

THE CLINICAL PROTEOMIC TUMOR ANALYSIS CONSORTIUM SCIENTIFIC SYMPOSIUM

Wednesday, October 16, 2019

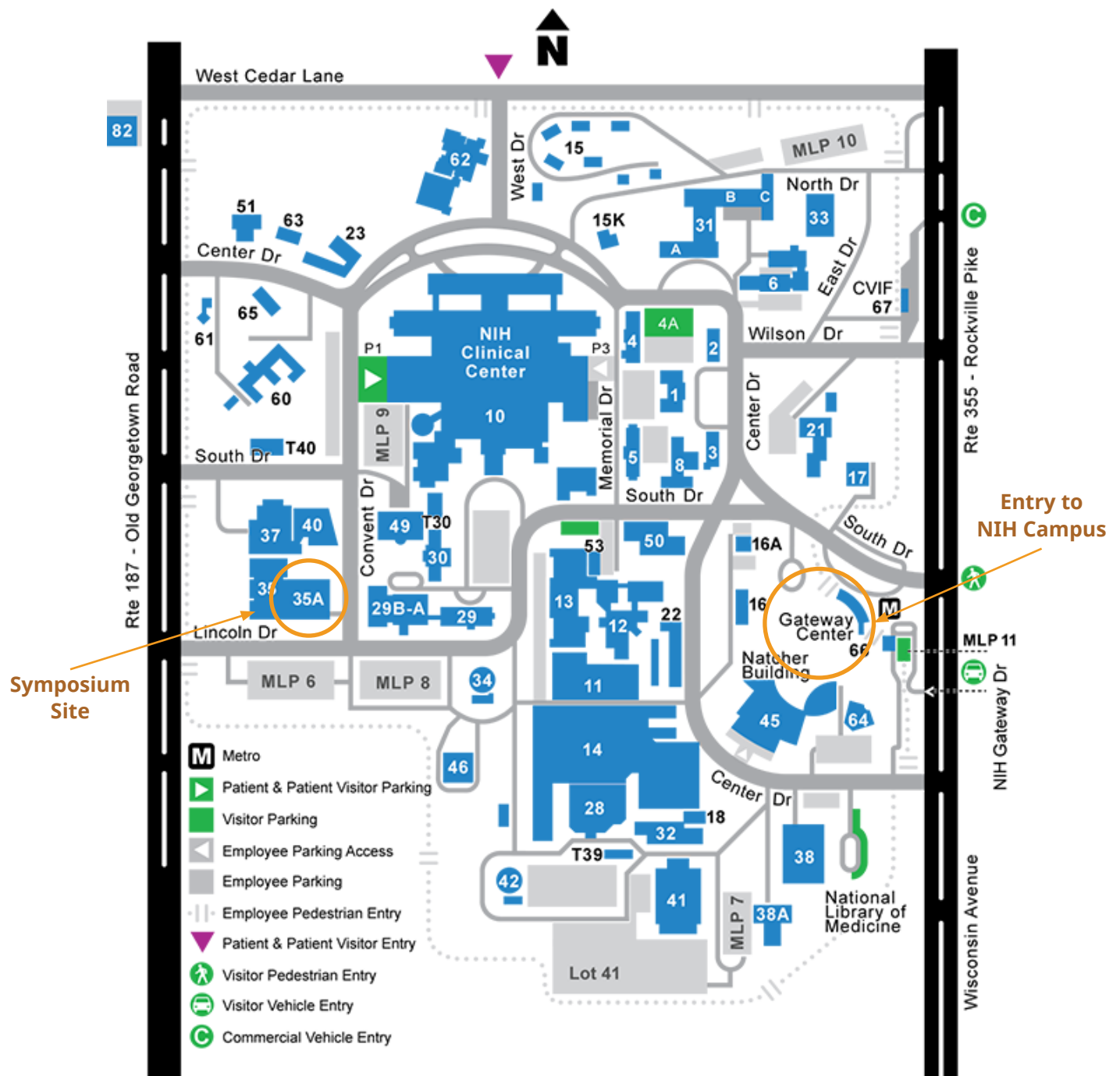
National Institutes of Health
John Edward Porter Neuroscience
Research Center
35 Convent Drive
Bethesda, Maryland 20892



NATIONAL CANCER INSTITUTE
Office of Cancer Clinical
Proteomics Research



NATIONAL INSTITUTES OF HEALTH, BETHESDA CAMPUS



WELCOME

Dear All:

On behalf of the National Cancer Institute (NCI), the Office of Cancer Clinical Proteomics Research (OCCPR) team welcomes you to the Clinical Proteomic Tumor Analysis Consortium (CPTAC) 2019 Scientific Symposium. We are excited to share the recent advances our program has made in the field of proteomics, proteogenomics and precision oncology in the fight against cancer.

CPTAC represents a coordinated effort dedicated to the application of proteogenomics, a technique that applies deep, comprehensive proteomics (proteins) with genomics (DNA and RNA) to the analysis of cancer. In this unique approach, the program aims to reignite the field of protein science as a means to realize the promise of genomics to advance biology and promote human health.

The CPTAC program began as a part of the Clinical Proteomic Technologies for Cancer (CPTC) initiative at the NCI. Its purpose was to understand and address sources of analytical error in proteomic technologies through standards and technology development, in order to produce quality data (rigor & reproducibility) that would complement genomic and transcriptomic analyses.

The CPTAC program grew to become its own entity as a collaborative network to apply the program's state-of-the-art standardized proteomic workflows to genomically-characterized tumors from The Cancer Genome Atlas (TCGA). Focusing on colorectal, ovarian and breast cancers, the programs' new goals were to identify and quantitate proteins that derive from alterations in cancer genomes in order to better understand the proteogenomic cellular complexity of cancer not fully elucidated through genomics.

As the CPTAC program matured, it has expanded its efforts to the comprehensive proteogenomic characterization of additional cancer types and to the support of clinically-relevant research projects that elucidate biological mechanisms of resistance/response/toxicity in support of NCI-sponsored clinical trials.

Lastly, as a part of its mission, the CPTAC program distributes all data, computational tools, and reagents (antibodies and standards), to the research community through public databases (and collaborative partnerships) to empower academic and industry scientists worldwide so to maximize utility and public benefit and to accelerate cancer research and advance patient care.

We are excited to have you join us as we share the wealth of knowledge we have gained throughout the years at this special Symposium.

Sincerely,



Henry Rodriguez, Ph.D., M.B.A.
Director, Office of Cancer Clinical Proteomics Research

CPTAC OVERVIEW

The goals of CPTAC are to accelerate the understanding of tumor biology (genotype-to-phenotype) through community resources that serve as a foundation for “hypothesis-driven science by others,” and to support clinical research projects that address mechanisms of treatment response, resistance, and/or toxicity. To achieve these goals, **CPTAC has two united and coordinated programs: The Tumor Characterization Program and the Translational Research Program.** In the Tumor Characterization Program, CPTAC expands deep comprehensive proteogenomic analysis to prospectively-collected treatment-naïve tumors. In the Translational Research Program, CPTAC partners with NCI-sponsored clinical trials to support clinically-relevant research projects that elucidate biological mechanisms of response, resistance, and/or toxicity.

All genomics, proteomics, imaging data, assays, and reagents are made publicly available to the research community to maximize utility and public benefit. Specifically, genomic data is sent to the Genomic Data Commons (GDC), proteomics data is sent to the CPTAC Data Portal and the Proteomics Data Commons (PDC), and imaging data sent to The Cancer Imaging Archive (TCIA). CPTAC also makes antibodies, fit-for-purpose assays, and corresponding protocols available to the cancer research community through the CPTAC Antibody and Assay Portals. This unique combination of coordinating research approaches, sharing research data, and combining proteomic with genomic analysis, should allow the CPTAC program to produce a more unified understanding of cancer biology with translational potential.

