Nucleic Acids Research, 2024, **52**, D1155–D1162
https://doi.org/10.1093/nar/gkad824
Advance access publication date: 12 October 2023
Database issue

CancerProteome: a resource to functionally decipher the proteome landscape in cancer

Dezhong Lv[1,†], Donghao Li[1,†], Yangyang Cai[1,†], Jiyu Guo[2,†], Sen Chu[1], Jiaxin Yu[1], Kefan Liu[2], Tiantongfei Jiang[1], Na Ding[1], Xiyun Jin[3,*], Yongsheng Li[2,*] and Juan Xu[1,*]

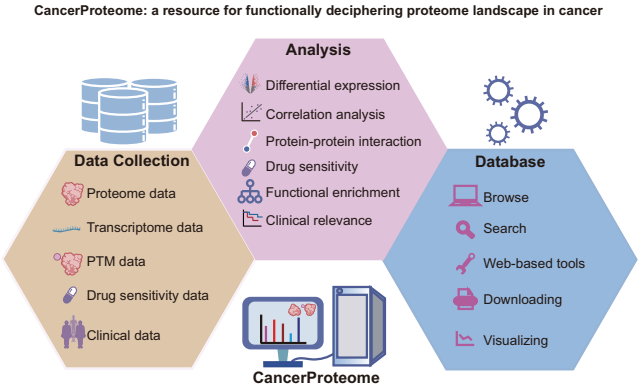
[1]College of Bioinformatics Science and Technology, Harbin Medical University, Harbin, Heilongjiang Province 150081, China
[2]School of Interdisciplinary Medicine and Engineering, Harbin Medical University, Harbin, Heilongjiang Province 150081, China
[3]School of Life Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang Province 150000, China

[*]To whom correspondence should be addressed. Tel: +86 13654559904; Fax: +86 13654559904; Email: xujuanbiocc@ems.hrbmu.edu.cn
Correspondence may also be addressed to Yongsheng Li. Tel: +86 13604805482; Email: liyongsheng@ems.hrbmu.edu.cn
Correspondence may also be addressed to Xiyun Jin. Tel: +86 18845643158; Email: vaejin_8811@163.com
[†]Joint Authors

Abstract

Advancements in mass spectrometry (MS)-based proteomics have greatly facilitated the large-scale quantification of proteins and microproteins, thereby revealing altered signalling pathways across many different cancer types. However, specialized and comprehensive resources are lacking for cancer proteomics. Here, we describe CancerProteome (http://bio-bigdata.hrbmu.edu.cn/CancerProteome), which functionally deciphers and visualizes the proteome landscape in cancer. We manually curated and re-analyzed publicly available MS-based quantification and post-translational modification (PTM) proteomes, including 7406 samples from 21 different cancer types, and also examined protein abundances and PTM levels in 31 120 proteins and 4111 microproteins. Six major analytical modules were developed with a view to describe protein contributions to carcinogenesis using proteome analysis, including conventional analyses of quantitative and the PTM proteome, functional enrichment, protein–protein associations by integrating known interactions with co-expression signatures, drug sensitivity and clinical relevance analyses. Moreover, protein abundances, which correlated with corresponding transcript or PTM levels, were evaluated. CancerProteome is convenient as it allows users to access specific proteins/microproteins of interest using quick searches or query options to generate multiple visualization results. In summary, CancerProteome is an important resource, which functionally deciphers the cancer proteome landscape and provides a novel insight for the identification of tumor protein markers in cancer.

Graphical abstract



Introduction

Precision medicine depends on the recognition of specific molecular alterations from which patients are stratified and effective therapeutic options selected. In recent years, both genomics and transcriptomics have become systematic in the exploration of cancer vulnerability (1,2), however, the intricate nature of cancer suggests that genomics alone cannot sufficiently guide the clinical treatment of cancer patients (3).

Proteins are executors encoded by a cell's genome and are integral to many different biological processes (4). While often used as a surrogate for protein expression, average RNA expression data has proved to be a poor predictor of protein expression (4). Directly assaying proteins provides unique insights into the dynamic molecular behavior of cells and enhances our understanding of genotype-to-phenotype relationships beyond the genome and transcriptome. Importantly,

Received: August 12, 2023. **Revised:** September 7, 2023. **Editorial Decision:** September 11, 2023. **Accepted:** September 20, 2023

recent studies have revealed that the protein-coding capacity of the human genome has been largely under-reported (5–7), with an increasing number of novel functional microproteins encoded from non-coding regions (8) and characterized as either oncogenic drivers or tumor suppressors in cancer (9). Thus, comprehensive analyses of proteomic datasets have allowed the identification of new cancer biomarkers and potentially improved diagnostic and treatment choices for clinicians and patients. However, high-quality proteomic datasets have been lagged behind RNA expression profiling approaches.

