020**, *183*, 1436.

(26) Liu, H.; Ma, Z.; Lih, T. M.; Chen, L.; Hu, Y.; Wang, Y.; Sun, Z.; Huang, Y.; Xu, Y.; Zhang, H. New Method Enhanced Extraction of Protein Signatures of Renal Cell Carcinoma from Proteomics Data. *J. Proteome Res.* **2025**, *24*, 5710−5726.

(27) Leitão, M. d. C.; Cabral, L. S.; Piva, L. C.; Queiroz, P. F. D. S.; Gomes, T. G.; De Andrade, R. V.; Perez, A. L. A.; De Paiva, K. L. R.; Báo, S. N.; Reis, V. C. B.; Moraes, L. M. P.; Togawa, R. C.; Barros, L. M. G.; Torres, F. A. G.; Pappas, G. J., Júnior; Coelho, C. M. SHIP identifies genomic safe harbors in eukaryotic organisms using genomic general feature annotation. *Sci. Rep.* **2025**, *15*, No. 7193, DOI: 10.1038/s41598-025-91249-9.

(28) Sayler, A. L.; Dean, H.; Hammond, J. R. Impact of the loss of slc43a3 on 6-mercaptopurine absorption and tissue distribution in mice. *Drug Metab. Dispos.* **2025**, *53*, No. 100054, DOI: 10.1016/j.dmd.2025.100054.

(29) Li, L.; Wu, Z.; Guarracino, A.; Villani, F.; Kong, D.; Mancieri, A.; Zhang, A.; Saba, L.; Chen, H.; Brozka, H.; Vales, K.; Senko, A. N.; Kempermann, G.; Stuchlik, A.; Pravenec, M.; Lechner, J.; Prins, P.; Mathur, R.; Lu, L.; Yang, K.; Peng, J.; Williams, R. W.; Wang, X. Genetic modulation of protein expression in rat brain. *iScience* **2025**, *28*, No. 112079, DOI: 10.1016/j.isci.2025.112079.

(30) Yu, X.-Q.; Yang, H.; Feng, H.; Hou, J.; Tian, J.; Niu, S.; You, C.; Tao, X.; Zhang, S.; Wang, Z.; He, Y. Targeting efflux pumps prevents the multi-step evolution of high-level resistance to fluoroquinolone in Pseudomonas aeruginosa. *Microbiol. Spectr.* **2025**, *13*, No. e02981-24, DOI: 10.1128/spectrum.02981-24.

(31) Van Wijk, K. J.; Leppert, T.; Sun, Z.; Kearly, A.; Li, M.; Mendoza, L.; Guzchenko, I.; Debley, E.; Sauermann, G.; Routray, P.; Malhotra, S.; Nelson, A.; Sun, Q.; Deutsch, E. W. Detection of the Arabidopsis Proteome and Its Post-translational Modifications and the Nature of the Unobserved (Dark) Proteome in PeptideAtlas. *J. Proteome Res.* **2024**, *23*, 185.

(32) Schwenk, J. M.; Omenn, G. S.; Sun, Z.; Campbell, D. S.; Baker, M. S.; Overall, C. M.; Aebersold, R.; Moritz, R. L.; Deutsch, E. W. The Human Plasma Proteome Draft of 2017: Building on the Human Plasma PeptideAtlas from Mass Spectrometry and Complementary Assays. *J. Proteome Res.* **2017**, *16*, 4299.

(33) Wang, M.; Wang, J.; Carver, J.; Pullman, B. S.; Cha, S. W.; Bandeira, N. Assembling the Community-Scale Discoverable Human Proteome. *Cell Syst.* **2018**, *7*, 412.

(34) Zahn-Zabal, M.; Michel, P.; Gateau, A.; Nikitin, F.; Schaeffer, M.; Audot, E.; Gaudet, P.; Duek, P. D.; Teixeira, D.; Rech de laval, V.; Samarasinghe, K.; Bairoch, A.; Lane, L. The neXtProt knowledgebase in 2020: data, tools and usability improvements. *Nucleic Acids Res.* **2019**, *48*, D328 DOI: 10.1093/nar/gkz995.

(35) Farrah, T.; Deutsch, E. W.; Omenn, G. S.; Campbell, D. S.; Sun, Z.; Bletz, J. A.; Mallick, P.; Katz, J. E.; Malmström, J.; Ossola, R.; Watts, J. D.; Lin, B.; Zhang, H.; Moritz, R. L.; Aebersold, R. A High-Confidence Human Plasma Proteome Reference Set with Estimated Concentrations in PeptideAtlas. *Mol. Cell. Proteomics* **2011**, *10*, No. M110.006353, DOI: 10.1074/mcp.M110.006353.

