Pan-cancer Proteomics Analysis to Identify Tumor-Enriched and Highly Expressed Cell Surface Antigens as Potential Targets for Cancer Therapeutics

Jixin Wang,[1] Wen Yu,[2] Rachel D'Anna,[1] Anna Przybyla,[3] Matt Wilson,[4] Matthew Sung,[3] John Bullen,[5] Elaine Hurt,[5] Gina DAngelo,[6] Ben Sidders,[7] Zhongwu Lai,[8] Wenyan Zhong[9]*

[1]Oncology Data Science and [6]Late Oncology Statistics, Oncology R&D, and [2]Data Science and AI, BioPharmaceuticals R&D, AstraZeneca, Gaithersburg, MD; [3]Immune-oncology Discovery, [4]Early TDE Discovery, [5]Early TTD Discovery, and [7]Oncology Data Science, Oncology R&D, AstraZeneca, Cambridge, UK;[8]Oncology Data Science, Oncology R&D, AstraZeneca, Waltham, MA; [9]Oncology Data Science, Oncology R&D, AstraZeneca, New York, NY

Running title: Tumor-enriched, highly expressed cell surface antigen identification

***Corresponding author:**

Wenyan Zhong

AstraZeneca, One MedImmune Way, Gaithersburg, MD 20878, USA

Phone: (301) 398-4209

Fax: (301) 398-9000

E-mail: wenyan.zhong@astrazeneca.com

ABSTRACT

The National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC) provides unique opportunities for cancer target discovery using protein expression. Proteomics data from CPTAC tumor types have been primarily generated using a multiplex tandem mass tag (TMT) approach, which is designed to provide protein quantification relative to reference samples. However, relative protein expression data is suboptimal for prioritization of targets within a tissue type, which requires additional reprocessing of the original proteomics data to derive absolute quantitation estimation. We evaluated the feasibility of using differential protein analysis coupled with intensity-based absolute quantification (iBAQ) to identify tumor-enriched and highly expressed cell surface antigens, employing tandem mass tag (TMT) proteomics data from CPTAC. Absolute quantification derived from TMT proteomics data was highly correlated with that of label-free proteomics data from the CPTAC colon adenocarcinoma cohort, which contains proteomics data measured by both approaches. We validated the TMT-iBAQ approach by comparing the iBAQ value to the receptor density value of HER2 and TROP2 measured by flow cytometry in about 30 selected breast and lung cancer cell lines from the Cancer Cell Line Encyclopedia. Collections of these tumor-enriched and highly expressed cell surface antigens could serve as a valuable resource for the development of cancer therapeutics, including antibody-drug conjugates and immunotherapeutic agents.

INTRODUCTION

Biologics-based cancer therapeutics including antibody drug conjugates and immunotherapies are proven modalities for the treatment of cancer. Most of these therapeutics often target proteins preferentially expressed on the surface of cancer cells (1). For many years, drug candidate targets for these modalities have been discovered through mining of large RNA expression databases such as The Cancer Genome Atlas (TCGA) as a surrogate approach in the absence of a large protein expression database (2). Recent advancements in proteogenomic characterization of large collections of cancer patient samples represented by The National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium (CPTAC) (3) has provided unique opportunities for cancer target discovery using protein expression.

The CPTAC serves as a repository of global proteomics data of tumor and normal adjacent tissue (NAT) samples from more than 10 tumor types, including colon adenocarcinoma (COAD) (4), breast cancer (BRCA) (5), lung adenocarcinoma (LUAD) (6), ovarian cancer (OV) (7), glioblastoma multiforme (GBM) (8), clear-cell renal cell carcinoma (ccRCC) (9), pancreatic ductal adenocarcinoma (PDA) (10), uterine corpus endometrial carcinoma (UCEC) (11), head and neck squamous-cell carcinoma (HNSCC) (12), and lung squamous-cell carcinoma (LSCC) (13). However, proteomics data from these tumor types have been primarily generated by the multiplex tandem mass tag (TMT) approach, which was designed to provide relative protein quantification to reference samples. In addition to preferential expression in tumor compared to normal tissues, ideal targets for biological therapeutics often requires high level expression of these targets on the tumor cell surface. The relative protein expression data from CPTAC hinders the direct utilization of the proteomics data for target prioritization and additional reprocessing of the original proteomics data to derive absolute quantitation estimation is required.

Mass spectrometry–based proteomics methods have been widely used for relative and absolute quantification of proteomes and posttranslational modifications (14). Both label-free and label-based quantification methods have been applied in proteomics studies. Several algorithms for label-free absolute protein quantification have been developed, including TOP3 (15), label-free absolute quantification (LFAQ) (16), absolute protein expression (APEX) (17), total protein approach (TPA) (18-21), and intensity-based absolute quantification (iBAQ) (22). The TPA approach has been applied to the data-independent acquisition (DIA)–based label-free quantification (LFQ) method, and the tandem mass spectrometry–based DIA-TPA approach is comparable to the mass spectrometry–based data-dependent acquisition (DDA)–TPA method (23).

In addition to these methods, FragPipe (24) with the MSFragger search engine can provide equivalent protein intensity data based on the intensities of the precursors and TMT reporters, which can be used as input to derive estimated absolute quantitation by using previously published absolute quantification algorithms. This approach has previously been applied to CPTAC ccRCC data processing (9). Although the intensity calculated by FragPipe reflects the absolute intensity of the protein summarized from intensities from all ionized peptides. However, this value cannot be used to directly compare the protein abundance across different proteins, since different proteins have different number of ionizable peptides. In this study, we developed a computational approach to estimate absolute protein abundance from TMT proteomics data. We reprocessed TMT proteomics data with FragPipe (25) and calculated the TMT-TPA and TMT-iBAQ values for 10 tumor types. Two methods for calculating estimated absolute protein abundance were compared using label-free global proteomics data from the CPTAC COAD data set, which contains both label-free and TMT proteomics data. We also compared the LFQ-TPA

(label-free) and the TMT-TPA (TMT) value of the CPTAC COAD data set. The iBAQ approach was evaluated by using receptor density values (26) measured by flow cytometry and assessing the potential application of the iBAQ data for comparing protein expression levels by indication to identify tumor-enriched and highly expressed cell surface antigens. Finally, by combining differential analysis of tumor and NAT samples with surface protein prediction, we were able to identify pan-cancer tumor-enriched and highly expressed proteins as potential cancer antigens for target discovery (27, 28), cancer vaccine development (29), and cancer biology studies (30).

EXPERIMENTAL SECTION

Re-analysis of CPTAC and CCLE data sets

Raw data files from the CPTAC were downloaded from https://proteomic.datacommons.cancer.gov/pdc/ with PDC study identifier COAD (PDC000116), BRCA (PDC000120), ccRCC (PDC000127), PDA (PDC000393), OV (PDC000118), LSCC (PDC000234), GBM, LUAD (PDC000153), HNSCC (PDC000221), and UCEC (PDC000125), and data files from the Cancer Cell Line Encyclopedia (CCLE) were downloaded from https://massive.ucsd.edu/ProteoSAFe/dataset.jsp?task=02cd1b6a7c674f3ebdbed300b5d9aa57 (MassIVE repository accession MSV000085836). The CPTAC COAD data set contains both label-free and TMT proteomics data. LFQ was carried out on the unlabeled data (100 tumor samples) of COAD from CPTAC, using MaxQuant (31). TMT 10-plex mzML files from CPTAC COAD, BRCA, ccRCC, PDA, OV, LSCC, GBM, LUAD, HNSCC, and UCEC and CCLE were analyzed via the FragPipe computational platform (version 15) with MSFragger (version 3.2) and Philosopher (version 3.4.13) (25) for protein identification and quantitation,

using the recommended search parameters with global normalization. Protein grouping was performed.

