nature portfolio

https://doi.org/10.1038/s41568-022-00446-5

Supplementary information

Cancer proteogenomics: current impact and future prospects

In the format provided by the authors and unedited

Supplementary table 1

Study	PMID	Cancer type	Tumors	Matched Control tissues	Treatment	Omics data	Specific insights
Zhang et al., 2014	25043054	CRC	95	-	treatment naive	mutations CNV RNA proteome	First proteogenomic study of human cancer; Messenger RNA transcript abundance does not reliably predict protein abundance; proteomics data enables prioritization of candidate copy number drivers
Vasaikar et al., 2019	31031003	CRC	110	110	treatment naive	mutations CNV RNA mi-RNA proteome phospho	Rb phosphorylation is an oncogenic driver and a putative target in colon cancer; Increased glycolysis in MSI-H tumors is associated with immune suppression
Li et al., 2020	32888432	CRC	146 (76 primary; 70 metastatic) 43 liver metastases	145 146 remote normal	treatment naive	mutations CNV RNA proteome phospho	Phosphoproteomic pattern distinguishes metastasis and predicts drug response; Workflow from generation of large omics datasets to in vivo drug testing models; Improves the selection of treatment strategies for patients without druggable mutation
Zhang et al., 2016	27372738	OV	174	-	treatment naive	mutations CNV RNA proteome phospho	Acetylation of histone H4 correlates with homologous repair deficiency status; Protein and phosphoprotein abundance identifies pathways associated with survival
McDermott et al., 2020	32529193	OV	83	20 (relevant normal tissue)	treatment naive	mutations CNV RNA proteome phospho	Comparison of ovarian cancer and normal precursors identifies key signaling pathways; Mitotic and replication stress appears to drive increased chromosomal instability
Hu et al., 2020	33086064	OV	83	23 (relevant normal tissue)	treatment naive	proteome glycoproteome	Integrated proteomic and glycoproteomic analysis reveals tumor-specific glycosylation, and glycosylation enzymes correlated with glycosylation alterations
Dou et al., 2020	32059776	EC	95	49	treatment naive	methylation mutations CNV RNA mi-RNA proteome phospho acetylome	EC cells uses multiple mechanisms to suppress antigen processing and presentation, which may contribute to immune evasion; QKI, circRNAs, and miRNAs form a potential feedback loop to promote EMT
Mun et al., 2019	30645970	GC	80	80	treatment naive	mutations RNA proteome phospho glyco- proteome	mRNA-protein correlation nominates genes with high association with patient survival
Clark et al., 2019	31675502	ccRCC	103	80	treatment naive	methylation mutations CNV RNA proteome phospho	Use of microenvironment cell signatures to delineate four immune-based ccRCC subtypes associated with distinct cellular pathways
Gao et al., 2019	31585088	HCC	159	-	treatment naive	mutations RNA proteome phospho	Proteogenomic characterization of HCC suggests stratification of patient survival and treatment based on proteomic subtypes
Mertins et al., 2016	27251275	BRCA	105	3 (unmatched normal tissue)	treatment naive	mutations CNV RNA proteome phospho	Proof of concept that integrated analyses of global proteomic and phosphoproteomic with genomic and transcriptomic data can narrow down candidate driver genomic events and generate therapeutic hypotheses