(36) Deutsch, E. W.; Omenn, G. S.; Sun, Z.; Maes, M.; Pernemalm, M.; Palaniappan, K. K.; Letunica, N.; Vandenbrouck, Y.; Brun, V.; Tao, S.; Yu, X.; Geyer, P. E.; Ignjatovic, V.; Moritz, R. L.; Schwenk, J.

https://doi.org/10.1021/acs.jproteome.5c01259
J. Proteome Res. 2026, 25, 2901−2914

M. Advances and Utility of the Human Plasma Proteome. *J. Proteome Res.* **2021**, *20*, 5241.
(37) Khan, A. A.; Mangalaparthi, K. K.; Advani, J.; Prasad, T.S. K.; Gowda, H.; Jain, D.; Chatterjee, A. Data from quantitative proteomic analysis of lung adenocarcinoma and squamous cell carcinoma primary tissues using high resolution mass spectrometry. *Data in Brief* **2018**, *19*, 1631.
(38) Franciosa, G.; Nieddu, V.; Battistini, C.; Caffarini, M.; Lupia, M.; Colombo, N.; Fusco, N.; Olsen, J. V.; Cavallaro, U. Quantitative Proteomics and Phosphoproteomics Analysis of Patient-Derived Ovarian Cancer Stem Cells. *Mole. Cell. Proteomics* **2025**, *24*, 100965.
(39) Hu, Y.; Pan, J.; Shah, P.; Ao, M.; Thomas, S. N.; Liu, Y.; Chen, L.; Schnaubelt, M.; Clark, D. J.; Rodriguez, H.; Boja, E. S.; Hiltke, T.; Kinsinger, C. R.; Rodland, K. D.; Li, Q. K.; Qian, J.; Zhang, Z.; Chan, D. W.; Zhang, H.; Pandey, A.; Paulovich, A.; Hoofnagle, A.; Zhang, B.; Mani, D. R.; Liebler, D. C.; Ransohoff, D. F.; Fenyo, D.; Tabb, D. L.; Levine, D. A.; Kuhn, E.; White, F. M.; Whiteley, G. A.; Zhu, H.; Shih, I.; Bavarva, J.; Mcdermott, J. E.; Whiteaker, J.; Ketchum, K. A.; Clauser, K. R.; Ruggles, K.; Elburn, K.; Ding, L.; Hannick, L.; Zimmerman, L. J.; Watson, M.; Thiagarajan, M.; Ellis, M. J. C.; Oberti, M.; Mesri, M.; Sanders, M. E.; Borucki, M.; Gillette, M. A.; Snyder, M.; Edwards, N. J.; Vatanian, N.; Rudnick, P. A.; Mcgarvey, P. B.; Mertins, P.; Townsend, R. R.; Thangudu, R. R.; Smith, R. D.; Rivers, R. C.; Slebos, R. J. C.; Payne, S. H.; Davies, S. R.; Cai, S.; Stein, S. E.; Carr, S. A.; Skates, S. J.; Madhavan, S.; Liu, T.; Chen, X.; Zhao, Y.; Wang, Y.; Shi, Z. Integrated Proteomic and Glycoproteomic Characterization of Human High-Grade Serous Ovarian Carcinoma. *Cell Rep.* **2020**, *33*, No. 108276, DOI: 10.1016/j.celrep.2020.108276.
(40) Wu, D.; Gong, P.; Zeng, Q.; Zhang, W.; Xie, F.; Zhou, X. Prognostic implication of proteomic profiles in head and neck squamous cell carcinoma. *Clinica Chimica Acta* **2020**, *509*, 304.
(41) Stella, M.; Chinello, C.; Cazzaniga, A.; Smith, A.; Galli, M.; Piga, I.; Grasso, A.; Grasso, M.; Del Puppo, M.; Varallo, M.; Bovo, G.; Magni, F. Histology-guided proteomic analysis to investigate the molecular profiles of clear cell Renal Cell Carcinoma grades. *Journal of Proteomics* **2019**, *191*, 38.
(42) Hong, Z.; Wang, T.; Wang, W.; Jing, H.; Tang, H.; Xu, M.; Pan, C.; Mu, X.; Zhang, D.; Gao, G.; Gao, Z.; Luo, H.; Zhou, Y. Proteomic Profiling and Tumor Microenvironment Characterization Reveal Molecular and Immunological Hallmarks of Left-Sided and Right-Sided Colon Cancer Tumorigenesis. *J. Proteome Res.* **2023**, *22*, 2973.
(43) Coles, G. L.; Cristea, S.; Webber, J. T.; Levin, R. S.; Moss, S. M.; He, A.; Sangodkar, J.; Hwang, Y. C.; Arand, J.; Drainas, A. P.; Mooney, N. A.; Demeter, J.; Spradlin, J. N.; Mauch, B.; Le, V.; Shue, Y. T.; Ko, J. H.; Lee, M. C.; Kong, C.; Nomura, D. K.; Ohlmeyer, M.; Swaney, D. L.; Krogan, N. J.; Jackson, P. K.; Narla, G.; Gordan, J. D.; Shokat, K. M.; Sage, J. Unbiased Proteomic Profiling Uncovers a Targetable GNAS/PKA/PP2A Axis in Small Cell Lung Cancer Stem Cells. *Cancer Cell* **2020**, *38*, 129.