The label-free data were searched in MaxQuant against a collection of 73045 protein sequences downloaded from Uniprot on June 22, 2018, with full tryptic specificity and up to two miss-cleavage sites, fixed modification of “Carbamidomethyl (C)”, and variable modifications of “Oxidation (M), Acetyl (Protein N-term) and Deamidation (NQ)”. The mass tolerance was set to 20 ppm for both the precursors and the fragments. The protein identifications were filtered at the 1% false discovery rate (FDR) level as determined by MaxQuant.

The TMT-labelled data, they were searched in FragPipe (version 15.0 with MsFragger version 3.2 and Philosopher version 3.4.13) against protein sequences (reviewed, 20420 entries) downloaded from Uniprot by FragPipe on July 15, 2021, with full tryptic specificity and up to two miss cleavages, fixed modification of Cys residue by 57.02146 daltons (Da), and Lys residue by 229.16293 Da, variable modifications of Met residue by 15.9949 Da, protein N-terminus by 42.0106 Da, and peptide N-terminus or Ser residue by 229.16293 Da. The mass tolerance was set to 20 ppm for both the precursors and the fragments. The protein identifications were filtered to 1% FDR as determined by FragPipe from the embedded decoy sequences.

LFQ-TPA and LFQ-iBAQ values of the CPTAC COAD cohort were calculated from label-free raw intensity using equations 1 and 2 (see below) respectively. TMT-TPA, TMT-iBAQ, and TMT-iBAQ-derived copy number values of all 10 CPTAC indications and CCLE were calculated from protein abundance output of FragPipe using equations 1, 2, and 3 (see below), respectively.

Equation 1 (TPA method (19)):

$$P_i = \frac{Intensity_i}{\sum_i (Intensity_i) \cdot MW_i}$$

Equation 2 (iBAQ method (22)):

$$P_i = \frac{\sum_k Intensity_k}{No.of\ theoretical\ peptides}$$

Equation 3 (iBAQ-derived copy number estimation (32, 33)):

$$riBAQ = \frac{iBAQ}{\sum_i iBAQ}$$

$$iBAQ - derived\ copy\ number\ = riBAQ * 2 * 10\text{^}9$$

where *Pi* = predicted absolute abundance of protein *i*, and *k* = peptide and MW = molecular weight. TPA units are expressed as femtomoles per gram. "No. of theoretical peptides" refers to the number of theoretical peptides when that protein is typically digested by trypsin. iBAQ values were calculated based on UniProt protein accessions and reported using associated gene symbols. For iBAQ-derived copy number estimation, 2×10^9 as the total number of proteins in a cell was used based on published studies (33).)

Differential protein analysis

FragPipe protein abundance data from CPTAC tumor and NAT samples were used for differential protein analysis. Principal component analysis and unsupervised hierarchical clustering were used to evaluate batch effect and sample quality control. An initial filtering step was applied to the protein abundance data, and proteins that were quantified in at least 50% of the samples were used for further analysis. Missing value visualization was performed, and k–

nearest neighbor (KNN) imputation was applied to the data with the R package MSnbase (version 2.20.4) (34). Previous studies have shown that KNN imputation performs better in TMT proteomics data (35, 36). Differential protein expression analysis was performed with R (version 4.1.1), using Empirical Bayes statistics on protein-wise linear models with limma (37) embedded in the differential protein analysis (DEP) package (version 1.16.0) (38). Proteins that were significantly upregulated in tumor compared with NAT were identified with a threshold of adjusted $P = 0.01$ and fold change = 1.5. Cell surface proteins were predicted using surface prediction consensus (SPC) score downloaded from SurfaceGenie (39-41), and proteins with SPC ≥ 1 were annotated as potential surface proteins. Differentially upregulated cell surface proteins were ranked by their iBAQ values in tumor tissues to obtain highly expressed, tumor-enriched cell surface antigens.

Flow experiment and analysis methodology

Datopotamab and trastuzumab deruxtecan (T-DXd) (human immunoglobulin G1 [IgG1]) were labeled with Alexa Fluor 647, using the Alexa Fluor protein labeling kit (A20173; Invitrogen). Cells from CCLE cell lines were harvested with Accutase (A1110501; Gibco) and stained with Live/Dead Fixable Violet Dead Cell Stain (L34955; Invitrogen) for 20 minutes on ice and protected from light. Cells were then washed and stained with 20 μg of datopotamab−Alexa Fluor 647 or T-DXd−Alexa Fluor 647 per mL for 1 hour on ice and protected from light. Cells were fixed with 2% paraformaldehyde (AAJ19943K2; Thermo Fisher Scientific) for 20 minutes at room temperature and stored in phosphate-buffered saline (PBS) at 4°C overnight.

Quantum Simply Cellular anti-human IgG beads (816; Bangs Laboratories) were stained with 20 μg of datopotamab−Alexa Fluor 647 or T-DXd−Alexa Fluor 647 in PBS for 30 minutes at room temperature. Beads were then washed three times in PBS and transferred to flow tubes.

Beads and cells were run on an LSRFortessa flow cytometer (BD Biosciences), using the same voltage settings. Flow data analysis was performed on an FCS Express flow cytometer (De Novo Software) and bead median fluorescence intensity values were used to plot the ABC standard curves for datopotamab and T-DXd. The lot-specific template provided with the bead kit was used to obtain cell ABC values by extrapolating from the standard curves.

Correlation analysis

Correlations of FragPipe protein abundance with global normalization and LFQ intensity, iBAQ vs. TPA, LFQ-TPA vs. TMT-TPA, and LFQ-iBAQ vs. TMT-iBAQ, as well as receptor density vs. iBAQ, and protein abundance by FragPipe vs derived iBAQ data were determined with Pearson correlation.

RESULTS

Differentially up-regulated proteins in tumor compared to NAT

We first identified proteins that were upregulated in tumor tissues compared with NAT. The workflow used in this study is outlined in Figure 1. We re-analyzed the TMT 10-plex mzML files of 10 indications (COAD, BRCA, LUAD, OV, GBM, ccRCC, PDA, UCEC, HNSCC, and LSCC) from CPTAC through FragPipe and MSFragger to obtain protein abundance values for identified proteins expressed in tumor and NAT. The protein summary and quantification from FragPipe and MSFragger output for CPTAC 10 indications are shown in Supplemental Table S1.

Missing values are commonly observed in proteomics data due to a variety of reasons, including protein digestion loss, measurement failure, protein abundances below the instrument limit of detection, poor ionization efficiency, and signal-to-noise ratio (42-44). It is critical to apply the correct missing value imputation method before performing differential abundance analysis. Our evaluation of the missing values showed a missing-at-random pattern across tumor and normal conditions but a missing-not-at-random pattern at the multiplex level. Therefore, we chose a KNN algorithm to impute missing values (45) and identified a range of differentially upregulated proteins across various tumor types (Table 1). As an example, Supplemental Figure S1 shows a missing data pattern for the CPTAC ccRCC indication.