Huang et al., 2017	28348404	BRCA PDX	24	-	treatment naive	mutations CNV RNA proteome phospho	Overexpression of key proteins, such as AKT, and phosphosites on ARAF, BRAF, HSP90AB1 are not readily explainable by genomic analysis
Mundt et al., 2018	29472518	BRCA TNBC PDX	6	-	Buparlisib (pan-PI3K inhibitor)	RNA proteome phospho	Integrative phosphoproteogenomic analysis uncovers potential intrinsic resistance mechanisms of TNBC tumors to PI3K inhibition.
Johansson et al., 2019	30962452	BRCA	45	-	treatment naive	mutations CNV RNA proteome metabolome	First study to recapitulate known function-enriched stratifications of breast tumors based on unbiased analyses of proteome profiles
Satpathy et al., 2020	31988290	BRCA ERBB2+	14	-	baseline and on-treatment trastuzumab, pertuzumab, chemo- therapy	mutations CNV RNA protein phospho	Application of a microscaled proteogenomic workflow to core biopsies from patients in a clinical trial uncovers potential causes of treatment resistance including the absence of ERBB2 amplification, insufficient ERBB2 activity for therapeutic sensitivity despite ERBB2 amplification, and alternative drivers, and suggests therapeutic alternatives.
Krug et al., 2020	33212010	BRCA	122	-	treatment naive	mutations CNV RNA proteome phospho acetylome	First deep-scale characterization of the acetylproteome in a large set of breast tumors exposed new details of breast cancer subtype-specific metabolism; Proteogenomic readout of Rb1 and ERBB2 enhances assessment of protein status with direct clinical implications..
Gillette et al., 2020	32649874	LUAD	110	101	treatment naive	methylation mutations CNV RNA mi-RNA proteome phospho acetylome	Deep-scale proteogenomic analysis of LUAD in a heterogeneous population of smokers and non-smokers from Asian and Western countries identifies ALK Y1507 as a diagnostic marker and SND1 phosphoprotein overexpression as potentially oncogenic in ALK fusion driven tumors, neutrophil degranulation as a potential immunosuppressive mechanism in STK11 mutant immune cold tumors, and Shp2 inhibition as a therapeutic alternative in EGFR mutant tumors.
Chen et al., 2020b	32649875	LUAD	103	103	treatment naive	mutations CNV RNA proteome phospho	First deep proteogenomic landscape of predominantly non-smoking lung adenocarcinoma in East Asia identifies age-, sex-, and environmental carcinogen-related mutagenic processes; notes potential therapeutic vulnerabilities in APOBEC-high tumors, and observes a subset of early stage tumors with "late-like" features and worse patient outcomes
Xu et al., 2020	32649877	LUAD	103	103	treatment naive	mutations CNV RNA proteome phospho	Identification of three molecular tumor subtypes in Chinese cohort and a subset of tumors with high expression of normal lung signature proteins associated with better patient survival; nomination of potential prognostic biomarkers and drug targets.
Stewart et al. 2019	31395880	LSCC	108	-	treatment naive	mutations CNV RNA proteome	Three proteomic subtypes with immune features (Inflamed-subtype) and metabolic vulnerabilities (TP63, PSAT1, and TFRC) in Redox-subtype