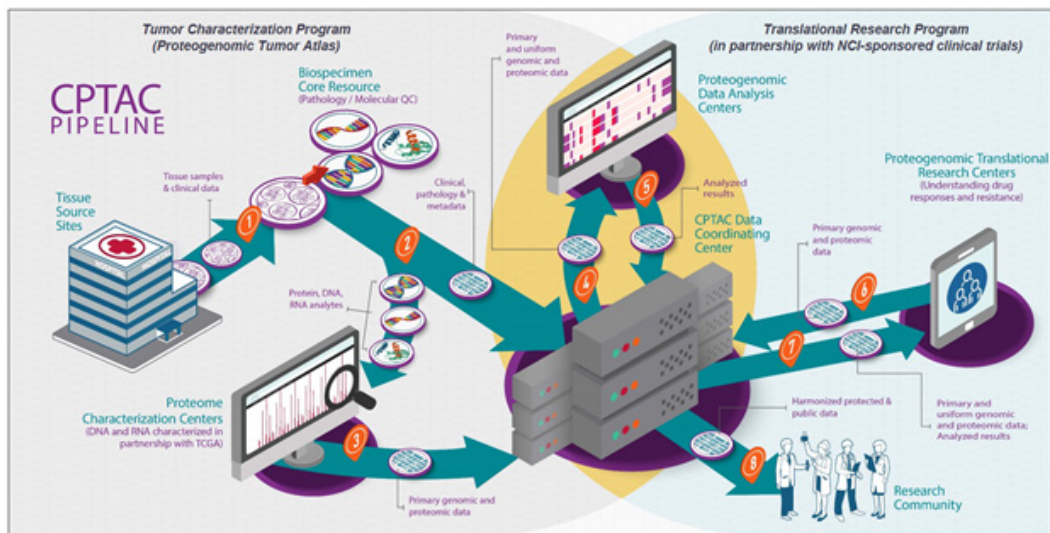
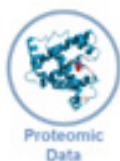


Figure 1. Illustration of the CPTAC Pipeline, which works to transform tissue samples into publicly available resources and reagents for the cancer research community.



TEAM LEADERSHIP BIOSKETCHES

Proteome Characterization Center Leadership Biosketches

(PCC teams are listed alphabetically)

Broad Institute

Steven Carr, Ph.D.

Steven Carr is senior director of proteomics at the Broad Institute and an institute scientist. Research in his lab focuses on detecting and quantifying post-translational modifications (phosphorylation, acetylation, ubiquitylation, methylation, etc.) in the proteome, developing new technologies to quantify proteins in cells and biofluids with high sensitivity and specificity, improving informatics for peptide and protein assignment using mass spectrometry (MS) data, and integrating MS-derived data with genomic data.

Michael Gillette, M.D., Ph.D.

Dr. Michael Gillette is an instructor at the Dana-Farber Cancer Institute and senior group leader in the Biomarker Discovery Program in the Proteomics Platform at the Broad Institute, a Pulmonary and Critical Care Medicine Attending and Associate Physician at the Massachusetts General Hospital, and Assistant Professor of Medicine at Harvard Medical School. He is a research scientist with expertise in application and development of MS-pattern-based biomarker discovery.

Johns Hopkins University

Daniel Chan, Ph.D., DABCC, FACB

Dr. Chan is the director of the Clinical Chemistry Division and Co-Director of the Pathology Core Laboratory of the Johns Hopkins Hospital, as well as a professor of Pathology, Oncology, Radiology and Radiological Science, and Urology at the Johns Hopkins University School of Medicine. He also serves as director of the Center for Biomarker Discovery and Translation. Dr. Chan is a diplomate of the American Board of Clinical Chemistry, a fellow of the National Academy of Clinical Biochemistry, and an active member of several professional societies, including the American Association for Clinical Chemistry and US HUPO. The focus of his research is the development and application of proteomic and immunologic techniques in the diagnosis, management and understanding of cancer.

Hui Zhang, Ph.D.

Dr. Hui Zhang is a professor of pathology and associate professor of oncology at the Johns Hopkins University School of Medicine. She is also the director of Mass Spectrometry Core Facility, Center for Biomarker Discovery and Translation. She specializes in clinical chemistry, with particular emphasis on protein modification on the proteome scale and the effects of modification on protein function and disease progression.

Zhen Zhang, Ph.D.

Dr. Zhang is the associate director of the Department of Pathology's Center for Biomarker Discovery and Translation, as well as an associate professor of pathology and oncology at the Johns Hopkins University School of Medicine. An expert in bioinformatics and biostatistics, he is an investigator in both the Division of Clinical Chemistry and the Sidney Kimmel Comprehensive Cancer Center. His research focuses on the design of large-scale discovery studies for biomarkers and other molecular targets, and their translation into clinical applications.

Pacific Northwest National Laboratory

Tao Liu, Ph.D.

Dr. Liu is a senior scientist in the Integrative Omics Group in the Biological Sciences Division at PNNL, and principle investigator of the National Cancer Institute's Early Detection Research Network Biomarker Reference and Resource Center at PNNL. His work at PNNL has centered on developing and applying innovative methods and technologies including targeted proteomics (e.g., selected reaction monitoring) that enable high-throughput, sensitive, and quantitative LC-MS-based proteomics measurements of complex biological and clinical samples.

Richard Smith, Ph.D.

Richard D. Smith, Ph.D., is a Battelle Fellow and chief scientist in the Biological Sciences Division and director of Proteomics Research at Pacific Northwest National Laboratory (PNNL). Dr. Smith also is director of the NIH Research Resource for Integrative Biology, an adjunct faculty member in the chemistry departments at Washington State University and the University of Utah, and an affiliate faculty member in the Department of Chemistry at the University of Idaho and the Department of Molecular Microbiology & Immunology at Oregon Health & Science University. His research interest has broadly involved the development and application of advanced methods, instrumentation, and informatics capabilities and their applications in biological research, with particular emphasis on proteomics and metabolomics.

Proteogenomic Data Analysis Center Leadership Biosketches

(PDAC teams are listed alphabetically)

Baylor College of Medicine

Bing Zhang, Ph.D.

Dr. Bing Zhang is a McNair Scholar, a Cancer Prevention Research Institute of Texas (CPRIT) Scholar, and professor in Molecular and Human Genetics at the Lester and Sue Smith Breast Center within the NCI-designated Dan L. Duncan Comprehensive Cancer Center at Baylor College of Medicine. Dr. Zhang's research interest lies in developing computational and statistical approaches that help translate multidimensional omics data into biological and clinical insights.

Broad Institute

D.R. Mani, Ph.D.

D.R. Mani is a principal computational scientist in the Proteomics Platform at the Broad Institute. His recent research has focused on the design and implementation of innovative algorithms to enable proteogenomic analysis, pattern-based discovery of proteomic biomarker candidates, evaluation of data quality, assessment of variability and reproducibility in mass spectrometry based assays, and data visualization.

Gad Getz, Ph.D.

Gad Getz directs the Cancer Genome Computational Analysis group at the Broad Institute, where he is also an institute member. In addition to his role at the Broad, Getz is a co-principal investigator in the Genome Data Analysis Center of the NCI/NHGRI The Cancer Genome Atlas project; a co-leader of the International Cancer Genome Consortium Pan-Cancer Analysis of Whole Genomes project; a co-principal investigator of the Broad-led NCI Cloud Pilot; and a member of various NCI advisory committees. In addition, Getz directs the Bioinformatics Program at the Massachusetts General Hospital Cancer Center and Department of Pathology and serves as an associate professor of pathology at Harvard Medical School. Getz is also the inaugural incumbent of the Paul C. Zamecnik Chair in Oncology at the MGH Cancer Center. The Getz Laboratory specializes in cancer genome analysis.

Chet Birger, Ph.D.

Dr. Birger is the associate director at Broad Institute, and principal architect of the Cancer Genome Computational Analysis (CGCA) Group, developing the Broad's NCI Cloud Pilot. His research includes development of cloud-based analysis platforms for cancer genome data.

Icahn School of Medicine at Mount Sinai

Pei Wang, Ph.D.

Dr. Wang is an associate professor of Genetics and Genomics at Icahn Medical School at Mount Sinai, and a member of the Icahn Institute for Data Science and Genomic Technology. Her research focuses on developing statistical and computational methods to address scientific questions based on data from high throughput biology/genetics experiments.

Eric Schadt, Ph.D.

Dr. Eric Schadt, is Dean of the Department of Genetics and Genomics Sciences, professor of Predictive Health and Computational Biology, and a member of the Icahn Institute of Data Science and Genomic Technology at the Icahn School of Medicine at Mount Sinai. He is also founder and CEO of Sema4, a Mount Sinai Venture. Dr. Schadt research focuses on the generation and integration of very large-scale sequence variation, molecular profiling and clinical data in disease populations for constructing molecular networks that define disease states and link molecular biology to physiology.