Estimation of absolute protein abundance

In addition to differential protein analysis, we estimated absolute protein abundance using TPA and iBAQ approaches. Both methods have been previously applied and validated in estimating absolute protein abundance with label-free proteomics data (23, 46). The CPTAC COAD data set contains both label-free and TMT global proteomics data from the same tumor samples and thus is an ideal data set with which to compare absolute protein abundance calculated from both platforms. The data source and analysis workflow are shown in Supplemental Figure S2. The correlation between FragPipe protein abundance and label-free quantification intensity is 0.67 as shown in Supplemental Figure S3.

We then compared iBAQ and TPA methods for calculating estimated absolute protein abundance, using label-free and TMT global proteomics data from the COAD data set. For label-free proteomics data, LFQ-TPA was highly correlated with LFQ-iBAQ (r = 0.95), as shown in Figure 2A. A similar trend was observed for TMT-TPA and TMT-iBAQ correlation (r = 0.98) in

the TMT proteomics data, suggesting that the TPA and iBAQ methods are highly comparable. Next, we compared LFQ-TPA (label-free) with TMT-TPA (TMT), and LFQ-iBAQ with TMT-iBAQ, using overlap samples and proteins from the COAD data set. The LFQ-iBAQ and TMT-iBAQ methods of the overlap samples and proteins for the COAD data set were well correlated (r = 0.71). TMT-iBAQ greatly underestimated protein abundance compared with LFQ-iBAQ (Figure 2B), although a higher correlation was observed between LFQ-TPA and TMT-TPA (r = 0.76). Further, we investigated the correlation between protein abundance reported by FragPipe and derived TMT-iBAQ data and the correlation ranges from 0.85 to 0.88 (R > 0.5) across 10 indications (Supplemental Figure S4). These results suggest that both TPA and iBAQ can be applied to TMT proteomics data to estimate absolute protein abundance. We chose to use iBAQ to rank protein expression for identifying highly expressed cell surface antigens because this method is used in most of the protein copy number estimation methods reported in the literature (47-49).

Comparison of TMT-iBAQ absolute protein abundance estimation to receptor density values measured by flow cytometry

To evaluate iBAQ-based protein abundance estimation for TMT proteomics data, we compared experimentally generated receptor density values of HER2 and TROP2 of CCLE cell lines with iBAQ values derived from CCLE TMT proteomics data processed through FragPipe. iBAQ quantity and receptor density for HER2 and TROP2 were highly correlated, with Pearson r = 0.82 and r = 0.83, respectively (Figure 3).

Identifying highly expressed, tumor-enriched cell surface antigens

To identify highly expressed, tumor-enriched cell surface antigens, we ranked the upregulated and predicted cell surface proteins by their iBAQ values (Supplemental Table S2). The top five ranked antigens based on differential expression analysis and iBAQ ranking are listed in Table 1. For COAD, we identified 236 upregulated proteins at the threshold of FDR-adjusted $P < 0.01$ and fold change > 1.5. Of these 236 proteins, 46 were predicted to be expressed on the cell surface, having a SPC score of ≥1. We also evaluated well-known tumor-enriched and highly expressed cell surface antigen HER2 in BRCA. ERBB2 was highly expressed in the HER2 subtype, and iBAQ values were higher in tumor samples than in normal samples (Supplemental Figure S5), confirming that it is a highly expressed cell surface antigen in breast cancer.

DISCUSSION

We built a comprehensive dataset containing differentially expressed tumor cell surface antigens ranked by their absolute protein expression across 10 tumor types via a newly developed computational approach (Figure 1), which represents the first surfaceome dataset containing absolute protein expression quantification derived from mass spectrometry–based proteomics data. With the addition of proteomics and phosphoproteomics characterization, CPTAC has provided the cancer community with an unprecedented view of cancer disease biology at the protein level. By combining genetic and transcriptomics data, cancer researchers have unveiled many novel mechanisms of cancer, discovered novel subtypes, and provided richer disease context for current treatments (50, 51). However, the process and parameter settings are generally not the same across tumor types in the CPTAC dataset (51) and the protein expression values are relative measurements, thus presenting challenges for studies that require cross-

indication and cross-protein comparisons. Reprocessing of TMT global proteomics data with the same pipeline is necessary to avoid potential confounding factors. In addition, developing a method for estimating the absolute protein quantitation would allow cross-protein comparisons. The ability to compare and rank protein expression levels is necessary for prioritizing cancer surface antigens for novel biologics-based target identification.

To address the aforementioned challenges, we reprocessed CPTAC data sets across 10 indications with a common pipeline, FragPipe. We also incorporated missing value evaluation, implemented missing value imputation, and used a model-based method (DEP) for differential analysis, taking advantage of these recently developed proteomics data analysis methods. Through this analysis, we identified a significant number (28–417) of tumor-overexpressed proteins across 10 indications (Table 1), providing a comprehensive view of proteins that could be related to tumor pathogenesis obtained through a consistent methodology.

Although several methods have been developed and validated for absolute protein quantification from label-free proteomics data (19, 22, 52-54), there is a lack of evidence that these methods can be used for absolute quantification of TMT proteomics data. Here, we used CPTAC COAD data containing both label-free and TMT proteomics data to show that the TMT-TPA and TMT-iBAQ methods are well correlated with label-free–based LFQ-TPA and LFQ-iBAQ approaches. This finding provides a strong rationale for ranking protein expression levels using iBAQ quantification, which will ultimately aid in the identification of tumor-enriched and highly expressed cell surface antigens. To further evaluate our approach of using TMT-iBAQ for estimating absolute protein quantification, we reprocessed CCLE TMT proteomics data using the same pipeline and calculated iBAQ values for all identified proteins. Our analysis revealed a

high correlation between iBAQ values and receptor density of HER2 and TROP2 in about 30 selected breast and lung cancer cell lines (26, 55), providing additional confirmation for using the iBAQ absolute quantitation method for TMT proteomics data.

A limitation of our study is that we tested our approach using only HER2 and TROP2 receptor density data in a limited number of cell lines. Additional orthogonal measurements of protein absolute quantification for a larger number of proteins will help to further validate the iBAQ methodology for the estimation of absolute protein abundance using TMT proteomics data.

In summary, we have established a novel computational approach for identifying tumor-selective and highly expressed surface antigens from proteomics data, thereby providing a unique data resource for integrating potential cancer disease vulnerabilities employing biologics-based cancer therapeutics. Our results demonstrate that differential protein expression analysis combined with iBAQ ranking can aid in identifying tumor-enriched and highly expressed cell surface antigens. We evaluated the TMT-iBAQ and TMT-TPA approaches with label-free data from the CPTAC COAD data set and found that TPA and iBAQ could be applied to TMT proteomics data to derive estimated absolute protein abundance. We further validated the iBAQ protein abundance approach with receptor density data from the CCLE cell line. Collectively, the data shared in this report provide a method for the identification of tumor-enriched and highly expressed cell surface antigens with iBAQ and DEP analysis. FragPipe-derived DEP, TMT-TPA, iBAQ, and iBAQ-derived copy numbers are useful data sources for cancer biology research and the identification of potential targets. The methodology and the data sources in this study can serve as valuable resources for cancer research.