Satpathy et al., 2021	34358469	LSCC	108	99	treatment naive	methylation mutations CNV RNA mi-RNA proteome phospho acetylome ubiquitylome	Subtype with EMT and phosphoprotein signatures; Potential therapeutic vulnerabilities included survivin, NSD3, LSD1, and EZH2; Rb phosphorylation nominated as a biomarker for trials with CDK4/6 inhibitors; Detailed immune landscape analysis highlighted targetable points of immuneregulation
Archer et al., 2018	30205044	MB	45	-	treatment naive	methylation mutations CNV RNA proteome phospho acetylome	Subgroups of SHH tumors are distinctly regulated on post-transcriptional and post-translational levels despite showing similar RNA expression.
Forget et al., 2018	30205043	MB	41	-	treatment naive	methylation mutations CNV RNA proteome phospho	Highly divergent posttranscriptional pathway regulation in MB subgroups; Phosphoproteomic profiles reveal specific kinase activity in MB subgroups; Identification of aberrant ERBB4-SRC signaling as a hallmark of group 4 MBs; Over expression of activated SRC in the developing cerebellum induces MB
Rivero-Hinojosa et al,, 2018	29880060	MB	41	-	treatment naive	methylation mutations CNV RNA proteome	Explored effect of stability on correlation, showing that genes with similar stability in both mRNA and protein tend to have higher RNA-protein correlation; Influence of CNA and DNA methylation on the proteome; Identified EIF4F cap-dependent translation pathway as a druggable pathway in MB
Petralia et al., 2020	33242424	PBT	218	-	treatment naive; post-treatment	mutations CNV RNA proteomics phospho	Comprehensive proteogenomic survey across seven histologies of brain tumors uncovers foundational pediatric brain tumor biology and informs rational treatment selection
Wang et al., 2021	33577785	GBM	99	10 unmatched GTEx normla brain samples	treatment naive	methylation mutations CNV RNA snRNA proteome phospho acetylome lipidome metabolome	Comprehensive analysis of adult glioblastoma utilising 10 different omic platforms, including metabolomics, lipidomics, and single nuclei RNA-Seq; identificatio of signaling hubs, such as Phosphorylated PTPN11 and PLCG1, which could provide therapeutic vulnerabilities.
Yanovich-Arad et al., 2021	33657365	GBM	87	-	treatment naive	RNA sc-RNA proteome	Integration of proteomics and transcriptomics to study survival in GBM patients; Proteomics-based classification identifies three subtypes; Correlation analysis reveals survival patterns unique to the protein level; Incorporating single-cell published data associates subpopulations with survival
Huang et al. 2021	33417831	HNSCC	108	66	treatment naive	mutations CNV RNA mi-RNA proteome phospho	Widespread deletion of immune modulatory genes accounts for loss of immunogenicity; multi-omic analysis identifies therapeutic hypotheses that may inform more precise approaches to treatment of HPV-negative HNSCC.
Sinha et al., 2019	30889379	Prostate cancer	76	-	treatment naive	methylation mutations CNV RNA proteome	ETS fusions have divergent effects on transcriptome and proteome; Combining genomics and proteomics improves biomarker performance
Latonen et al. ,2018	29563510	Prostate cancer	38	-	treatment naive; post-treatment	methylation CNV RNA proteome	Genomic (Mutation, copy number) and epigenomic (DNA methylation) events that influence RNA expression do not correlate well with protein expression; study of miRNA-protein correlation identified regulatory events; pathway analysis of proteome detected metabolic shifts in the TCA cycle during cancer development and progression
Roper et al., 2019	30840888	metastatic thoracic cancer	40	-	chemotherapy and/or targeted therapy (range 2–8 lines of therapy)	mutations CNV RNA proteome	Identification of APOBEC mutagenesis, driven by germline variants, mutant TP53, and the tumor microenvironment, and late copy-number alterations as key mechanisms underlying proteogenomic evolution and heterogeneity.
Yang et al., 2019	30944321	ALL	27	-	treatment naive	mutations CNV RNA structural variants proteome	CTCF and cohesin have very low expression in hyperdiploid ALL compared to ETV6/RUNX1-positive ALL, leading to loss of insulation at topologically associating domain (TAD) borders; genes with low mRNA-protein correlations tend to be target of ubiquitin-proteasome pathway