Recent improvements in mass spectrometry (MS)-based proteomics have facilitated the measurement of global protein and microprotein abundance, and also post-translational modifications (PTMs) (10). Thousands of cell lines and clinical samples derived from tumor tissue biopsies have undergone standardized quantification procedures. Several proteome databases have been developed to highlight dynamics/changes of protein abundances in normal or disease states, or identify correlations with drugs, including the Human Protein Atlas (11), ProteomicsDB (12), TCRD and Pharos (13) and Expression Atlas (14). Human Protein Atlas (HPA) is a spatial map of the human proteome, which has constituted a tool for researchers studying the location and expression of proteins in human tissues and cells (11). TCRD and Pharos have produced two major releases, which incorporated human and viral-human protein–protein interactions (PPIs), protein–disease and protein–phenotype associations (13). Additionally, qPTM (15), dbPTM (16) and VPTMdb (17) have been introduced to specifically quantify the PTM events under different conditions or explore associations with molecular features. In particular, dbPTM database has integrated experimentally validated PTMs and also provides PTMdisease associations based on non-synonymous single nucleotide polymorphisms (16). VPTMdb is constructed for collecting systematic information of PTMs in human viruses and infected host cells (17). Despite the valuable information these data resources provide, they were primarily constructed for general purposes, while a specialized and comprehensive cancer proteome resource remains to be developed.

Here, we describe the CancerProteome database (http://bio-bigdata.hrbmu.edu.cn/CancerProteome), which functionally deciphers or visualizes the proteome landscape in cancer. Proteins and microproteins (or (micro-)protein, hereafter) expression and PTM levels were measured by re-analyzing raw MS datasets against canonical proteins and our integrated microprotein theory library. CancerProteome provided not only dysregulation information on (micro-)protein activities across cancer types, but also provided multiple insights toward carcinogenesis. (Micro-)proteins with differential expression/modifications in different cancer types were comprehensively identified, and functional enrichment analyses performed. We also investigated correlations between (micro-)protein abundances and corresponding transcript levels using RNA sequencing (RNA-seq) integrative data analyses and PTMs regulatory relationships with (micro-)protein abundances. Moreover, functional associations between (micro-)proteins were identified by integrating co-expression information in different cancers with known PPIs. Finally, we identified (micro-)proteins related to drug sensitivity and clinical relevance.

Materials and methods

Collection of MS data

We collected all available clinical samples and cell lines proteome datasets from seven widely used proteome resources, such as CPTAC (https://pdc.cancer.gov), PRIDE (18), MassIVE.quant (19), PeptideAtlas (20), jPOST (21), Panorama (22) and iProX (23). Raw data and clinically related information were downloaded for in-depth analyses. In total, we collected 102 datasets, including 7406 samples across 21 different cancer types (Supplementary Table S1). Both quantitative and PTM proteomes across 21 cancer types were re-analyzed. Additionally, we acquired transcriptome datasets of corresponding patients from 10 cancer types, including 1560 tumor and 368 control samples.

Known cancer genes

To identify correlations between known cancer genes and proteins from multi-omics data, CancerProteome compiled known cancer genes from several databases, including COSMIC (24), Lnc2cancer 3.0 (25) and OncoKB (26). In total, 3749 known cancer genes were collected. Also, genes encoding proteins that overlapped with known cancer genes were similarly collected and stored in CancerProteome.

Drug-related information

To uncover all pairwise associations between (micro-)proteins and drug sensitivity in cancer cell lines (4), CancerProteome retrieved half-maximal inhibitory concentration (IC_{50}) measurements for each drug across all cell lines from the DepMap Portal (https://depmap.org/portal/). In total, 413 unique drugs were included in CancerProteome and the natural log of raw IC_{50} data, was used for subsequent analyses.

Constructing a (micro-)proteins theory library

We integrated all ribo-seq-supported open reading frames (ORFs) from RPFdb (27), nuORFdb (28), TransLnc (29), and IEAtlas (30), along with their basic annotations. All ORFs with 'NTG' start codons and 'TAA/TGA/TAG' stop codons were included in our analyses. Based on ORF genome coordinates and corresponding annotation files, we generated ORF sequences using the 'getblast' function in the R 'bedtoolsr' package using default parameters. Only ORFs that generated peptides $\geq$8 amino acids (aa) but $\leq$100 aa were retained and defined as microproteins according to previous studies (31,32). Those microproteins whose sequences were entirely contained within other (micro-)proteins were removed. After integration with protein information from the human UniProt database (33), a benchmarked protein and microprotein theory library was constructed for further analysis.

Database searches for MS data

CancerProteome re-analyzed MS-based proteomic data using MaxQuant (v.2.1.0.0) (34), referencing the benchmarked protein and microprotein theory library. Variable modifications were set at oxidized methionine, protein N-terminal acetylation and asparagine and glutamine deamidation. Carbamidomethyl cysteine was searched for as a fixed modification (10,35). For the phosphoproteome dataset, variable modifications included serine, threonine, and tyrosine phos-

phorylation (36). For the glycopeptide enriched data analyses, asparagine (Asn) deamidation in $H_2{}^{18}O$ (^{18}O tag of Asn, +2.9890 Da) was set as an additional variable modification (37). For acetylproteome and ubiquitylproteome datasets, searches were completed using acetylation and diglycine of lysine residues as variable modifications, respectively (36,38).