New York University; Washington University in St. Louis; and Pacific Northwest National Lab

David Fenyö, Ph.D.

Dr. Fenyö is currently a professor in the Department of Biochemistry and Molecular Pharmacology, Interim Director for the Center for Health Informatics and Bioinformatics at New York University Langone Medical Center, director for the Biomedical Informatics Training Program and Biomedical Informatics Shared Resources, Co-Director for the Biomedical Informatics Core at the Clinical and Translational Science Institute, and adjunct faculty at Rockefeller University. Dr. David Fenyö's research focuses on providing a detailed understanding of the dynamics of cellular processes.

Li Ding, Ph.D.

Dr. Li Ding is the assistant director of the McDonnell Genome Institute at Washington University School of Medicine, and director of Computational Biology and Oncology. She is also an associate professor in the Department of Medicine and Genetics. Dr. Ding's research focuses on identifying and characterizing somatic/germline genetic changes relevant to cancer initiation and progression as well as drug response by integrating various data types including DNA, RNA, and proteomics data.

Samuel Payne, Ph.D.

Dr. Payne is a bioinformatics professor at Brigham-Young University. Dr. Payne's research interests are focused on algorithms for proteomics data analysis and subsequent interpretation and integration with an emphasis on integrative omics, active data, and metaproteomics data analysis.

University of Michigan

Alexey Nesvizhskii, Ph.D.

Dr. Nesvizhskii is a professor of Computational Medicine and Bioinformatics and professor of Pathology at the University of Michigan Medical School. He also serves as the director of the Proteomics Resource Facility. Dr. Nesvizhskii's research interest is in the field of quantitative proteomics, with a focus on the development of computational methods for processing and extracting biological information from complex proteomic datasets.

Arul Chinnaiyan, M.D., Ph.D.

Dr. Chinnaiyan is a Howard Hughes Medical Institute Investigator, American Cancer Society Research Professor, and a S.P. Hicks Endowed Professor of Pathology and Urology at the University of Michigan. He also serves as the inaugural director of the Michigan Center for Translational Pathology, focusing his research on functional genomic, proteomic and bioinformatics approaches to study cancer for the purposes of understanding cancer biology as well as to discover clinical biomarkers.

S. Mohan Dhanasekaran, Ph.D.

Dr. Dhanasekaran is an associate research scientist in the Department of Pathology and Michigan Center for Translational Pathology (MCTP) at the University of Michigan Medical School. His research interests include cancer genomics, gene fusions in cancer and cancer epigenetics with a focus on the role of DNA methylation and polycomb group proteins in cancer progression.

Proteogenomic Translational Research Center Leadership Biosketches

(PTRC teams are listed alphabetically)

Baylor College of Medicine; Broad Institute

Matthew J. Ellis, MB, BChir, BSc., PhD, FRCP

Dr. Ellis, a McNair Scholar, serves as the director and professor of the Lester and Sue Smith Breast Center and associate director of Precision Medicine at the Dan.L Duncan Comprehensive Cancer Center at Baylor College of Medicine. He is currently co-chair of the Translational Medicine Committee for the NRG cooperative group and co-leader for The Cancer Genome Atlas Breast Project. The Ellis lab is interested in delineating mechanisms of resistance to standard-of-care therapeutics and metastasis for each breast tumor subtype.

Steven Carr, Ph.D.

See aforementioned under proteome characterization center leadership biosketches.

Fred Hutchinson Cancer Research Center, University of Alabama at Birmingham

Amanda Paulovich, M.D., Ph.D.

Dr. Paulovich is the director of the Clinical Research Proteomics Platform at Brotman Baty Institute for Precision Medicine, and a member of the Clinical Research Division at Fred Hutchinson as well as a member of the Fred Hutchinson/University of Washington Cancer Consortium. Additionally, Dr. Paulovich is a professor in the Department of Medicine/Division of Oncology, associate faculty member of the Molecular and Cellular Biology Program at the University of Washington School of Medicine, and the Aven Foundation Endowed Chair at Fred Hutch. An oncologist by training, Dr. Paulovich's research focuses on developing technologies and strategies for translation of novel diagnostics and therapeutics to enable precision medicine and measure clinically relevant phenotypes in patients.

Michael Birrer, M.D., Ph.D.

Dr. Birrer is the director of the University of Alabama at Birmingham Comprehensive Cancer Center and is a professor of medicine in the Division of Hematology-Oncology, Pathology, and Obstetrician and Gynecology Oncology. He holds the Evalina B. Spencer Chair in Oncology and is a senior scientist at the O'Neal Comprehensive Cancer Center. The Birrer laboratory focuses on characterizing the molecular origin of gynecological cancers that includes identifying and characterizing oncogenic mutations and tumor suppressor genes in ovarian, endometrial and cervical cancers.

Pacific Northwest National Laboratory; Oregon Health & Science University

Karin Rodland, Ph.D.

Dr. Rodland is a laboratory fellow and chief scientist for Biomedical Research at Pacific Northwest National Laboratory, and a member of the board of directors for the U.S. Human Proteome Organization. Dr. Rodland also holds a joint appointment as Affiliate Professor of Cell, Developmental, and Cancer Biology at Oregon Health & Science University, and is a Fellow of the American Association for the Advancement of Science. Dr. Rodland's current research focuses on improving the ability to identify and validate biomarkers of disease by combining expert knowledge of cellular pathways with statistical approaches.

Brian J. Druker, M.D.

Dr. Druker holds multiple positions at Oregon Health and Science University in the School of Medicine. He is a professor of Medicine in the Division of Hematology and Medical Oncology, associate Dean of Oncology in the Office of the Dean and Director of Oregon Health and Science University Knight Cancer Institute. Additionally, Dr. Druker is the JELD-WEN Chair of Leukemia Research. Dr. Druker's research focuses on activated tyrosine kinases with an emphasis on signal transduction, cellular transformation and its application to cancer therapies.

PARTICIPANT LIST

Matthew Anderson, M.D., Ph.D.

University of South Florida

Mingghui Ao

Johns Hopkins University

Ozgun Babur, Ph.D.

Oregon Health & Science University

Jill Barnholtz-Sloan, Ph.D.

Case Western Reserve University

Jodi Basner, Ph.D.

Clarivate

Oliver Bathe, M.D., FRCS

University of Calgary

Michelle Berny-Lang, Ph.D.

Center for Strategic Scientific Initiatives, NCI

Chet Birger, Ph.D.

Broad Institute

Michael Birrer, M.D., Ph.D.

University of Alabama at Birmingham

Lili Blumenberg

New York University Medical Center

Simina Boca, Ph.D.

Georgetown University

William Bocik, Ph.D.

Leidos/Frederick National Laboratory for Cancer Research - Center for Strategic Scientific Initiatives, NCI

Hannah Boekweg

Brigham Young University

Emily Boja, Ph.D.

Center for Strategic Scientific Initiatives, NCI

Melissa Borucki, M.S., M.B.A.

Leidos

Meghan Burke, Ph.D.

National Institute of Standards and Technology

Shuang Cai, Ph.D.

ESAC, Inc.

Anna Calinawan

Icahn School of Medicine at Mount Sinai

Liwei Cao, Ph.D.

Johns Hopkins University

Song Cao, Ph.D.

Washington University School of Medicine

Steven Carr, Ph.D.

Broad Institute

Daniel Chan, Ph.D., DABCC, FACB

Johns Hopkins University

Feng Chen, Ph.D.

Washington University School of Medicine

Xi Chen, Ph.D.

University of Miami

KyungCho (Ryan) Cho, Ph.D.

Johns Hopkins University

Shrabanti Chowdhury, Ph.D.

Ichan School of Medicine at Mount Sinai

Marcin Cieslik, Ph.D.

University of Michigan

David Clark, Ph.D.

Johns Hopkins University

Karl Clauser, Ph.D.

Broad Institute

Simona Colantonio, Ph.D.

Leidos/Frederick National Laboratory for Cancer Research, NCI

Antonio Colaprico, Ph.D.