Data Availability: The FragPipe output of protein abundance estimation, derived TMT-TPA, iBAQ, iBAQ-derived copy number and differential protein expression data for CPTAC ten indications can be downloaded from https://doi.org/10.5281/zenodo.7991979.

Supplemental data: This article contains supplemental data.

Author Contributions

J. W. and W. Z. conception and design; A. P., M. W., and M. S. flow experiment; J. W., Y. W., R. D., J. B., E. H., G. D., B. S., Z. L., and W. Z. analysis and interpretation of data. All authors contributed to the writing, review, and/or revision of the manuscript, have approved the final version of the manuscript, and agree to be accountable for all aspects of the work.

Acknowledgments: We gratefully acknowledge the CPTAC for providing open-source proteomics data and Deborah Shuman of AstraZeneca for editing the manuscript and formatting the figures.

Conflict of Interest: All authors are employees of AstraZeneca and may have stock ownership, options, or interests in the company. This study was funded by AstraZeneca.

Abbreviations: The abbreviations used are: APEX, absolute protein expression; BRCA, breast cancer; CCLE, Cancer Cell Line Encyclopedia; ccRCC, clear-cell renal cell carcinoma; COAD, colon adenocarcinoma; CPTAC, Clinical Proteomic Tumor Analysis Consortium; DEP, differential protein analysis; DIA, data-independent acquisition; FDR, false discovery rate; GBM, glioblastoma multiforme; HNSCC, head and neck squamous-cell carcinoma; iBAQ, intensity-based absolute quantification; IgG, immunoglobulin G; KNN, k–nearest neighbor; LFQ, label-free quantification; LSCC, lung squamous-cell carcinoma; LUAD, lung

adenocarcinoma; NAT, normal adjacent tissue; OV, ovarian cancer; PBS, phosphate-buffered saline; PDA, pancreatic ductal adenocarcinoma; SPC, surface prediction consensus; T-DXd, trastuzumab deruxtecan; TMT, tandem mass tag; TPA, total protein approach; UCEC, uterine corpus endometrial carcinoma.

REFERENCES

1. Orentas, R., Yang, J., Wen, X., Wei, J., Mackall, C., and Khan, J. (2012) Identification of cell surface proteins as potential immunotherapy targets in 12 pediatric cancers. *Frontiers in Oncology* 2, 194

2. Jiang, J., Yuan, J., Hu, Z., Zhang, Y., Zhang, T., Xu, M., Long, M., Fan, Y., Tanyi, J. L., Montone, K. T., Tavana, O., Vonderheide, R. H., Chan, H. M., Hu, X., and Zhang, L. (2022) Systematic illumination of druggable genes in cancer genomes. *Cell Rep* 38, 110400

3. Lindgren, C. M., Adams, D. W., Kimball, B., Boekweg, H., Tayler, S., Pugh, S. L., and Payne, S. H. (2021) Simplified and unified access to cancer proteogenomic data. *J Proteome Res* 20, 1902-1910

4. 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. T., 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., and Zhang, B. (2019) Proteogenomic analysis of human colon cancer reveals new therapeutic opportunities. *Cell* 177, 1035-1049

5. 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., and Gillette, M. A. (2020) Proteogenomic landscape of breast cancer tumorigenesis and targeted therapy. *Cell* 183, 1436-1456

6. Gillette, M. A., Satpathy, S., Cao, S., Dhanasekaran, S. M., Vasaikar, S. V., Krug, K., Petralia, F., Li, Y., Liang, W. W., Reva, B., Krek, A., Ji, J., Song, X., Liu, W., Hong, R., Yao, L., Blumenberg, L., Savage, S. R., Wendl, M. C., Wen, B., Li, K., Tang, L. C., MacMullan, M. A., Avanessian, S. C., Kane, M. H., Newton, C. J., Cornwell, M., Kothadia, R. B., Ma, W., Yoo, S., Mannan, R., Vats, P., Kumar-Sinha, C., Kawaler, E. A., Omelchenko, T., Colaprico, A., Geffen, Y., Maruvka, Y. E., da Veiga Leprevost, F., Wiznerowicz, M., Gümüş, Z. H., Veluswamy, R. R., Hostetter, G., Heiman, D. I., Wyczalkowski, M. A., Hiltke, T., Mesri, M., Kinsinger, C. R., Boja, E. S., Omenn, G. S., Chinnaiyan, A. M., Rodriguez, H., Li, Q. K., Jewell, S. D., Thiagarajan, M., Getz, G., Zhang, B., Fenyö, D., Ruggles, K. V., Cieslik, M. P., Robles, A. I., Clauser, K. R., Govindan, R., Wang, P., Nesvizhskii, A. I., Ding, L., Mani, D. R., and Carr, S. A. (2020) Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell* 182, 200-225

7. 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. F., Payne, S. H., Huang, C., Wang, L. B., 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., and Rodland, K. D. (2020) Proteogenomic characterization of ovarian HGSC implicates mitotic kinases, replication stress in observed chromosomal instability. *Cell Rep Med* 1, 100004

8. 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.-W., Song, X., Dou, Y., Wendl, M. C., Caravan, W., Liu, W., Cui Zhou, D., Ji, J., Tsai, C.-F., 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, 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., and Zhu, J. (2021) Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell* 39, 509-528

9. 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., 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. C., 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., and Zhang, H. (2019) Integrated proteogenomic characterization of clear cell renal cell carcinoma. *Cell* 179, 964-983

10. 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., and Zhang, H. (2021) Proteogenomic characterization of pancreatic ductal adenocarcinoma. *Cell* 184, 5031-5052

11. 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.-F., 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.-Y., Reva, B., Liao, Y., Rykunov, D., Colaprico, A., Chen, X. S., Czekański, A., Jędryka, 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., Fenyö, D., Agarwal, A., Anurag, M., Avtonomov, D., Birger, C., Birrer, M. J., Boca, S. M., Bocik, W. E., Borate, U., Borucki, M., Burke, M. C., Cai, S., Calinawan, A., Carr, S. A., Carter, S., Castro, P., Cerda, S., Chaikin, M., Chan, D. W., Chan, D., Charamut, A., Chen, F., Chen, J., Chen, L., Chen, L. S., Chesla, D.,

Chheda, M. G., Chinnaiyan, A. M., Chowdhury, S., Cieslik, M. P., Clark, D. J., Cottingham, S., Culpepper, H., Day, J., De Young, S., Demir, E., Dhanasekaran, S. M., Dhir, R., Domagalski, M. J., Dottino, P., Druker, B., Duffy, E., Dyer, M., Edwards, N. J., Edwards, R., Elburn, K., Field, J. B., Francis, A., Gabriel, S., Geffen, Y., Geiszler, D., Gillette, M. A., Godwin, A. K., Grady, P., Hannick, L., Hariharan, P., Hilsenbeck, S., Hindenach, B., Hoadley, K. A., Hong, R., Hostetter, G., Hsieh, J. J., Hu, Y., Ittmann, M. M., Jaehnig, E., Jewell, S. D., Ji, J., Jones, C. D., Karabon, R., Ketchum, K. A., Khan, M., Kim, B.-J., Krek, A., Krubit, T., Kumar-Sinha, C., Leprevost, F. D., Lewis, M., Li, Q. K., Li, Y., Liu, H., Lubinski, J., Ma, W., Madan, R., Malc, E., Malovannaya, A., Mareedu, S., Markey, S. P., Marrero-Oliveras, A., Martignetti, J., McDermott, J., McGarvey, P. B., McGee, J., Mieczkowski, P., Modugno, F., Montgomery, R., Newton, C. J., Omenn, G. S., Paulovich, A. G., Perou, A. M., Petralia, F., Piehowski, P., Polonskaya, L., Qi, L., Richey, S., Robinson, K., Roche, N., Rohrer, D. C., Schadt, E. E., Schnaubelt, M., Shi, Y., Skelly, T., Sokoll, L. J., Song, X., Stein, S. E., Suh, J., Tan, D., Tansil, D., Teo, G. C., Thangudu, R. R., Tognon, C., Traer, E., Tyner, J., Um, K. S., Valley, D. R., Vatanian, N., Vats, P., Velvulou, U., Vernon, M., Wang, L.-B., Wang, Y., Webster, A., Westbrook, T., Wheeler, D., Whiteaker, J. R., Wilson, G. D., Zakhartsev, Y., Zelt, R., Zhang, H., Zhang, Y., Zhang, Z., and Zhao, G. (2020) Proteogenomic characterization of endometrial carcinoma. *Cell* 180, 729-748