Data normalization and missing value imputation

For each identified (micro-)protein, expression abundance was considered as the protein intensity after performing $\log_2$ transformation and normalization steps. A median centering method was used to normalize MS data by label-free technologies, while reference channel intensity was used for MS data assayed by labeling technologies, and normalized ratios were calculated as the subtraction between $\log_2$ report ion intensities and $\log_2$ reference intensity. Final expression abundance as assayed by labeling technologies was calculated as the sum of normalized ratios and $\log_2$-transformed mean reference intensity (39). To ensure samples had sufficient data for imputation, we retained proteins and modification sites having $\leq$50% missing data. Based on the 'impute.knn' function in the R package, missing values were imputed in the cancer proteome.

Differential expression analysis

We used the eBayes method in the 'limma' R package to implement differential expression analyses across tumor and control samples of quantitative and PTM proteomes, and the transcriptome (40). The false discovery rate (FDR) threshold was set to 0.05 and expression/PTM changes required 1.5-fold changes. Differential analyses were also performed based on clinical features in datasets without normal samples, such as pre- and post-treatment, pre- and on-treatment, responsive and unresponsive to drugs, classic and desmo, squamous and translational.

Correlation analysis

We obtained the transcriptome data from the supplementary tables of associated publications. If the transcriptome data was not normalized, we normalized the expression across samples by absolute deviations from the median within each cancer type and dataset. A rich body of researches demonstrate that the mRNA and protein abundances frequently have a poorer correlation than expected, a consequence of both translational and post-translational regulation. Spearman's correlation analysis was used to measure correlations between protein abundance and corresponding transcript expression of a given protein, or to measure correlations between protein abundance and drug IC_{50} values. Additionally, the 'lm' method in R was used to assess correlations between protein abundance and modification levels at sites.

Protein–protein association analysis

Protein co-expression and PPIs were used to evaluate protein–protein associations in different cancers. First, significant co-expressed (micro-)protein pairs were identified in the specific cancer using Spearman's correlation and aforementioned thresholds. We also assembled and integrated known PPIs from STRING (41), MINT (42) and IntAct (43), and only high-confidence interactions were retained in STRING.

Functional enrichment analysis

To further analyze the molecular functions of quantitative and PTM proteomes and the transcriptome, function enrichment analyses were performed for (micro-)proteins with differential expression/modifications using the 'enricher' method in the clusterProfiler R package (44).

Clinical relevance

Clinical information of cancer patients was obtained from the clinical proteomic tumor analysis consortium (CPTAC) and supplementary tables of associated literature. The survival time, tumor or normal, treatment information and response to therapy were obtained. We next performed Cox regression analyses and log-rank tests to identify survival-related (micro-)proteins and PTM sites using univariate and multivariate Cox regression analyses. Cancer patients were ranked based on the abundance of a given protein or site, and survival differences between high- and low-expression groups were evaluated using the Kaplan–Meier method.

Database implementation

The CancerProteome backed server was constructed and accessed based on Java Server Pages with the Tomcat container (v.6.0). CancerProteome used the MySQL database (v.5.5.48) for documenting and managing all metadata and the web frontend was constituted with HTML, JavaScript and CSS code, containing jQuery (v.3.3.1), Datatable (v.1.10.25) and ECharts (v.5.5.1) plugins to visualize all analysis results and multiple statistical tables. The R framework (v.3.6.3) was performed for statistical analyses. CancerProteome was tested on several popular web browsers, including Google Chrome (preferred), Firefox and Apple Safari browsers.

Database content

CancerProteome was used to curate and re-analyze publicly available MS-based raw cancer datasets at the proteome level (Figure 1A and Supplementary Table S1). Quantitative proteome data came from 2708 tumor samples and 1752 control samples across 21 cancer types. The PTM proteome consisted of 1632 tumor samples and 1012 control samples across nine cancer types, and 302 proteomes spanning 14 cancer types in cell lines. Most cancers have thousands of samples (Figure 1B), and thousands of proteins/PTM sites were quantified in each cancer (Figure 1C), in total including 84 257 PTM sites, 31 120 proteins and 4111 microproteins. We also observed that a large number of microproteins were expressed in cancer (Figure 1D). Next, cancer-related (micro-)proteins were further identified from three aspects, with 13 673 proteins and 742 microproteins. As shown in Figure 1E, proteome contributions were different across cancer types. In total, 20 737 differential PTMs were identified and differential distributions across cancers were revealed (Figure 1F). We found that PTM sites on microproteins were also cancer-related and involved in 16.6% microproteins. Furthermore, (micro-)protein abundances correlations with corresponding transcript levels were calculated using RNA-seq data integrative analysis, and 15 816 co-expressed pairs were identified. The underlying regulatory effects of PTMs on (micro-)protein abundances were also estimated from 29 156 correlations, indicating wide protein PTMs across the different cancers.