University of Miami

Emily Coolbaugh

University of Michigan

Aubrey Coulas

Duke University

Daniel Cui-Zhou

Washington University in St. Louis

Felipe da Veiga Leprevost, Ph.D.

University of Michigan

Tanja Davidsen, Ph.D.

Center for Biomedical Informatics and Information
Technology, NCI

Emek Demir, Ph.D.

Oregon Health & Science University

Saravana Dhanasekaran, Ph.D.

University of Michigan

Li Ding, Ph.D.

Washington University School of Medicine

Marcin Domagalski, Ph.D.

ESAC, Inc.

Yongchao Dou, Ph.D.

Baylor College of Medicine

Brian Druker, M.D.

Oregon Health & Science University

Maureen Dyer, Ph.D.

Leidos

Nathan Edwards, Ph.D.

Georgetown University

Kimberly Elburn, M.S., PMP

Leidos

Carol Elliot, M.S., CG(ASCP), CCRP

St. Joseph's Hospital and Medical Center

Matthew Ellis, M.B., B.Chir., B.Sc., Ph.D., FRCP

Baylor College of Medicine

Adel El-Naggar, M.D., Ph.D.

University of Texas MD Anderson Cancer Center

David Fenyő, Ph.D.

New York University

Brenda Fevrier-Sullivan, CAP

National Cancer Institute, NIH

Barbara Foster, Ph.D.

Roswell Park Comprehensive Cancer Center

Alicia Francis, M.S.

ESAC, Inc.

Kar-Ming Fung, M.D., Ph.D.

University of Oklahoma Health Sciences Center

Gad Getz, Ph.D.

Broad Institute

Michael Gillette, M.D., Ph.D.

Broad Institute

Charles Goldthwaite, Ph.D.

National Cancer Institute, NIH

Ramaswamy Govindan, M.D.

Washington University School of Medicine

Zeynep Gumus, Ph.D.

Icahn School of Medicine at Mount Sinai

Sean Hanlon, Ph.D.

Center for Strategic Scientific Initiatives, NCI

Linda Hannick, Ph.D.

Leidos/Center for Strategic Scientific Initiatives, NCI

Dawn Hayward

Center for Strategic Scientific Initiatives, NCI Fellow

David Heiman

Broad Institute

Beth Hermes

St. Joseph's Hospital and Medical Center

Tara Hiltke, Ph.D.

Center for Strategic Scientific Initiatives, NCI

Runyu Hong

New York University School of Medicine

Galen Hostetter, M.D.

Van Andel Institute

Ralph Hruban, M.D.

Johns Hopkins University

Yingwei Hu, Ph.D.

Johns Hopkins University

Antonio Lavarone, M.D.

Columbia University Medical Center

Eric Jaehnig Ph.D., M.B.E.

Baylor College of Medicine

Scott Jewell, Ph.D.

Van Andel Research Institute

Jiayi Ji

Icahn School of Medicine at Mount Sinai

Selim Kalayci

Icahn School of Medicine at Mount Sinai

Alla Karpova, M.S.

Washington University in St. Louis

Emily Kawaler, M.S.

New York University

La'Toya Kelly

Center for Strategic Scientific Initiatives, NCI

Karen Ketchum, Ph.D.

ESAC, Inc.

Christopher Kinsinger, Ph.D.

Center for Strategic Scientific Initiatives, NCI

Justin Kirby

Leidos

John Koomen, Ph.D.

H. Lee Moffitt Cancer Center and Research Institute

Ramani Kothadia

Broad Institute

Azra Krek, Ph.D.

Icahn School of Medicine at Mount Sinai

Karsten Krug, Ph.D.

Broad Institute

Yelena Krutikova, M.D., Ph.D.

ProteoGenex, Inc.

Toan Le, M.S.

ESAC, Inc.

Kai Li

Baylor College of Medicine

Yize Li

Washington University School of Medicine

Qin Li, M.Sc.

BioMatrix

Qing Kay Li, M.D., Ph.D.

Johns Hopkins University

Wen-Wei Liang

Washington University in St. Louis

Tung-Shing Lih, Ph.D.

Johns Hopkins University

Caleb Lindgren

Brigham Young University

W. Marston Linehan, M.D.

National Cancer Institute, NIH

Tao Liu, Ph.D.

Pacific Northwest National Laboratory

Wenke Liu, Ph.D.

New York University

Jiang Long, M.D.

Shanghai Cancer Center Fudan University

Yin Lu

ESAC, Inc.

Weiping Ma, Ph.D.

Icahn School of Medicine at Mount Sinai

Shiyong Ma, Ph.D.

Johns Hopkins School of Medicine

D.R. Mani, Ph.D.

Broad Institute

Annette Marrero-Oliveras, M.S.

Center for Strategic Scientific Initiatives, NCI

Michaela McCowin

Brigham Young University

Jason McDermott, Ph.D.

Pacific Northwest National Laboratory

Peter McGarvey, Ph.D.

Georgetown University School of Medicine

Wilson McKerrow, Ph.D.

New York University School of Medicine

Anand Merchant, M.D., Ph.D.

Frederick National Laboratory for Cancer Research, NCI

Mehdi Mesri, M.Med.Sci., Ph.D.

Center for Strategic Scientific Initiatives, NCI

Helen Moore, Ph.D.

National Cancer Institute, NIH

Mike Moser, Ph.D.

Roswell Park Comprehensive Cancer Center

Alexey Nesvizhskii, Ph.D.

University of Michigan

Chelsea Newton

Van Andel Research Institute

Kristen Nyce, M.F.A.

ESAC, Inc.

Lindsey Olsen

Brigham Young University

Gil Omenn, M.D., Ph.D.

University of Michigan

Zoe Osborne

University of Calgary

Amanda Paulovich, M.D., Ph.D.

Fred Hutchinson Cancer Research Center

Samuel Payne, Ph.D.

Pacific Northwest National Laboratory

Francesca Petralia, Ph.D.

Icahn School of Medicine at Mount Sinai

Paul Piehowski, Ph.D.

Pacific Northwest National Laboratory

Samuel Pugh

Brigham Young University

Liqun Qi

Frederick National Laboratory for Cancer Research, NCI

Adam Resnick, Ph.D.

Children's Hospital of Philadelphia

Boris Reva, Ph.D.

Icahn School of Medicine at Mount Sinai

Christopher Ricketts, Ph.D.

National Cancer Institute, NIH

Anders Ruita

WikiPathways

Anna Roberts-Pilgrim, Ph.D.

Center for Strategic Scientific Initiatives, NCI

Karna Robinson, M.P.H.

Leidos

Ana Robles, Ph.D.

Center for Strategic Scientific Initiatives, NCI

Karin Rodland, Ph.D.

Pacific Northwest National Laboratory

Henry Rodriguez, Ph.D., M.S., M.B.A.

Center for Strategic Scientific Initiatives, NCI

Dan Rohrer, M.S., M.B.A.

Van Andel Research Institute

Brian Rood, M.D.

Children's National Medical Center

Kelly Ruggles, Ph.D.

New York University

Dmitry Rykunov, Ph.D.

Icahn School of Medicine at Mount Sinai

Shankha Satpathy, Ph.D.

Broad Institute

Sara Savage, Ph.D.

Baylor College of Medicine

Michael Schnaubelt

Johns Hopkins University

Yvonne Shutack, M.D.

Clinical Proteomic Tumor Analysis Consortium, NCI

Richard Smith, Ph.D.

Pacific Northwest National Laboratory

Kellie Soafer

Duke University

Lori Sokoll, Ph.D.

Johns Hopkins University

Phillip Storm, M.D.

Children's Hospital of Philadelphia

James Suh, M.D.

Frederick National Laboratory for Cancer Research, NCI

Wei Sun, Ph.D.

Peking Union Medical College

Mathangi Thiagarajan, M.S.

Leidos/Center for Strategic Scientific Initiatives, NCI

Nicole Tignor, Ph.D.

Icahn School of Medicine at Mount Sinai

Chia-Feng Tsai, Ph.D.

Pacific Northwest National Laboratory

Shirley Tsang, Ph.D.

BioMatrix

Dana Valley

Van Andel Research Institute

Rodrigo Vargas Egue

Johns Hopkins University

Rajwanth Veluswamy, M.D.