Supplementary table 2

Software	Website	PMID	Brief Description
Data Access			
PDC	https://proteomic.datacommons.cancer.gov/pdc/		Making cancer-related proteomic datasets easily accessible to the public and facilitating multi-omic integration through interoperability with other resources.
GDC	https://gdc.cancer.gov/about-gdc/contributed-genomic-data-cancer-research/clinical-proteomic-tumor-analysis-consortium-cptac		Harmonizing DNA sequences from CPTAC whole genome sequencing, whole exomes sequencing, and RNAseq using GDC pipelines.
TCIA	https://wiki.cancerimagingarchive.net/display/Public/CPTAC+Imaging+Proteomics		Hosting both the radiology and pathology imaging data generated by CPTAC samples.
LinkedOmics	http://www.linkedomics.org/	29136207	Direct downloads for standardized multi-omic data matrices from the TCGA and CPTAC projects.
CPTAC (Python package)	https://paynelab.github.io/cptac/	33560848	Accessing and interacting with CPTAC data in python.
cBioPortal (Wu et al. 2019)	http://cbioportal.org/	31308250	Open source platform for exploring, visualizing, and analyzing multidimensional cancer genomics and clinical data, including CPTAC proteomic data for breast, colon and ovarian cancer.
Panorama	https://panoramaweb.org/home/project-begin.view?	25102069	Web application for storing, sharing, analyzing, and reusing targeted assays created and refined with Skyline. Panorama allows laboratories to store and organize curated results contained in Skyline documents with fine-grained permissions, which facilitates secure sharing of published and unpublished data via a web-browser interface.
Data Analysis - Proteomics Data Processing			
CDAP	https://pdc.cancer.gov/data-dictionary/harmonization.html		The CPTAC Common Data Analysis Platform for LC-MS/MS data.
Philosopher	https://philosopher.nesvilab.org/	32669682	A versatile toolkit for shotgun proteomics data analysis
MSFragger	http://msfragger.nesvilab.org/	28394336	Ultrafast database search tool for peptide identification in mass spectrometry-based proteomics across a wide range of datasets and applications. The speed of MSFragger makes it particularly suitable for the analysis of large datasets (including timsTOF PASEF data), for enzyme unconstrained searches (e.g. peptidome), for 'open' database searches (i.e. precursor mass tolerance set to hundreds of Daltons) for identification of modified peptides, and for glycopeptide identification (N-linked and O-linked) with MSFragger Glyco mode.
FragPipe	https://fragpipe.nesvilab.org/		Suite of computational tools enabling comprehensive and flexible analysis of mass spectrometry-based proteomics data. It is powered by MSFragger - an ultrafast search engine suitable for both conventional and 'open' (mass-tolerant) peptide identification. FragPipe includes the Philosopher toolkit for downstream post-processing of MSFragger search results (PeptideProphet, iProphet, ProteinProphet), FDR filtering, label-based quantification, and report generation. Also included in the FragPipe binary are Crystal-C and PTM-Shepherd to aid interpretation of open search results, TMT-Integrator for TMT/iTRAQ isobaric labeling-based quantification, IonQuant for label-free quantification, SpectraST and EasyPQP spectral library building modules, and DIA-Umpire SE module for direct analysis of data independent acquisition data.
Spectrum Mill	https://proteomics.broadinstitute.org/		Collection of tools for high- throughput processing of MS/MS and MS spectra to provide protein and peptide identifications and relative quantitation.
Skyline	https://skyline.ms/project/home/software/Skyline/begin.view	20147306	Windows client application for building Selected Reaction Monitoring (SRM) / Multiple Reaction Monitoring (MRM), Parallel Reaction Monitoring (PRM), DIA/SWATH and targeted DDA quantitative methods and analyzing the resulting mass spectrometry data.
Data Analysis – Proteomics Post-Processing			
TMT-Integrator	https://github.com/Nesvilab/TMT-Integrator		Normalizes and combines channel abundances from multiple TMT or iTRAQ-labeled proteomics samples, generating quantification reports as specified by the user. TMT-Integrator currently provides four quantification options: gene, protein, peptide, and modified site levels.
PTM-Shepherd	https://ptmshepherd.nesvilab.org/	33568339	Automated characterization of PTM profiles detected in open (mass-tolerant) proteomics searches based on attributes such as amino acid localization, fragmentation spectra similarity, retention time shifts, and relative modification rates. PTM-Shepherd can also perform multi-experiment comparisons for studying changes in modification profiles, e.g. in data generated in different laboratories or under different conditions.
IonQuant	http://ionquant.nesvilab.org/	32616513, 33813065	IonQuant is a label free quantification tool for shotgun proteomics. It supports timsTOF PASEF and non-timsTOF (e.g. Orbitrap) data, as well as matching-between-runs (MBR) and light/heavy chemical labeling.
Data QC			
COSMO	https://github.com/bzhanglab/COSMO		Identifying and correcting sample mislabeling in multi-omics data.
DREAM AI	https://github.com/WangLab-MSSM/DreamAI	https://doi.org/10.1101/2020.07.21.214205	An ensemble approach to missing value imputation for proteomics data developed by the NCI-CPTAC DREAM Proteogenomics Challenge in 2016