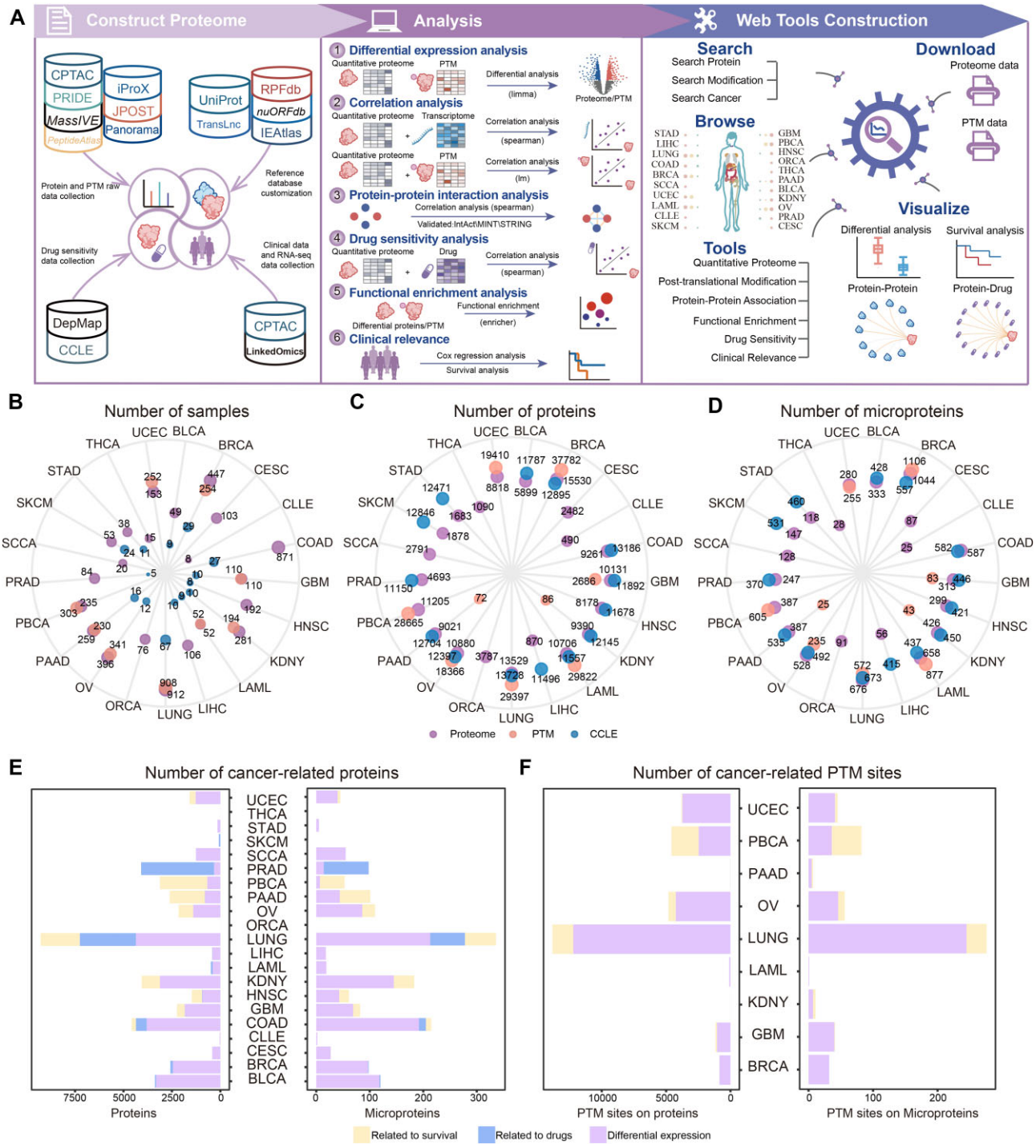


Figure 1. Schematic showing the CancerProteome design. **(A)** Data collection and workflow showing the construction of the CancerProteome resource. **(B)** Number of samples in the proteome, PTM and cell line proteomes. **(C)** Number of proteins detected in the proteome, PTM and cell line proteomes. **(D)** Number of microproteins identified in the proteome, PTM and cell line proteomes. **(E)** Number of cancer-related proteins across the different cancer types. **(F)** Number of cancer-related PTM sites across the different cancer types.

To identify functional abnormalities caused by proteomic abnormalities, we used differential (micro-)protein/PTM sites to perform functional enrichment analysis in each cancer. Considering that microprotein function associations were unknown, co-expression and PPIs were analyzed in each cancer. CancerProteome collected 54 146 014 co-expression signatures in cancer, which were supported by known PPIs (Supplementary Fig. S1A). Moreover, 12 437 drug–protein associations were used to estimate potential (micro-)protein roles related to drug sensitivity (Supplementary Fig. S1B).