12. 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. Y., Wyczalkowski, M. A., Li, K., Colaprico, A., Li, Q. K., Clark, D. J., Hu, Y., Cao, L., Pan, J., Wang, Y., Cho, K. C., Shi, Z., Liao, Y., Jiang, W., Anurag, M., Ji, J., Yoo, S., Zhou, D. C., Liang, W. 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., and Zhang, B. (2021) Proteogenomic insights into the biology and treatment of HPV-negative head and neck squamous cell carcinoma. *Cancer Cell* 39, 361-379

13. 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.-K., 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., and Hanhan, Z. (2021) A proteogenomic portrait of lung squamous cell carcinoma. *Cell* 184, 4348-4371

14. Ankney, J. A., Muneer, A., and Chen, X. (2018) Relative and absolute quantitation in mass spectrometry-based proteomics. *Annu Rev Anal Chem* 11, 49-77

15. Wilhelm, M., Schlegl, J., Hahne, H., Gholami, A. M., Lieberenz, M., Savitski, M. M., Ziegler, E., Butzmann, L., Gessulat, S., Marx, H., Mathieson, T., Lemeer, S., Schnatbaum, K., Reimer, U., Wenschuh, H., Mollenhauer, M., Slotta-Huspenina, J., Boese, J. H., Bantscheff, M., Gerstmair, A., Faerber, F., and Kuster, B. (2014) Mass-spectrometry-based draft of the human proteome. *Nature* 509, 582-587

16. Chang, C., Gao, Z., Ying, W., Fu, Y., Zhao, Y., Wu, S., Li, M., Wang, G., Qian, X., Zhu, Y., and He, F. (2019) LFAQ: Toward Unbiased Label-Free Absolute Protein Quantification by Predicting Peptide Quantitative Factors. *Anal Chem* 91, 1335-1343

17. Lu, P., Vogel, C., Wang, R., Yao, X., and Marcotte, E. M. (2007) Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* 25, 117-124

18. Wiśniewski, J. R., Ostasiewicz, P., Duś, K., Zielińska, D. F., Gnad, F., and Mann, M. (2012) Extensive quantitative remodeling of the proteome between normal colon tissue and adenocarcinoma. *Mol Syst Biol* 8, 611

19. Wiśniewski, J. R. (2017) Label-free and standard-free absolute quantitative proteomics using the "total protein" and "proteomic ruler" approaches. *Methods Enzymol* 585, 49-60

20. Wiśniewski, J. R., and Rakus, D. (2014) Multi-enzyme digestion FASP and the 'Total Protein Approach'-based absolute quantification of the *Escherichia coli* proteome. *J Proteomics* 109, 322-331

21. Sánchez, B. J., Lahtvee, P. J., Campbell, K., Kasvandik, S., Yu, R., Domenzain, I., Zelezniak, A., and Nielsen, J. (2021) Benchmarking accuracy and precision of intensity-based absolute quantification of protein abundances in Saccharomyces cerevisiae. *Proteomics* 21, e2000093

22. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature* 473, 337-342

23. He, B., Shi, J., Wang, X., Jiang, H., and Zhu, H. J. (2019) Label-free absolute protein quantification with data-independent acquisition. *J Proteomics* 200, 51-59

24. Yu, F., Teo, G. C., Kong, A. T., Li, G. X., Demichev, V., and Nesvizhskii, A. I. (2022) One-stop analysis of DIA proteomics data using MSFragger-DIA and FragPipe computational platform. *bioRxiv*, 2022.2010.2028.514272

25. Kong, A. T., Leprevost, F. V., Avtonomov, D. M., Mellacheruvu, D., and Nesvizhskii, A. I. (2017) MSFragger: ultrafast and comprehensive peptide identification in mass spectrometry–based proteomics. *Nature Methods* 14, 513-520

26. Nusinow, D. P., Szpyt, J., Ghandi, M., Rose, C. M., McDonald, E. R., III, Kalocsay, M., Jane-Valbuena, J., Gelfand, E., Schweppe, D. K., Jedrychowski, M., Golji, J., Porter, D. A., Rejtar, T., Wang, Y. K., Kryukov, G. V., Stegmeier, F., Erickson, B. K., Garraway, L. A.,

Sellers, W. R., and Gygi, S. P. (2020) Quantitative proteomics of the cancer cell line encyclopedia. *Cell* 180, 387-402

27. Lee, J. K., Bangayan, N. J., Chai, T., Smith, B. A., Pariva, T. E., Yun, S., Vashisht, A., Zhang, Q., Park, J. W., Corey, E., Huang, J., Graeber, T. G., Wohlschlegel, J., and Witte, O. N. (2018) Systemic surfaceome profiling identifies target antigens for immune-based therapy in subtypes of advanced prostate cancer. *Proc Natl Acad Sci U S A* 115, E4473-e4482

28. Feola, S., Chiaro, J., Martins, B., and Cerullo, V. (2020) Uncovering the tumor antigen landscape: what to know about the discovery process. *Cancers (Basel)* 12, 1660

29. Buonaguro, L., and Tagliamonte, M. (2020) Selecting target antigens for cancer vaccine development. *Vaccines (Basel)* 8, 615

30. Karhemo, P. R., Ravela, S., Laakso, M., Ritamo, I., Tatti, O., Mäkinen, S., Goodison, S., Stenman, U. H., Hölttä, E., Hautaniemi, S., Valmu, L., Lehti, K., and Laakkonen, P. (2012) An optimized isolation of biotinylated cell surface proteins reveals novel players in cancer metastasis. *J Proteomics* 77, 87-100

31. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnol* 26, 1367-1372

32. Shin, J. B., Krey, J. F., Hassan, A., Metlagel, Z., Tauscher, A. N., Pagana, J. M., Sherman, N. E., Jeffery, E. D., Spinelli, K. J., Zhao, H., Wilmarth, P. A., Choi, D., David, L. L., Auer, M., and Barr-Gillespie, P. G. (2013) Molecular architecture of the chick vestibular hair bundle. *Nat Neurosci* 16, 365-374

33. Milo, R. (2013) What is the total number of protein molecules per cell volume? A call to rethink some published values. *Bioessays* 35, 1050-1055

34. Gatto, L., and Lilley, K. S. (2012) MSnbase: an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation. *Bioinformatics* 28, 288-289