Icahn School of Medicine at Mount Sinai

Alexander Voss

University of Kansas Cancer Center

Liang-Bo Wang

Washington University School of Medicine

Pei Wang, Ph.D.

Icahn School of Medicine at Mount Sinai

Ying Wang

New York University

Guanghai Wang, Ph.D.

National Institute of Standards and Technology

Bin Wang, M.D.

Peking University International Hospital

Bo Wen, Ph.D.

Baylor College of Medicine

Jeffrey Whiteaker, Ph.D.

Fred Hutchinson Cancer Research Center

Maciej Wiznerowicz, M.D., Ph.D.

International Institute for Molecular Oncology

Kamila Wiznerowicz

International Institute for Molecular Oncology

Matthew Wyczalkowski, Ph.D.

Washington University School of Medicine

Jingxuan Yang, Ph.D.

The University of Oklahoma Health Sciences Center

Seungyeul Yoo, Ph.D.

Ichan School of Medicine

Bing Zhang, Ph.D.

Baylor College of Medicine

Hui Zhang, Ph.D.

Johns Hopkins University

Zhen Zhang, Ph.D.

Johns Hopkins University

POSTERS

Topic	Team	Title	Poster #
Bioinformatic Tools: Data Analysis and Data Access	PGDAC: New York University/ Washington University/ Brigham Young University	Unified access to cancer proteogenomics data	1
		Data-driven discovery of phosphorylation modules in cancer	2
		PTMcosmos: A Web Portal of Post-Translational Modifications and Proteogenomic Resources in Cancer	3
		Predicting and Visualizing Mutations in Cancer Histopathology Images Using Deep Learning	4
	PGDAC: Baylor College of Medicine	Bioinformatics tools to integrate and understand molecular changes associated with oncogenic processes.	5
		OmicsEV: a tool for comprehensive evaluation of omics data processing methods	6
		PrecisionFDA NCI-CPTAC Multi-omics Enabled Sample Mislabeling Correction Challenge	7
	PGDAC: Mount Sinai	iProFun: An integrative analysis tool to screen for Proteogenomic Functional traits	8
		The CPTAC Network Exploration Portal and CPTAC Data Browser	9
	PGDAC: University of Michigan	Michigan Proteomics Pipeline	10
	PGDAC: Broad Institute	Interactive Co-Occurrence Mutation (iCoMut) Plots Provide a Comprehensive Visualization of Mixed Genomics and Proteomics Data	11
		Broad GDAC Terra/FireCloud Pipeline Brings Proven, Scalable and Reproducible Computational Methods to the Interpretive Analysis of Proteogenomics Data	12
		A fully automated, cloud-enabled pipeline for unsupervised clustering of multi-omics datasets	13
		Proteogenomic Tumor Analysis in the Cloud	14
	National Institute of Standards and Technology	In Depth Analysis of Peptide Modifications - Estimating Modification-specific FDR and Variation Across Samples	15
	Leidos	The Comprehensive Data Resource (CDR): An informatics system for complex, multi-omics research studies	16
	The Cancer Imaging Archive	The Cancer Imaging Archive (TCIA): CPTAC Radiology and Pathology Imaging Research Efforts	17
	PCC: Pacific Northwest National Laboratory	Evaluation of differential peptide loading on TMT-based proteomic and phosphoproteomic data quality	18
	DCC: Enterprise Science and Computing/Georgetown	"Proteomic Resources for the Cancer Research Community: The CPTAC Assay Portal, Data Portal, and Clinical Dashboard"	19
		Engaging the Protein Science Community to Expand Protein Literature Representation and Annotations in UniProt	20
Head & Neck Cancer	PGDAC: Baylor College of Medicine	Proteogenomic Characterization of HPV-negative Head and Neck Squamous Cell Carcinoma	21
Kidney Cancer	PCC: Johns Hopkins University	Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma	22

Topic	Team	Title	Poster #
Endometrial Cancer	PGDAC: Baylor College of Medicine	Proteogenomic Characterization of Endometrial Carcinoma	23
Pancreatic Cancer	PCC: Johns Hopkins University	Integrated Proteogenomic Characterization of Pancreatic Ductal Adenocarcinoma	24
AML	PTRC: Pacific Northwest National Laboratory/ Oregon Health and Science University	Global and Phosphoproteomic Analysis of AML Cell Line Response to Kinase Inhibitor Treatment	25
Ovarian Cancer	PTRC: Fred Hutchinson Cancer Center/University of Alabama	Proteogenomic approach to identify mechanisms of platinum refractoriness in high-grade serous ovarian cancers	26
Pediatric Brain Cancers	Tissue Source Site: Biomatrix/China	Urinary protein biomarker for pediatric medulloblastoma	27
	PGDAC: Mount Sinai	Integrated Proteogenomic Characterization of Pediatric Brain Tumors	28
Cancer Immunotherapy	PGDAC: Baylor College of Medicine	Cancer neoantigen prioritization through sensitive and reliable proteogenomics analysis	29
	PGDAC: Mount Sinai	Proteogenomics Characterization of Immune Infiltration and Immune Evasion in Tumor Tissue	30
Pancancer Studies	Tissue Source Site: Poland	Cancer Stemness Defined by CPTAC Proteogenomics Guides Identification of Protein Targets Towards Personalized Therapies for Cancer Patients	31
	PGDAC: New York University/ Washington University/ Brigham Young University	Mutational signatures and environmental factor influence across cancer types	32
		Aberrations in acetylation of metabolic proteins and histone complexes across cancer types	33
		DNA methylation instability and its functional impacts in human cancers	34
		LINE-1 activity in cancer is correlated across life-cycle stages and with interferon signatures	35
Targetted Proteomics	PTRC: Fred Hutchinson Cancer Center/University of Alabama	Multiplexed quantification of phosphorylation signaling dynamics in the DNA damage response using peptide immunoaffinity enrichment coupled to quantitative mass spectrometry	36
Technology Development	PGDAC: New York University/ Washington University/ Brigham Young University	Single-cell Omics and Proteomic Approaches to Study Pancreatic Ductal Adenocarcinoma Heterogeneity and Treatment Response	37
	Tissue Source Site: Biomatrix/China	Microfluidic system of zebrafish PDX models facilitates fast-track regimen decision for pancreatic cancer	38
	PTRC: Broad Institute & Baylor College of Medicine	Microscaled Proteogenomic Methods for Precision Oncology	39
	PCC: Pacific Northwest National Laboratory	Comprehensive nanoproteomics analysis enabled by an intelligent Boosting to Amplify Signal with Isobaric Labeling (iBASIL) strategy	40
		Carrier-assisted single-tube processing approach for targeted proteomics analysis of small number of cells	41
		An integrative analysis of tumor proteomic and phosphoproteomic profiles to examine the relationships between kinase activity and phosphorylation	42

ABSTRACTS

Bioinformatic Tools: Data Analysis and Data Access

PGDAC: New York University/Washington University in St. Louis/Brigham Young University:

1 Unified Access to Cancer Proteogenomics Data

Caleb M. Lindgren, Hannah Boekweg, David W. Adams, Sadie Taylor, Samuel H. Payne

Biology Department, Brigham Young University, Provo UT

Cancer data has many audiences. Although a small number of scientists may be involved in generating the clinical and molecular characterization of cancer cohorts, the data is a national resource that should be broadly distributed to facilitate analysis by a wider community of researchers, data scientists and even the lay public. Therefore, expanding access to data is an important goal for publicly funded research.

To seamlessly enable analysis, data must be distributed in a convenient and accessible manner. Although storing data in supplemental tables or cloud-based archives is fine for historical records, it is not the optimal dissemination method for active collaborations or ongoing analyses; researchers need to be able to access data in their analytical software via APIs. No convenient dissemination method currently exists for the quantitative molecular data tables that are the primary input of data interpretation and analysis codes.

We present an example of a unified API for accessing proteogenomic data from CPTAC cancer cohorts, a model that could be adopted by NCI for childhood cancer data. Each CPTAC dataset contains comprehensive genomics, transcriptomics, proteomics, and clinical data for a tumor-specific cohort - e.g., ovarian, endometrial, colon, etc. This data is contained within a Python package, *cptac*, that is freely distributed through the Python Package Index (PyPI). Our package removes many common barriers to analysis by automating all data loading and formatting to make the data ready for statistical and visual analytics. Additionally, the package handles complex merging between data tables and includes common algorithms for analyses. The module contains extensive tutorials and documentation to assist users in understanding the data and analysis methodology.