Web interface

CancerProteome offers a user-friendly interface that allows users to search, browse, visualize and download data (Figures 1A and 2). The 'Browse' page allows users to explore

Figure 2. The CancerProteome user interface. **(A)** Search page for proteins, PTMs and cancers by genes or resource. **(B)** Global results for differential expression and functional analysis. **(C)** Differential PTM sites in cancer. **(D)** Differential expression analysis results. **(E)** Basic information on proteins identified in the cancer proteome. **(F)** Protein expression abundance and the associations with transcripts in cancer. **(G)** The modification levels of PTM sites derived from proteins and associations between protein expression and PTMs. **(H)** PPIs and drug associations. **(I)** The clinical relevance of protein expressions identified in CancerProteome.

differentially expressed proteins, differential PTM sites and proteins that correlate with drugs across different cancers, by clicking on different tabs (Figure 1A). Multiple visualization results show how proteins contribute to carcinogenesis processes. Moreover, we developed six flexible tools to retrieve and analyze CancerProteome data (Figure 1A), including conventional analyses of both quantitative and PTM proteomes, functional enrichment analyses, protein–protein associations by integrating known interactions with co-expression, drug sensitivity and clinical relevance analyses.

CancerProteome also allows users to access each (micro-)protein entry for the cancer of interest via browse, quick search or query options. The 'Search' page provides three query options (Figure 2A) to search (micro-)proteins, PTM sites or cancer. CancerProteome provides a full list of differentially expressed (micro-)proteins when users search for a cancer of interest, as well as corresponding volcano figures that visualize differential expression results (Figure 2B). Enriched molecular functions from four aspects are also provided (Figure 2B). By searching for a (micro-)protein or PTM site of interest using two query options, users can obtain corresponding differential results across different cancers. We also highlight information on protein type, PTM type, site information, differential expression results and known cancer genes (Figure 2C, D).

Detailed (micro-)protein information (six types) is provided for users by clicking on relevant tabs (Figure 2E–I). CancerProteome provides basic information on (micro-)proteins in a table (Figure 2E) and identifies differentially expressed/modified (micro-)proteins (Figure 2F, G). Furthermore, to investigate associations between proteins and corresponding transcripts and PTM sites, (micro-)protein expression correlations with corresponding transcript levels from RNA-seq are explored (Figure 2F), as well as with PTM sites (Figure 2G). To compare (micro-)proteins expression/modification levels with corresponding PTM sites, modification levels at all PTM sites derived from (micro-)proteins in many cancers are also provided (Figure 2F, G). (Micro-)protein functional roles are further extrapolated by integrating known PPIs with co-expression information for each cancer, or from drug sensitivity perspectives (Figure 2H). CancerProteome allows users to compare survival differences between low and high-expression groups across diverse cancer types (Figure 2I). Multiple visualization results are provided to understand their roles in cancer. Differential proteins/sites and basic information are also provided on the 'Download' page and a detailed tutorial is provided on the 'Help' page to allow users better understand CancerProteome.

Case study

Studies have shown that GOLPH3 promotes metastasis and tumorigenicity in non-small-cell lung cancer and is highly expressed in other cancers (45). When querying the proteins encoded by *GOLPH3* in CancerProteome using the 'Search Protein' tab in all cancer types, we retrieved 35 entries from 15 cancer types (Figure 2D). In particular, detailed information for Golgi phosphoprotein 3 encoded by GOLPH3 was investigated, including basic information and functional analyses (Figure 2E–I). We found that the Golgi phosphoprotein 3 was significantly upregulated in the PDC000234 lung cancer dataset, and the corresponding *GOLPH3* gene was consistently different in the transcriptome (Figure 2F). Moreover, lower Golgi phosphoprotein 3 expression was associated with increased survival in cancer patients (Figure 2I), consistent with previous data (45). We also investigated the *APOC1* gene, which has important roles in proliferation and metastasis in several cancers (46). We identified the APOC1_odORF_MP_2 microprotein, which originated from *APOC1*, and was differentially expressed in multiple cancer types, such as the LUNG PDC000234 dataset (Supplementary Fig. S2). Different microprotein expression and potential functional data were successfully retrieved by CancerProteome (Supplementary Fig. S2), suggesting important roles in cancer.

Conclusions and future development

MS-base proteome datasets offer an effective method to decode the protein regulatory mechanism in human complex diseases. However, there is not yet a specialized, comprehensive resource for cancer proteome. By collecting and re-analyzing the publicly available MS-based raw datasets against canonical protein and our integrated benchmarked-microprotein library, CancerProteome provides a comprehensive proteome landscape to functionally decipher and visualize the (micro-)proteins across many different cancers. It is noteworthy that (micro-)proteins roles were further extrapolated by integrating known PPIs with co-expression information in different cancers, or by viewing drug sensitivity data. CancerProteome also allows users to conduct comparative analysis between tumor and control samples, as well as within different types of cancer. This module enables the decoding of the functional implications associated with differential (micro-)proteins. In the (micro-)proteins atlas in CancerProteome, the biological function of several (micro-)proteins have been validated in previous studies, such as the canonical protein Golgi phosphoprotein 3 derived from *GOLPH3* and the microproteins APOC1_odORF_MP_2 encoded by *APOC1*.