35. Gao, Q., Zhu, H., Dong, L., Shi, W., Chen, R., Song, Z., Huang, C., Li, J., Dong, X., Zhou, Y., Liu, Q., Ma, L., Wang, X., Zhou, J., Liu, Y., Boja, E., Robles, A. I., Ma, W., Wang, P., Li, Y., Ding, L., Wen, B., Zhang, B., Rodriguez, H., Gao, D., Zhou, H., and Fan, J. (2019) Integrated proteogenomic characterization of HBV-related hepatocellular carcinoma. *Cell* 179, 561-577

36. Palstrøm, N. B., Matthiesen, R., and Beck, H. C. (2020) Data Imputation in Merged Isobaric Labeling-Based Relative Quantification Datasets. *Methods Mol Biol* 2051, 297-308

37. Smyth, G. K. (2004) Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology* 3

38. Zhang, X., Smits, A. H., van Tilburg, G. B. A., Ovaa, H., Huber, W., and Vermeulen, M. (2018) Proteome-wide identification of ubiquitin interactions using UbIA-MS. *Nat Protoc* 13, 530-550

39. Waas, M., Snarrenberg, S. T., Littrell, J., Jones Lipinski, R. A., Hansen, P. A., Corbett, J. A., and Gundry, R. L. (2020) SurfaceGenie: a web-based application for prioritizing cell-type-specific marker candidates. *Bioinformatics* 36, 3447-3456

40. Díaz-Ramos, M. C., Engel, P., and Bastos, R. (2011) Towards a comprehensive human cell-surface immunome database. *Immunol Lett* 134, 183-187

41. da Cunha, J. P., Galante, P. A., de Souza, J. E., de Souza, R. F., Carvalho, P. M., Ohara, D. T., Moura, R. P., Oba-Shinja, S. M., Marie, S. K., Silva, W. A., Jr., Perez, R. O., Stransky, B.,

Pieprzyk, M., Moore, J., Caballero, O., Gama-Rodrigues, J., Habr-Gama, A., Kuo, W. P., Simpson, A. J., Camargo, A. A., Old, L. J., and de Souza, S. J. (2009) Bioinformatics construction of the human cell surfaceome. *Proc Natl Acad Sci U S A* 106, 16752-16757

42. McGurk, K. A., Dagliati, A., Chiasserini, D., Lee, D., Plant, D., Baricevic-Jones, I., Kelsall, J., Eineman, R., Reed, R., Geary, B., Unwin, R. D., Nicolaou, A., Keavney, B. D., Barton, A., Whetton, A. D., and Geifman, N. (2020) The use of missing values in proteomic data-independent acquisition mass spectrometry to enable disease activity discrimination. *Bioinformatics* 36, 2217-2223

43. Jin, L., Bi, Y., Hu, C., Qu, J., Shen, S., Wang, X., and Tian, Y. (2021) A comparative study of evaluating missing value imputation methods in label-free proteomics. *Sci Rep* 11, 1760

44. Gardner, M. L., and Freitas, M. A. (2021) Multiple imputation approaches applied to the missing value problem in bottom-up proteomics. *Int J Mol Sci* 22, 9650

45. Gao, Q., Zhu, H., Dong, L., Shi, W., Chen, R., Song, Z., Huang, C., Li, J., Dong, X., Zhou, Y., Liu, Q., Ma, L., Wang, X., Zhou, J., Liu, Y., Boja, E., Robles, A. I., Ma, W., Wang, P., Li, Y., Ding, L., Wen, B., Zhang, B., Rodriguez, H., Gao, D., Zhou, H., and Fan, J. (2019) Integrated Proteogenomic Characterization of HBV-Related Hepatocellular Carcinoma. *Cell* 179, 561-577.e522

46. Millán-Oropeza, A., Blein-Nicolas, M., Monnet, V., Zivy, M., and Henry, C. (2022) Comparison of different label-free techniques for the semi-absolute quantification of protein abundance. *Proteomes* 10, 2

47. Krey, J. F., Scheffer, D. I., Choi, D., Reddy, A., David, L. L., Corey, D. P., and Barr-Gillespie, P. G. (2018) Mass spectrometry quantitation of proteins from small pools of developing auditory and vestibular cells. *Sci Data* 5, 180128

48. Orsburn, B. C. (2021) Evaluation of the sensitivity of proteomics methods using the absolute copy number of proteins in a single cell as a metric. *Proteomes* 9, 34

49. Orsburn, B. C. (2021) Evaluation of the Sensitivity of Proteomics Methods Using the Absolute Copy Number of Proteins in a Single Cell as a Metric. *Proteomes* 9

50. Rodriguez, H., Zenklusen, J. C., Staudt, L. M., Doroshow, J. H., and Lowy, D. R. (2021) The next horizon in precision oncology: Proteogenomics to inform cancer diagnosis and treatment. *Cell* 184, 1661-1670

51. Mani, D. R., Krug, K., Zhang, B., Satpathy, S., Clauser, K. R., Ding, L., Ellis, M., Gillette, M. A., and Carr, S. A. (2022) Cancer proteogenomics: current impact and future prospects. *Nat Rev Cancer* 22, 298-313

52. Al Shweiki, M. H. D. R., Mönchgesang, S., Majovsky, P., Thieme, D., Trutschel, D., and Hoehenwarter, W. (2017) Assessment of Label-Free Quantification in Discovery Proteomics and Impact of Technological Factors and Natural Variability of Protein Abundance. *Journal of Proteome Research* 16, 1410-1424

53. Chang, C., Gao, Z., Ying, W., Fu, Y., Zhao, Y., Wu, S., Li, M., Wang, G., Qian, X., Zhu, Y., and He, F. (2019) LFAQ: Toward Unbiased Label-Free Absolute Protein Quantification by Predicting Peptide Quantitative Factors. *Analytical Chemistry* 91, 1335-1343

54. Nguyen, T. M. T., Kim, J., Doan, T. T., Lee, M.-W., and Lee, M. (2020) APEX Proximity Labeling as a Versatile Tool for Biological Research. *Biochemistry* 59, 260-269

55. Ghandi, M., Huang, F. W., Jané-Valbuena, J., Kryukov, G. V., Lo, C. C., McDonald, E. R., Barretina, J., Gelfand, E. T., Bielski, C. M., Li, H., Hu, K., Andreev-Drakhlin, A. Y., Kim, J., Hess, J. M., Haas, B. J., Aguet, F., Weir, B. A., Rothberg, M. V., Paolella, B. R., Lawrence, M. S., Akbani, R., Lu, Y., Tiv, H. L., Gokhale, P. C., de Weck, A., Mansour, A. A., Oh, C., Shih, J.,

Hadi, K., Rosen, Y., Bistline, J., Venkatesan, K., Reddy, A., Sonkin, D., Liu, M., Lehar, J., Korn, J. M., Porter, D. A., Jones, M. D., Golji, J., Caponigro, G., Taylor, J. E., Dunning, C. M., Creech, A. L., Warren, A. C., McFarland, J. M., Zamanighomi, M., Kauffmann, A., Stransky, N., Imielinski, M., Maruvka, Y. E., Cherniack, A. D., Tsherniak, A., Vazquez, F., Jaffe, J. D., Lane, A. A., Weinstock, D. M., Johannessen, C. M., Morrissey, M. P., Stegmeier, F., Schlegel, R., Hahn, W. C., Getz, G., Mills, G. B., Boehm, J. S., Golub, T. R., Garraway, L. A., and Sellers, W. R. (2019) Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* 569, 503-508

Table 1. Summary of differential expression analysis and top-ranked tumor enriched and highly expressed tumor antigens.