(44) Cochrane, D. R.; Negri, G. L.; Huvila, J.; Kalantari, F.; Farnell, D. A.; Mohammad, N.; Thompson, E.; Yang, W.; Lum, A.; Spencer, S. E.; Riley, R.; Jamieson, A.; Leung, S.; Chiu, D.; Chow, C.; Lim, J. L.P.; Kobel, M.; Kommoss, S.; Kommoss, F.; Gilks, B.; Hoang, L.; Huntsman, D. G.; Morin, G. B.; McAlpine, J. N. Proteomic analysis uncovers biological diversity in molecularly defined endometrial carcinomas. *Neoplasia* **2025**, *69*, 101229.
(45) Moreno-Ulloa, A.; Zarate-Cordova, V. L.; Ramirez-Sanchez, I.; Cruz-Lopez, J. C.; Perez-Ortiz, A.; Villarreal-Garza, C.; Diaz-Chavez, J.; Estrada-Mena, B.; Antonio-Aguirre, B.; Lopez-Almanza, P. X.; Lira-Romero, E.; Estrada-Mena, F.. J. Evaluation of a Proteomics-Guided Protein Signature for Breast Cancer Detection in Breast Tissue. *J. Proteome Res.* **2024**, *23*, 4907.
(46) Satpathy, S.; Jaehnig, E. J.; Krug, K.; Kim, B.; Saltzman, A. B.; Chan, D. W.; Holloway, K. R.; Anurag, M.; Huang, C.; Singh, P.; Gao, A.; Namai, N.; Dou, Y.; Wen, B.; Vasaikar, S. V.; Mutch, D.; Watson, M. A.; Ma, C.; Ademuyiwa, F. O.; Rimawi, M. F.; Schiff, R.; Hoog, J.; Jacobs, S.; Malovannaya, A.; Hyslop, T.; Clauser, K. R.; Mani, D. R.; Perou, C. M.; Miles, G.; Zhang, B.; Gillette, M. A.; Carr, S. A.; Ellis, M. J. Microscaled proteogenomic methods for precision oncology. *Nat. Commun.* **2020**, *11*, No. 532, DOI: 10.1038/s41467-020-14381-2.
(47) Feilen, T.; Rogg, M.; Andreev, G.; Pinter, N.; Wess, M.; Kossinger, A. L.; Diel, N.; Neumann-Haefelin, E.; Ganner, A.; Grabbert, M.; Schell, C.; Schilling, O. Proteomic characterization of the pseudocapsule of clear cell renal cell carcinoma in VHL disease reveals a distinct microenvironment at the tumor boundary zone. *Neoplasia* **2025**, *68*, 101214.
(48) Wang, Y.; Lih, T. M.; Lee, J. W.; Ohtsuka, T.; Hozaka, Y.; Mino-Kenudson, M.; Adsay, N. V.; Luchini, C.; Scarpa, A.; Maker, A. V.; Kim, G. E.; Paulino, J.; Chen, L.; Woo, J.; Jiao, L.; Sun, Z.; Goodman, D.; Pflüger, M. J.; Roberts, N. J.; Matthaei, H.; Goggins, M. G.; Wood, L. D.; Furukawa, T.; Zhang, H.; Hruban, R. H. Multi-omic profiling of intraductal papillary neoplasms of the pancreas reveals distinct patterns and potential markers of progression. *Cancer Cell* **2025**, *43*, 2156.
(49) Omenn, G. S.; Orchard, S.; Lane, L.; Lindskog, C.; Pineau, C.; Overall, C. M.; Budnik, B.; Mudge, J. M.; Packer, N. H.; Weintraub, S. T.; Roehrl, M. H. A.; Nice, E.; Guo, T.; Van Eyk, J. E.; Völker, U.; Zhang, G.; Bandeira, N.; Aebersold, R.; Moritz, R. L.; Deutsch, E. W. The 2024 Report on the Human Proteome from the HUPO Human Proteome Project. *J. Proteome Res.* **2024**, *23*, 5296.
(50) Baril, S. A.; Wilson, K. A.; Shaik, M. M.; Fukuda, Y.; Umans, R. A.; Barbieri, A.; Lynch, J.; Gose, T.; Myasnikov, A.; Oldham, M. L.; Wang, Y.; Zhu, J.; Fang, J.; Zuo, J.; Kalathur, R. C.; Ford, R. C.; Coffin, A.; Taylor, M. R.; O'mara, M. L.; Schuetz, J. D. The role of ATP-binding Cassette subfamily B member 6 in the inner ear. *Nat. Commun.* **2024**, *15*, No. 9885, DOI: 10.1038/s41467-024-53663-x.
(51) Duek, P.; Gateau, A.; Bairoch, A.; Lane, L. Exploring the Uncharacterized Human Proteome Using neXtProt. *J. Proteome Res.* **2018**, *17*, 4211.
(52) Otake, H.; Yamamoto, T.; Yamamoto, N.; Nakazawa, Y.; Miyata, Y.; Taga, A.; Sasaki, H.; Nagai, N. Changes in Protein Expression in Warmed Human Lens Epithelium Cells Using Shotgun Proteomics. *Medicina* **2025**, *61*, No. 286, DOI: 10.3390/medicina61020286.