We will update and integrate more datasets into CancerProteome by (i) continuously mining datasets to expand the current (micro-)protein atlas; (ii) expanding available ribo-seq-supported-ORF regions by integrating newly generated ribo-seq datasets and (iii) incorporating information on experimentally defined (micro-)proteins. These additions will enhance CancerProteome efficiency, provide an important database to investigate (micro-)proteins in different cancers, and identify potential roles as therapeutic cancer vaccines. In summary, CancerProteome is a valuable resource thatfunctionally deciphers the cancer proteome landscape, and provides novel insights for tumor protein marker-prioritization strategies in cancer.

Data availability

CancerProteome is an open resource which functionally deciphers the proteome landscape in cancer. It is freely available via: http://bio-bigdata.hrbmu.edu.cn/CancerProteome.

Supplementary data

Supplementary Data are available at NAR Online.

Acknowledgements

Author Contributions: D.L.: formal analysis, methodology, validation, writing—original draft. D.L.: formal analysis, methodology, validation, writing—original draft. Y.C.: formal analysis, methodology, validation, writing—original draft. J.G.: formal analysis, methodology, validation. S.C.: formal analysis, methodology, validation. J.Y.: formal analysis, validation. K.L.: formal analysis, validation. T.J.: formal analysis, validation. N.D.: formal analysis, validation. X.J.: formal analysis, visualization, writing—original draft. Y.L.: conceptualization, formal analysis, visualization, writing—review and editing. J.X.: conceptualization, formal analysis, visualization, writing—review and editing.

Funding

This work was supported by the National Natural Science Foundation of China [32322020, 32170676, 31970646 and 32060152]; Natural Science Foundation of Heilongjiang Province (Key Program) [ZD2023C007] and Heilongjiang Touyan Innovation Team Program. Funding for open access charge: National Natural Science Foundation of China.

Conflict of interest statement

None declared.