2 Data-Driven Discovery of Phosphorylation Modules in Cancer

Lili Blumenberg^{1,2}, Vlad Sviderskiy^{3,4}, Richard Possemato^{3,4}, Kelly Ruggles^{1,2}

¹Sackler Institute, Department of Medicine, ²Institute for Systems Genetics, ³Laura & Isaac Perlmutter Cancer Center, ⁴Department of Pathology, New York University School of Medicine, New York, NY

Nearly 40% of people will be diagnosed with cancer in their lifetime. Although many cancer-specific therapies have been developed, tumors with different molecular characteristics typically require personalized treatment strategies, even within a specific cancer subtype. Phosphoproteomics provides a particularly useful perspective on kinase signalling and potential vulnerabilities in cancer that can provide personalized information on targetable treatment pathways. Kinases in particular are eminently targetable and represent some of the most successful personalized cancer therapeutics developed to date, including ABL kinase in acute lymphoblastic leukemia, BRAF for melanoma and HER2 in breast cancer. However, discerning comprehensive phosphorylation signaling pathway activity and their respective signaling cascade events is challenging, especially since these pathways are often significantly modified in cancer. Additionally, genome-wide studies are difficult for clinicians to interpret, due to their high-dimensionality and co-linearity. To address these challenges, we developed an approach to determine co-regulated phosphorylation modules that are relevant in a pan-cancer and disease-specific populations. We applied this approach to phosphoproteomics data from a breast cancer cohort. This approach has identified canonical as well as new putative phosphorylation site modules. Following module identification, we nominated regulators of each module, by determining associations between the modules and

activity of kinases and phosphatases. In addition, we determined how these modules were regulated in prognostically important sample groups. Finally, we used these modules to determine which signaling pathways were active in each patient sample, a potential strategy for personalized treatment design.

3 PTMcosmos: A Web Portal of Post-Translational Modifications and Proteogenomic Resources in Cancer

Liang-Bo Wang^{1,2}, Daniel Cui Zhou^{1,2}, Li Ding^{1,2}

¹Department of Medicine and Genetics, Siteman Cancer Center, ²McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO

PTMcosmos is a comprehensive database with an interactive web portal designed to catalog and visualize post-translational modifications (PTMs) in humans. PTMcosmos currently stores 438,983 experimentally validated PTM sites, consisting of phosphorylation, acetylation, methylation, glycosylation, and ubiquitination sites. PTM sites are retrieved from the UniProt Knowledge Base, PhosphoSitePlus, and CPTAC phosphorylation and acetylation mass spectrometry data. We collect supporting evidence such as publications, experimental validation, and manual curations in order to provide a confidence metric, which is regularly maintained and updated. PTMcosmos hosts the entire spectrum of CPTAC experimental data, which contains somatic and germline variants, RNA sequencing, and protein and PTM mass spectrometry data. Additionally, PTMcosmos also contains TCGA and COSMIC variants, thus allowing for the collective integration and analysis of different data types. All the data is preprocessed from multiple sources in a fully automatic and reproducible manner. Complex ID conversions between Ensembl IDs, RefSeq IDs, and UniProt IDs are resolved based on sequence identity using UniParc. We have built an ensemble of interactive visualization tools and analysis modules that will allow investigators to investigate altered PTM functions directly from quantitative data (e.g., linear proximity to mutations).

The database is live at <https://ptmcosmos.wustl.edu> and is user and password protected. We will continue to develop PTMcosmos to better serve the CPTAC consortium and the broader research community. We will incorporate new CPTAC3 data as it becomes available, as well as other data, such as TCGA RPPA data and natural human variants from ExAC and gnomAD. In addition, we plan on implementing more features, such as visualizing changes in protein levels based on mutation status, among others, that will allow investigators to explore the data directly on PTMcosmos. We envision that PTMcosmos will serve both CPTAC and the wider research community to better understand the role of PTMs in cancer.

4 Predicting and Visualizing Mutations in Cancer Histopathology Images Using Deep Learning

Runyu Hong¹, Wenke Liu¹, Kay Li², Michael Gillette³, David Fenyö¹

¹New York University, New York, NY; ²Johns Hopkins University, Baltimore, MD; ³Broad Institute, Cambridge, MA

Deep-learning-based methods, in particular convolutional neural networks (CNNs), have shown outstanding performance in a broad range of computer vision tasks including ones in computation pathology. Here, we present a histopathology image analysis workflow, which shows the capability of predicting mutations and visualizing corresponding morphological features. Images first went through the quality control and color normalization process before tiled into 299 by 299 pixels pieces. Data were then separated into training, validation, and testing sets at per patient level. Inception and ResNet-architected CNN models were trained and the performance of the prediction was evaluated by area under ROC, area under PRC, and Top-1 accuracy. The activations and weights before the last layer of the model were output for visualization. Class activation mapping was used to visualize the attention of the model by projecting the weights back to the input tile. In addition, we made tSNE plots of these activations of sampled tiles. The clusters on these plots revealed some interesting morphological features that correlated with certain mutations, many of which correspond to other molecular level analyses. So far, the workflow has been applied to endometrial cancer and non-small-cell lung cancer samples in the CPTAC cohort and achieved decent performance for many critical mutations in both cancer types. For instance, an InceptionV3-architected model showed a per slide AUROC of 0.961. In the visualization tSNE plot, the positively predicted clusters showed that STK11 mutation led to less immune response in non-small-cell lung cancer samples. An implementation pipeline was also designed, which could potentially assist pathologists making quick and accurate decisions.

PGDAC: Baylor College of Medicine

5 Bioinformatics Tools to Integrate and Understand Molecular Changes Associated with Oncogenic Processes

Antonio Colaprico¹, Gabriel J. Odom¹, Yuguang Ban², Lizhong Liu¹, Xiaodian Sun², Alexander R. Pico³, Bing Zhang⁴, Lily Wang^{1*}, Xi Chen^{1*}

¹Department of Public Health Sciences, ²Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL; ³Institute for Data Science and Biotechnology, Gladstone Institute, San Francisco, CA; ⁴Department of Molecular and Human Genetics, Baylor College of Medicine, Houston

*senior co-authors

Recently, the Pan-Cancer Atlas initiative of The Cancer Genome Atlas (TCGA) created a comprehensive collection of 27 studies covering 11,000 patient tumors from 33 cancer types. These studies investigated cancer complexity from different angles and integrating multi-omics and clinical data. The computational tools developed from these studies have led to the identification of 299 cancer-driver genes and over 3,400 driver mutations. While these methods have uncovered the complexity of human cancers, it remains critical to clarify the functional consequences of these alterations.

In order to deal with the challenges of data retrieval and integration, TCGAbiolinks and DeepBlueR were developed to retrieve multi-omics data from several consortiums. Tumor-specific cancer-driver-gene events and downstream impact can be elucidated with MoonlightR by integrating those datasets. Moonlight distills literature findings, pathways, and multiple -omics data sources into a comprehensive assessment of a gene's roles and functions. The versatility of the TCGAbiolinks and MoonlightR tools has been illustrated across diverse studies, including oncogenic processes identification, oncogenic clinically actionable driver genes discovery and comprehensive immune landscape characterization.

However, most currently available pathway analysis software provide little or no functionalities for analyzing multiple types of -omics data simultaneously. In addition, most tools do not provide sample-specific estimates of pathway activities which are important for precision medicine. To address these challenges we developed pathwayPCA for integrative pathway analysis that utilized modern statistical methodology including supervised PCA and adaptive elastic-net PCA for principal component analysis. We applied MoonlightR and pathwayPCA to multiple TCGA and CPTAC tumors to illustrate pathway analysis with identification of cancer-driver-genes and estimating sample-specific pathway activities. We expect these tools to be useful for empowering the wide scientific community on the analyses and interpretation of the wealth of multi-omics data recently made available by TCGA, CPTAC and other large consortiums.