OmicsEV	https://bzhanglab.github.io/OmicsEV/		Comparing and evaluating data matrices generated from the same omics dataset using different tools, algorithms, or parameter settings.
MSInspector	https://skyline.ms/skyts/home/software/Skyline/tools/details.view?name=MSInspector		Python program for quality evaluation of the five assay characterization experiments outlined by CPTAC Assay Portal guidance document. MSInspector enables researchers to test their Skyline files for statistical calculation and data visualization through the built-in R scripts. The report file describes the details of any errors.
MassQC	https://massqc.proteomesoftware.com/		MassQC is an online Quality Control Tool that serves to diagnose liquid chromatography-mass spectrometry instrument hardware to ensure the instrument is running in a reproducible manner. Using data from CPTAC inter-lab studies, the National Institute of Standards and Technology(link is external) developed a number of metrics to assess instrument performance and ProteomeSoftware subsequently built a graphical user interface to commercialize this tool.
Data Analysis – Association Discovery and Visualization			
ProTrack-CCRCC	http://ccrcc.cptac-data-view.org/	32510176	Interactive visualization of CPTAC CCRCC multi-omics data.
ProTrack-PBT	http://pbt.cptac-data-view.org/	33242424	An Interactive Multi-Omics Data Browser for Proteogenomic Data of CPTAC-CBTTC pediatric brain tumor study
CPTAC-BRCA2016	https://proteomics.broadapps.org/CPTAC-BRCA2016/	27251275	Interactive visualization of retrospective CPTAC-BRCA multi-omics data.
CPTAC-BRCA2020	https://proteomics.broadapps.org/CPTAC-BRCA2020/	33212010	Interactive visualization of prospective CPTAC-BRCA multi-omics data.
CPTAC-LUAD2020	https://proteomics.broadapps.org/CPTAC-LUAD2020/	32649874	Interactive visualization of CPTAC-LUAD multi-omics data.
CPTAC-LSCC2021	https://proteomics.broadapps.org/CPTAC-LSCC2021/	34358469	Interactive visualization of CPTAC-LSCC multi-omics data.
LinkedOmics	http://www.linkedomics.org/	29136207	Analyzing multi-omics data from TCGA and CPTAC projects using association and pathway analysis, both within and across cancer types.
multiOmics Viz	https://www.bioconductor.org/packages/release/bioc/html/multiOmicsViz.html	https://doi.org/10.1101/2020.07.21.214205	Plotting the effect of one omics data on other omics data along the chromosome.
iProFun	https://rdrr.io/github/songxiaoyu/iProFun/	31227599	Characterizing functional consequences of DNA copy number and methylation alterations in tumors.
ProNetView	http://ccrcc.cptac-network-view.org/		Visualizing networks inferred based on proteogenomic data from CPTAC CCRCC project.
NetSAM	http://www.bioconductor.org/packages/release/bioc/html/NetSAM.html	23807191	Network seriation and modularization.
NetGestalt	http://www.netgestalt.org/	23807191	Visualizing and analyzing multi-omics data based on biological networks.
Panoptes	https://github.com/rhong3/Panoptes https://pypi.org/project/panoptes-he/		Analyzing pathological images using a novel InceptionResnet-based convolutional neural network architecture.
TSNet	https://github.com/petraf01/TSNet	29949994	A statistical model for cell type specific inference based on bulk tumor profiling data.
Data Analysis - Sequence-based Integration			
CustomProDB	http://bioconductor.org/packages/release/bioc/html/customProDB.html	24058055	Generating customized databases from DNA and RNA sequencing data for proteomics search.
QUILTS	http://openslice.fenyolab.org/cgi-bin/pyquilts_cgi.pl	26631509	Creating sample specific protein sequence databases using genomic and transcriptomic data.
PepQuery	http://pepquery.org	30610011	Proteomic validation of novel genomic alterations without customized database construction.
NeoFlow	https://github.com/bzhanglab/neoflow	32273506	A proteogenomics pipeline for neoantigen discovery.
Data Analysis - PTM and Pathway Analysis			
PTMsigDB	ptmsigdb.org	30563849	A collection of modification site-specific signatures of perturbations, kinase activities and signaling pathways curated from literature.
PTM-SEA	https://github.com/broadinstitute/ssGSEA2.0	30563849	A modified version of ssGSEA to perform site-specific signature analysis by scoring PTMsigDB's bi-directional signature-sets.
WebGestalt	http://www.webgestalt.org	31114916	A gene set analysis toolkit with new support for phosphosite enrichment analysis.
PTMcosmos	https://ptmcosmos.wustl.edu/		A database that stores post-translational modifications (PTMs) and cancer mutations in humans.
HotPho	https://github.com/ding-lab/HotPho_Analysis		Identify spatially interacting phosphorylation sites and mutations by co-clustering somatic mutations and phoshosites
Black Sheep	https://github.com/ruggleslab/blackSheep	34165986	Outlier analysis of proteogenomic datasets to identify samples with aberrantly high or low levels of genes, proteins or PTM sites.
WikiPathways	https://www.wikipathways.org/index.php/Portal:CPTAC	33211851	WikiPathways is a database of molecular pathway diagrams contributed and refined by the research community. A dedicated portal was established for CPTAC pathways (link is external) including the hallmarks of cancer and phosphorylation state information. Browse, download, analyze or draw your own pathways of interest.
Data Analysis - Proteogenomic Pipeline			
PANOPLY	https://github.com/broadinstitute/PANOPLY	34040252	A cloud-based platform for automated and reproducible proteogenomic data analysis