References

1. Li,Y., McGrail,D.J., Xu,J., Mills,G.B., Sahni,N. and Yi,S. (2018) Gene regulatory network perturbation by genetic and epigenetic variation. *Trends Biochem. Sci*, **43**, 576–592.
2. Li,Y., Zhang,Y., Li,X., Yi,S. and Xu,J. (2019) Gain-of-function mutations: an emerging advantage for cancer biology. *Trends Biochem. Sci*, **44**, 659–674.
3. Goncalves,E., Poulos,R.C., Cai,Z., Barthorpe,S., Manda,S.S., Lucas,N., Beck,A., Bucio-Noble,D., Dausmann,M., Hall,C., *et al.* (2022) Pan-cancer proteomic map of 949 human cell lines. *Cancer Cell*, **40**, 835–849.
4. Nusinow,D.P., Szpyt,J., Ghandi,M., Rose,C.M., McDonald,E.R. 3rd, Kalocsay,M., Jane-Valbuena,J., Gelfand,E., Schweppe,D.K., Jedrychowski,M., *et al.* (2020) Quantitative proteomics of the Cancer Cell Line encyclopedia. *Cell*, **180**, 387–402.
5. Xu,K., Jin,X., Luo,Y., Zou,H., Lv,D., Wang,L., Fu,L., Cai,Y., Shao,T., Li,Y., *et al.* (2023) Spatial transcriptome analysis of long non-coding RNAs reveals tissue specificity and functional roles in cancer. *J Zhejiang Univ Sci B*, **24**, 15–31.
6. Li,Y., Zhang,Y., Pan,T., Zhou,P., Zhou,W., Gao,Y., Zheng,S. and Xu,J. (2022) Shedding light on the hidden human proteome expands immunopeptidome in cancer. *Brief Bioinform*, **23**, bbac034.
7. Lv,D., Xu,K., Jin,X., Li,J., Shi,Y., Zhang,M., Jin,X., Li,Y., Xu,J. and Li,X. (2020) LncSpA: lncRNA spatial atlas of expression across normal and cancer tissues. *Cancer Res.*, **80**, 2067–2071.
8. Sandmann,C.L., Schulz,J.F., Ruiz-Orera,J., Kirchner,M., Ziehm,M., Adami,E., Marczenke,M., Christ,A., Liebe,N., Greiner,J., *et al.* (2023) Evolutionary origins and interactomes of human, young microproteins and small peptides translated from short open reading frames. *Mol. Cell*, **83**, 994–1011.
9. Chen,J., Brunner,A.D., Cogan,J.Z., Nunez,J.K., Fields,A.P., Adamson,B., Itzhak,D.N., Li,J.Y., Mann,M., Leonetti,M.D., *et al.* (2020) Pervasive functional translation of noncanonical human open reading frames. *Science*, **367**, 1140–1146.
10. Gao,Q., Zhu,H., Dong,L., Shi,W., Chen,R., Song,Z., Huang,C., Li,J., Dong,X., Zhou,Y., *et al.* (2019) Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell*, **179**, 561–577.
11. Thul,P.J. and Lindskog,C. (2018) The Human Protein Atlas: a spatial map of the human proteome. *Protein Sci.*, **27**, 233–244.
12. Lautenbacher,L., Samaras,P., Muller,J., Grafberger,A., Shraideh,M., Rank,J., Fuchs,S.T., Schmidt,T.K., The,M., Dallago,C., *et al.* (2022) ProteomicsDB: toward a FAIR open-source resource for life-science research. *Nucleic Acids Res.*, **50**, D1541–D1552.
13. Sheils,T.K., Mathias,S.L., Kelleher,K.J., Siramshetty,V.B., Nguyen,D.T., Bologa,C.G., Jensen,L.J., Vidovic,D., Koleti,A., Schurer,S.C., *et al.* (2021) TCRD and Pharos 2021: mining the human proteome for disease biology. *Nucleic Acids Res.*, **49**, D1334–D1346.
14. Moreno,P., Fexova,S., George,N., Manning,J.R., Miao,Z., Mohammed,S., Munoz-Pomer,A., Fullgrabe,A., Bi,Y., Bush,N., *et al.* (2022) Expression Atlas update: gene and protein expression in multiple species. *Nucleic Acids Res.*, **50**, D129–D140.
15. Yu,K., Wang,Y., Zheng,Y., Liu,Z., Zhang,Q., Wang,S., Zhao,Q., Zhang,X., Li,X., Xu,R.H., *et al.* (2023) qPTM: an updated database for PTM dynamics in human, mouse, rat and yeast. *Nucleic Acids Res.*, **51**, D479–D487.
16. Huang,K.Y., Lee,T.Y., Kao,H.J., Ma,C.T., Lee,C.C., Lin,T.H., Chang,W.C. and Huang,H.D. (2019) dbPTM in 2019: exploring disease association and cross-talk of post-translational modifications. *Nucleic Acids Res.*, **47**, D298–D308.
17. Xiang,Y., Zou,Q. and Zhao,L. (2021) VPTMdb: a viral posttranslational modification database. *Brief Bioinform*, **22**, bbaa251.
18. Perez-Riverol,Y., Bai,J., Bandla,C., Garcia-Seisdedos,D., Hewapathirana,S., Kamatchinathan,S., Kundu,D.J., Prakash,A., Frericks-Zipper,A., Eisenacher,M., *et al.* (2022) The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. *Nucleic Acids Res.*, **50**, D543–D552.
19. Choi,M., Carver,J., Chiva,C., Tzouros,M., Huang,T., Tsai,T.H., Pullman,B., Bernhardt,O.M., Huttenhain,R., Teo,G.C., *et al.* (2020) MassIVE.Quant: a community resource of quantitative mass spectrometry-based proteomics datasets. *Nat. Methods*, **17**, 981–984.
20. Desiere,F., Deutsch,E.W., King,N.L., Nesvizhskii,A.I., Mallick,P., Eng,J., Chen,S., Eddes,J., Loevenich,S.N. and Aebersold,R. (2006) The PeptideAtlas project. *Nucleic Acids Res.*, **34**, D655–D658.
21. Watanabe,Y., Yoshizawa,A.C., Ishihama,Y. and Okuda,S. (2021) The jPOST repository as a public data repository for shotgun proteomics. *Methods Mol. Biol.*, **2259**, 309–322.
22. 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., *et al.* (2014) Panorama: a targeted proteomics knowledge base. *J. Proteome Res.*, **13**, 4205–4210.
23. Ma,J., Chen,T., Wu,S., Yang,C., Bai,M., Shu,K., Li,K., Zhang,G., Jin,Z., He,F., *et al.* (2019) iProX: an integrated proteome resource. *Nucleic Acids Res.*, **47**, D1211–D1217.
24. Tate,J.G., Bamford,S., Jubb,H.C., Sondka,Z., Beare,D.M., Bindal,N., Boutselakis,H., Cole,C.G., Creatore,C., Dawson,E., *et al.* (2019) COSMIC: the catalogue of somatic mutations In cancer. *Nucleic Acids Res.*, **47**, D941–D947.
25. Gao,Y., Shang,S., Guo,S., Li,X., Zhou,H., Liu,H., Sun,Y., Wang,J., Wang,P., Zhi,H., *et al.* (2021) Lnc2Cancer 3.0: an updated resource for experimentally supported lncRNA/circRNA cancer associations and web tools based on RNA-seq and scRNA-seq data. *Nucleic Acids Res.*, **49**, D1251–D1258.
26. Chakravarty,D., Gao,J., Phillips,S.M., Kundra,R., Zhang,H., Wang,J., Rudolph,J.E., Yaeger,R., Soumerai,T., Nissan,M.H., *et al.* (2017) OncoKB: a precision oncology knowledge base. *JCO Precis Oncol*, *2017*.
27. Wang,H., Yang,L., Wang,Y., Chen,L., Li,H. and Xie,Z. (2019) RPFdb v2.0: an updated database for genome-wide information of translated mRNA generated from ribosome profiling. *Nucleic Acids Res.*, **47**, D230–D234.