6 OmicsEV: A Tool for Comprehensive Evaluation of Omics Data Processing Methods

Bo Wen^{1,2}, Kai Li^{1,2}, Bing Zhang^{1,2,#}

¹Lester and Sue Smith Breast Center, ²Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX

Correspondence should be addressed to B.Z. (bing.zhang@bcm.edu)

High-throughput technologies such as RNA-Seq and mass spectrometry generate vast amount of transcriptomic and proteomic data that contain valuable quantification information for genes and proteins. Many algorithms, software, and pipelines have been developed to analyze these data. However, how to select optimal algorithms, software and parameters for analyzing a specific large-scale omics dataset remains a significant challenge. In this study, we developed an R package named OmicsEV, which allows users to compare and evaluate different data matrices generated from the same omics dataset using different tools, algorithms, or parameter settings. We have implemented more than 20 evaluation metrics and all the evaluation results are included in an HTML-report, which facilitates the identification of the optimal analysis method for the omics dataset of under investigation. We demonstrated the utility of OmicsEV on three different large-scale datasets, including both proteomics and RNA-Seq datasets. OmicsEV is easy to install and easy to use. Only one main function is needed to perform the whole evaluation process. The source code can be downloaded at <https://github.com/bzhanglab/OmicsEV>.

7 PrecisionFDA NCI-CPTAC Multi-omics Enabled Sample Mislabeling Correction Challenge

Seungyeul Yoo^{1,#}, Zhiao Shi^{2,#}, Bo Wen^{2,#}, Zeke Maier³, Jun Zhu¹, Weiping Ma¹, Elaine Johanson⁴, Emily Boja⁵, Henry Rodriguez⁵, Pei Wang^{1,*}, Bing Zhang^{2,*}

¹Icahn School of Medicine at Mount Sinai, New York, NY; ²Baylor College of Medicine, Houston, TX; ³Booz Allen, McLean, VA; ⁴U.S. Food and Drug Administration, Washington, DC; ⁵Office of Cancer Clinical Proteomics Research, National Cancer Institute, National Institutes of Health, Bethesda, MD

Equal contribution; * Co-corresponding (Contacts: pei.wang@mssm.edu, bing.zhang@bcm.edu)

Simultaneous use of multi-omics platforms to characterize a large set of biological samples, as utilized in National Cancer Institute (NCI)'s The Cancer Genome Atlas (TCGA) and the Clinical Proteomic Tumor Analysis Consortium (CPTAC) projects, has been demonstrated as a powerful approach to understanding the molecular basis of cancer and speeding the translation of new discoveries to patient care. However, sample or data mislabeling due to human errors is known to be one of the obstacles in these large scale omics studies. Such errors, which sometimes are unavoidable, contribute to irreproducible results and invalid conclusions. But leveraging the rich information of multi-omics data of one sample, it is possible to pinpoint whether and where the errors have occurred. To seek for an optimal solution, in September 2018, the precisionFDA and NCI-CPTAC jointly launched the Multi-omics Enabled Sample Mislabeling Correction Challenge to call on the scientific community (wisdom of the crowd) to tackle the problem. The 3-month long challenge comprises two sub-challenges. In the first one, participants were presented with clinical and proteomics data for the same set of samples and asked to detect samples with labelling errors. In the second one, participants were further presented with RNA-Seq data from the same samples and asked to identify mislabeled samples, and suggest the correct sample labels. The challenge was one of the most popular precisionFDA challenges with 153 and 84 submitted entries for each sub-challenge, respectively. Currently, the organizing committee with members from the FDA and NCI-CPTAC is teaming up with top performing participants to benchmark submitted methods and derive an integrated solution for sample mislabeling correction. Successful completion of the method development will provide a useful open-source tool to the scientific community to accurately detect and correct mislabeled samples using rich multi-omics datasets.

PGDAC: Mount Sinai

8 iProFun: An integrative Analysis Tool to Screen for Proteogenomic Functional Traits

Jiayi Ji^{*}, Xiaoyu Song^{1,2}, Kevin J. Gleason³, Fan Yang⁴, John A. Martignetti⁵, Lin S. Chen³, Pei Wang⁵

¹Department of Population Health Science and Policy, ²The Tisch Cancer Institute, ⁵Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; ³Department of Public Health Sciences, The University of Chicago, Chicago, IL; ⁴Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado, Denver, CO; *Presenter

In this work, we propose iProFun, an integrative analysis tool to screen for *Proteogenomic Functional* traits perturbed by DNA copy number alterations (CNAs) and methylations. The goal is to characterize functional consequences of DNA copy number and methylation alterations in tumors and to facilitate screening for cancer drivers contributing to tumor initiation and progression. Specifically, we consider three functional molecular quantitative traits: mRNA expression levels, global protein abundances, and phosphoprotein abundances. We aim to identify those genes whose CNAs and/or DNA methylations have cis-associations with either some or all three types of molecular traits. In comparison with analyzing each molecular trait separately, the joint modeling of multi-omics data enjoys several benefits: iProFun experienced enhanced power for detecting significant cis-associations shared across different omics data types; and it also achieved better accuracy in inferring cis-associations unique to certain type(s) of molecular trait(s). For example, unique associations of CNAs/methylations to global/phospho protein abundances may imply post-translational regulations.

We applied iProFun to ovarian high-grade serous carcinoma tumor data from The Cancer Genome Atlas and Clinical Proteomic Tumor Analysis Consortium, and identified CNAs and methylations of 500 and 121 genes, respectively, affecting the cis-functional molecular quantitative traits of the corresponding genes. We observed substantial power gain via the joint analysis of iProFun.

For example, iProFun identified 117 genes whose CNAs were associated with phosphoprotein abundances by leveraging mRNA expression levels and global protein abundances. By comparison, analyses based on phosphoprotein data alone identified none. A network analysis of these 117 genes revealed the known oncogene AKT1 as a key hub node interacting with many of the rest. These and other genes identified by iProFun could serve as potential drug targets for ovarian cancer.

9 The CPTAC Network Exploration Portal and CPTAC Data Browser

Anna Calinawan^{1,2,*}, Selim Kalayci^{1,2,*}, Saravana Mohan Dhanasekaran³, Francesca Petralia^{1,2}, Pei Wang^{1,2}, Boris Reva^{1,2,*}, Zeynep H. Gümüş^{1,2,*}

¹Department of Genetics and Genomic Sciences, ²Icahn Institute for Data Science and Genomic Technology, Icahn School of Medicine at Mount Sinai, New York, NY, ³University of Michigan, Ann Arbor, MI

*denotes equal co-authorship; Contact email: zeynep.gumus@mssm.edu, anna.calinawan@mssm.edu

We have developed the interactive visual data exploration portals *CPTAC Data Browser* and *CPTAC Network Exploration Portal*. These portals will greatly enhance the sharing of large-scale proteomics resources by allowing researchers to intuitively browse, query, and download data and analysis results from CPTAC projects.

CPTAC Network Exploration Portal

CPTAC Data Analysis centers are increasingly utilizing network analyses to interpret the massive, multi-scale data that are being generated by the consortium. While there are several network exploration tools available in the field, to share and explore these integrated proteomic datasets and their associated metadata in a user-intuitive way, we need a simple, unified and custom interface that integrates multiple network data types.

To address this need, we are developing the CPTAC Network Exploration Portal. Users can interactively explore (i) complex CPTAC networks and (ii) modules within these networks in 3D with a standard web browser. Customized functionalities include the capability to (i) display peptides and clinical variables associated with each gene; (ii) filter the networks based on phenotype; and (iii) highlight pathways enriched within network. Furthermore, network views can be customized to match specific data features. The portal is available at <http://ccrcc.cptac-network-view.org/>.

CPTAC Data Browser

This web application provides comprehensive and granular visualization of CPTAC data as interactive heatmaps. The browser has been implemented for three tumor types, including kidney cancer, lung adenocarcinoma, and pediatric brain tumors. While each portal has been tailored to the needs of the underlying data, they serve the same basic function: enabling users to specify genes of interest and generate an interactive visualization.

The portal allows users to render and explore heatmaps with proteome, phosphoproteome, transcriptome, and methylome data. When users find associations they consider meaningful, they can both export the image and download the data set in exactly the sample order that produced the pattern of interest.

We will demonstrate findings regarding a novel translocation event in the clear cell Renal Cell Carcinoma study that was not published as a figure but can be readily visualized with the CPTAC Data Browser on <http://ccrcc.cptac-data-view.org>.