CPTAC indication	No. of upregulated proteins	No. of predicted cell surface proteins	Top 5 tumor-enriched antigens
BRCA	210	64	VAMP8, PYCARD, TMED2, IER3IP1, DPM3
COAD	236	46	ELANE, PRTN3, CEACAM5, FCER1G, GPRC5A
ccRCC	968	249	MIF, CRYAB, LGALS1, ANXA4, ANXA2
PDA	242	57	BGN, ANXA1, CD59, CTSE, THBS1
LUAD	463	115	HSPA5, HSPD1, SEC61B, PDIA4, SPCS3
UCEC	998	417	CALR, MT-CO2, CD9, ATP5IF1, COX6C
OV	159	28	CALR, HSPD1, HSP90AB1, C1QBP, PDIA4
HNSCC	477	185	ELANE, PRTN3, HLA-B, SLC2A1, TAPBP
GBM	207	52	APOA1, ANXA5, ANXA1, ANXA2, TF
LSCC	1065	209	HSP90AB1, HSPA5, SLC25A5, HSPD1, VDAC2

FIGURE LEGENDS

Figure 1. Workflow for identification of tumor-enriched and highly expressed cell surface antigens.

Figure 2. Estimation of absolute protein abundance of iBAQ derived from TMT and label-free proteomics data, respectively, in the CPTAC COAD data set. (A) LFQ-TPA and LFQ-iBAQ, as well as TMT-TPA and TMT-iBAQ, were highly correlated by Pearson correlation. (B). For overlap sample and proteins, the TPA and iBAQ methods showed relatively good correlation between LFQ and TMT data.

Figure 3. Validation of absolute protein abundance by the iBAQ method with CCLE cell line receptor density data. iBAQ quantity and receptor density for HER2 (A) and TROP2 (B) were highly correlated by Pearson correlation.

Figure 1

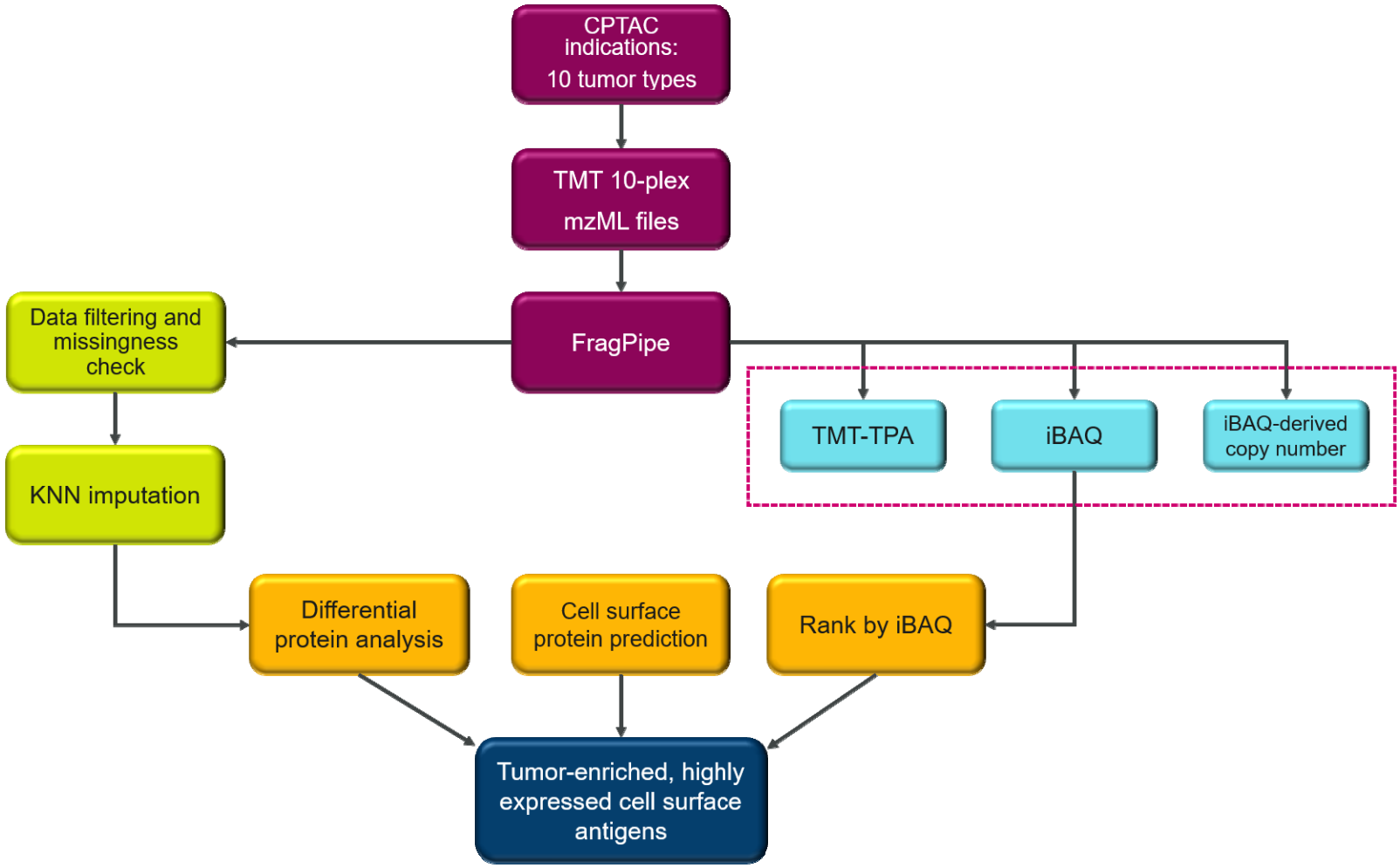
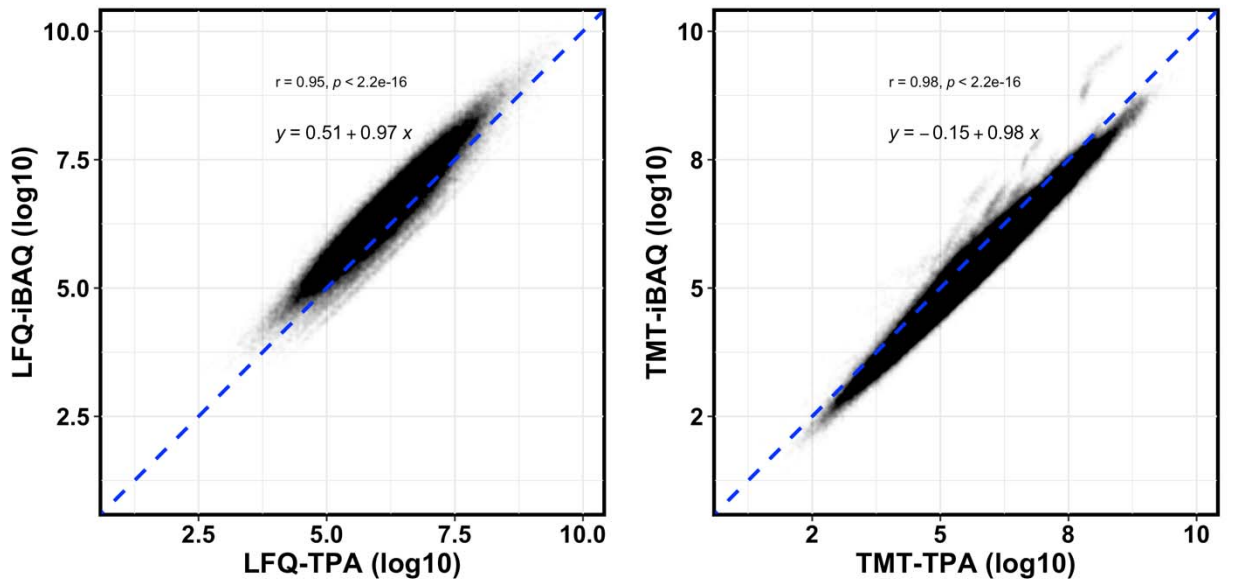