28. Ouspenskaia,T., Law,T., Clauser,K.R., Klaeger,S., Sarkizova,S., Aguet,F., Li,B., Christian,E., Knisbacher,B.A., Le,P.M., *et al.* (2022) Unannotated proteins expand the MHC-I-restricted immunopeptidome in cancer. *Nat. Biotechnol.*, **40**, 209–217.
29. Lv,D., Chang,Z., Cai,Y., Li,J., Wang,L., Jiang,Q., Xu,K., Ding,N., Li,X., Xu,J., *et al.* (2022) TransLnc: a comprehensive resource for translatable lncRNAs extends immunopeptidome. *Nucleic Acids Res.*, **50**, D413–D420.
30. Cai,Y., Lv,D., Li,D., Yin,J., Ma,Y., Luo,Y., Fu,L., Ding,N., Li,Y., Pan,Z., *et al.* (2023) IEAtlas: an atlas of HLA-presented immune epitopes derived from non-coding regions. *Nucleic Acids Res.*, **51**, D409–D417.
31. van Heesch,S., Witte,F., Schneider-Lunitz,V., Schulz,J.F., Adami,E., Faber,A.B., Kirchner,M., Maatz,H., Blachut,S., Sandmann,C.L., *et al.* (2019) The translational landscape of the Human heart. *Cell*, **178**, 242–260.
32. Duffy,E.E., Finander,B., Choi,G., Carter,A.C., Pritisanac,I., Alam,A., Luria,V., Karger,A., Phu,W., Sherman,M.A., *et al.* (2022) Developmental dynamics of RNA translation in the human brain. *Nat. Neurosci.*, **25**, 1353–1365.
33. UniProt,C. (2021) UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.*, **49**, D480–D489.
34. Cox,J. and Mann,M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.*, **26**, 1367–1372.
35. 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., *et al.* (2021) Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. *Cancer Cell*, **39**, 361–379.
36. Krug,K., Jaehnig,E.J., Satpathy,S., Blumenberg,L., Karpova,A., Anurag,M., Miles,G., Mertins,P., Geffen,Y., Tang,L.C., *et al.* (2020) Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell*, **183**, 1436–1456.
37. Zhang,Q., Ma,C., Chin,L.S. and Li,L. (2020) Integrative glycoproteomics reveals protein N-glycosylation aberrations and glycoproteomic network alterations in Alzheimer's disease. *Sci. Adv.*, **6**, eabc5802.
38. 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., *et al.* (2021) A proteogenomic portrait of lung squamous cell carcinoma. *Cell*, **184**, 4348–4371.
39. 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.Y., *et al.* (2019) Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell*, **179**, 964–983.
40. Jayavelu,A.K., Wolf,S., Buettner,F., Alexe,G., Haupl,B., Comoglio,F., Schneider,C., Doebele,C., Fuhrmann,D.C., Wagner,S., *et al.* (2022) The proteogenomic subtypes of acute myeloid leukemia. *Cancer Cell*, **40**, 301–317.
41. Szklarczyk,D., Kirsch,R., Koutrouli,M., Nastou,K., Mehryary,F., Hachilif,R., Gable,A.L., Fang,T., Doncheva,N.T., Pyysalo,S., *et al.* (2023) The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res.*, **51**, D638–D646.
42. Chatr-aryamontri,A., Ceol,A., Palazzi,L.M., Nardelli,G., Schneider,M.V., Castagnoli,L. and Cesareni,G. (2007) MINT: the Molecular INTeraction database. *Nucleic Acids Res.*, **35**, D572–D574.
43. Del Toro,N., Shrivastava,A., Ragueneau,E., Meldal,B., Combe,C., Barrera,E., Perfetto,L., How,K., Ratan,P., Shirodkar,G., *et al.* (2022) The IntAct database: efficient access to fine-grained molecular interaction data. *Nucleic Acids Res.*, **50**, D648–D653.
44. Wu,T., Hu,E., Xu,S., Chen,M., Guo,P., Dai,Z., Feng,T., Zhou,L., Tang,W., Zhan,L., *et al.* (2021) clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation (Camb)*, **2**, 100141.
45. Song,J.W., Zhu,J., Wu,X.X., Tu,T., Huang,J.Q., Chen,G.Z., Liang,L.Y., Zhou,C.H., Xu,X. and Gong,L.Y. (2021) GOLPH3/CKAP4 promotes metastasis and tumorigenicity by enhancing the secretion of exosomal WNT3A in non-small-cell lung cancer. *Cell Death. Dis.*, **12**, 976.
46. Ren,L., Yi,J., Yang,Y., Li,W., Zheng,X., Liu,J., Li,S., Yang,H., Zhang,Y., Ge,B., *et al.* (2022) Systematic pan-cancer analysis identifies APOC1 as an immunological biomarker which regulates macrophage polarization and promotes tumor metastasis. *Pharmacol. Res.*, **183**, 106376.

Received: August 12, 2023. **Revised:** September 7, 2023. **Editorial Decision:** September 11, 2023. **Accepted:** September 20, 2023