CPTAC indications:
10 tumor types
TMT 10-plex
mzML files
FragPipe
Data filtering and missingness check
KNN imputation
TMT-TPA
iBAQ
iBAQ-derived copy number
Differential protein analysis
Cell surface protein prediction
Rank by iBAQ
Tumor-enriched, highly expressed cell surface antigens

Figure 2

A


r = 0.95, p < 2.2e-16
y = 0.51 + 0.97 x
LFQ-iBAQ (log10)
LFQ-TPA (log10)
r = 0.98, p < 2.2e-16
y = −0.15 + 0.98 x
TMT-iBAQ (log10)
TMT-TPA (log10)

B

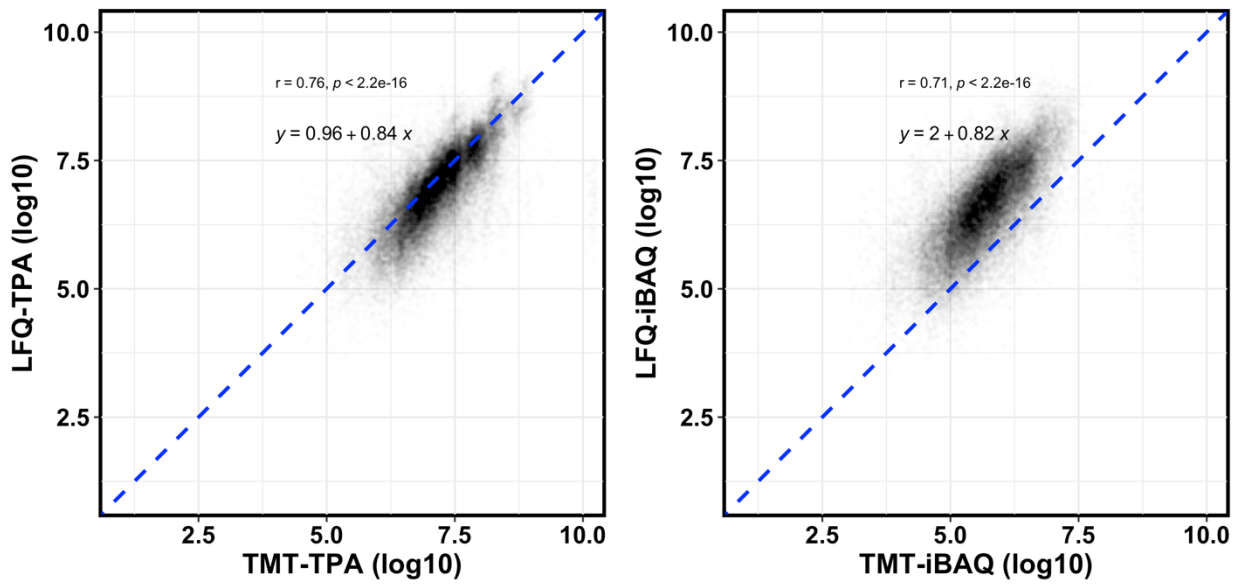
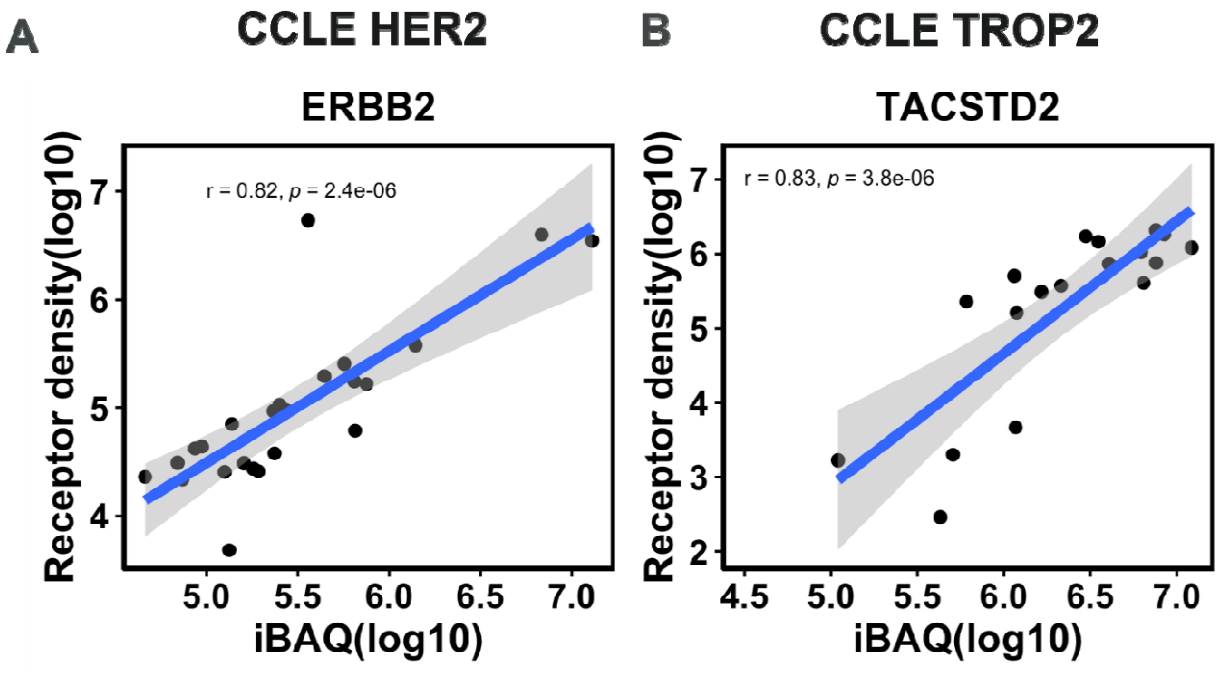

r = 0.76, p < 2.2e-16
y = 0.96 + 0.84 x
LFQ-TPA (log10)
TMT-TPA (log10)
r = 0.71, p < 2.2e-16
y = 2 + 0.82 x
LFQ-iBAQ (log10)
TMT-iBAQ (log10)

Figure 3


A
CCLE HER2
ERBB2
r = 0.82, p = 2.4e-06
Receptor density(log10)
iBAQ(log10)
B
CCLE TROP2
TACSTD2
r = 0.83, p = 3.8e-06
Receptor density(log10)
iBAQ(log10)