PGDAC: University of Michigan

10 Michigan Proteomics Pipeline

Felipe da Veiga Leprevost, Hui-Yin Chang, Avinash Kumar Shanmugam, Dattatreya Mellacheruvu, Guo Ci Teo, Dmitry M. Avtonomov, Fengchao Yu, Daniel Geiszler, Andy Kong, Alexey I. Nesvizhskii

University of Michigan Medical School, Ann Arbor, MI

We introduce the Michigan Proteomics Pipeline, a powerful solution for shotgun proteomics data analysis. By combining high-performance custom algorithms with third-party software, we provide a complete solution designed for the analysis of open (mass-tolerant) or closed (narrow mass tolerance) database searches. MSFragger provides ultrafast and comprehensive peptide identification using advanced ion indexing technology. The software is capable of executing both closed and open searches in minutes, and recent updates include support for vendor raw formats and the ability to automatically calibrate masses and optimize search parameters without compromising its speed and sensitivity. Post processing is done by Philosopher, a powerful, cross-platform, and installation-free software that provides easy access to a customizable selection of algorithms and third-party software necessary for both traditional and mass-tolerant LC-MS/MS analysis. Peptide validation and protein inference are performed using stand-alone versions of PeptideProphet and ProteinProphet, respectively. False discovery rate estimation is done simultaneously at PSM and protein levels. Peptides and proteins are quantified using both label-free and TMT labels. Individual data sets were unified and normalized using TMT-Integrator, which extracts and combines channel abundances from each TMT data set. PSM tables are integrated at the gene, peptide, phosphopeptide, and protein levels using normalization methods like MD (median centering) or GN (median centering with variance scaling), followed by the characterization of PTMs using PTM-Shepherd. The Michigan Proteomics Pipeline was successfully applied to the large-scale CCRCC (clear renal cell carcinoma) proteogenomic study, detecting and quantifying a wide set of genes and proteins, making it possible to understand the functional impact of their expression levels on the analyzed tissues.

PGDAC: Broad Institute

11 Interactive Co-Occurrence Mutation (iCoMut) Plots Provide a Comprehensive Visualization of Mixed Genomics and Proteomics Data

David Heiman¹, Katherine Huang¹, Karsten Krug¹, Karl Clauser¹, D. R. Mani¹, Gad Getz^{1,2,3}

¹Broad Institute, Cambridge, MA; ²Massachusetts General Hospital Cancer Center and Department of Pathology, Boston, MA; ³Harvard Medical School, Cambridge, MA

Modern analyses of cancer datasets typically involve dozens of large and diverse tables of data, the result of running complex analyses on many samples. Mutation Co-Occurrence (CoMut) plots, introduced by Nico Stransky*, have become a common method of summarizing these data. By aligning results across a common X axis of samples, a single graphic provides a comprehensive analysis profile of a cancer cohort, enabling identification of relationships between co-occurring results. However, these plots are static, and can be difficult to construct and reproduce. We have developed iCoMut, a javascript web application that produces interactive plots of co-occurring and mutually exclusive results, improving both analysis turnaround time and reproducibility. iCoMut supports a variety of customization and data manipulation options, and can be deployed for specific CPTAC data working groups. We have integrated several proteomics tracks into the iCoMut plots. To date, iCoMut has been used to generate a principal figure for multiple TCGA papers:

1. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas *Cell* (2017). Volume 171 Issue 4: p950-965
2. Integrated Molecular Characterization of Testicular Germ Cell Tumors *Cell Rep* (2018). Volume 23 Issue 11: p3392-3406

3. The Integrated Genomic Landscape of Thymic Epithelial Tumors *Cancer Cell* (2018). Volume 33 Issue 2: p244-258
4. Integrative Molecular Characterization of Malignant Pleural Mesothelioma *Cancer Discovery* (2018). Volume 8 Issue 12: p1548-1565

* Stransky, Nicolas, et al. "The mutational landscape of head and neck squamous cell carcinoma." *Science* 333.6046 (2011): 1157-1160

12 Broad GDAC Terra/FireCloud Pipeline Brings Proven, Scalable and Reproducible Computational Methods to the Interpretive Analysis of Proteogenomics Data

David Heiman¹, Yifat Geffen¹, Julian Hess¹, Liudmila Elagina¹, Chet Birger¹, Gad Getz^{1,2,3}

¹Broad Institute, Cambridge, MA; ²Massachusetts General Hospital Cancer Center and Department of Pathology, Boston, MA; ³Harvard Medical School, Cambridge, MA

The Broad Institute GDAC addresses three key challenges facing the cancer genomics research community: first, the technical challenge that nearly all cancer genomic data are noisy and complex, requiring an extensive set of computational procedures to render it suitable for meaningful downstream analyses; second, the biological challenge that cancer exhibits enormous genetic complexity, with any given tumor harboring a handful of cancer-causing genomic aberrations ("drivers") and 10,000s of innocent bystander mutations ("passengers") that have no oncogenic potential. Developing insights from cancer genome data thus requires not only an analytical framework that distinguishes drivers from passengers, but also downstream validations which in turn will enhance and refine the analyses – an iterative process that begins with easy access to credible data and sophisticated tools for the broader cancer community. Which leads to the third, a community challenge: for our research to be truly transformative it must be accessible, interpretable and plausible to the entire cancer community—not just to the principal investigators or those with specialized genomic or computational expertise.

The Broad GDAC pipeline ran on TCGA data for 5 years, during which our group ran pipelined analyses at unprecedented scale and complexity. At the height of sample characterization in TCGA we saw as many as 24K new data aliquots ingested per year, with as many as 6K pipelines per month executed upon them. In 2013 alone, we published 62 Firehose runs (>1 per week), and at the end were executing 1500 pipelines per analysis run upon ~80K data aliquots spread over 38 disease cohorts and 10 data modalities.

We have migrated the GDAC pipeline to Terra, a cloud-based run-execution framework, where we continue to refine it as new tools and technologies become available. The pipelines is now being used in several CPTAC Data Working Groups using their own customized data.

13 A Fully Automated, Cloud-Enabled Pipeline for Unsupervised Clustering of Multi-Omics Datasets

Karsten Krug, D. R. Mani

Broad Institute, Cambridge, MA

Non-negative matrix factorization (NMF) has emerged as a powerful method to find latent structure in large 'omics datasets by decomposing a non-negative data matrix V into two matrices W and H such that $V = W \cdot H$, given a factorization rank K which corresponds to the number of clusters. While matrix H provides information about the clustering of samples (pattern matrix), matrix W contains information about the contribution of each feature (e.g. protein, gene) to a particular cluster (loadings). Thus, in a single analysis, NMF can provide a clustering of sample into K groups as well as information about which features are most characteristic for each cluster.

Here we describe a cloud-based, fully automated analysis pipeline to perform integrative, unsupervised cluster analysis of multi-omics datasets (e.g., proteome, phosphoproteome, RNA expression) using NMF. Several methods for determining the

factorization rank K have been implemented and are used to automatically determine the optimal number of clusters. Cluster membership scores reflecting the relative contribution of samples to clusters are used to determine a set of core samples that are most representative for each cluster. Each cluster is characterized by over-representation analysis of clinical variables, and determination of the most representative proteogenomics features. All results are presented in publication ready vector graphics for which users can fine-tune color schemes for clinical variable used as annotation tracks. The pipeline is available on DockerHub and on Broad's Terra platform running on Google Cloud infrastructure.

We applied our pipeline to various multi-omics cancer datasets including Lung Adenocarcinoma (CPTAC-LUAD), prospective breast cancer (CPTAC-prospBRCA), Uterine Corpus Endometrial Carcinoma (CPTAC-UCEC) and Medulloblastoma to find meaningful clusters and identify characteristic features discriminating clusters in a fully automated fashion. This pipeline is being integrated into the Broad's ProteoGenomic Data Analysis Pipeline in the Cloud (PGDAC) to facilitate routine application to all proteogenomic datasets.

14 Proteogenomic Tumor Analysis in the Cloud

Ramani Kothadia¹, D. R. Mani¹, Michael Noble¹, Karsten Krug¹, David Heiman¹, Karl R. Clauser¹, Gad Getz^{1,2,3}, Steven A. Carr¹

¹Broad Institute, Cambridge, MA; ²Massachusetts General Hospital Cancer Center and Department of Pathology, Boston, MA; ³Harvard Medical School, Cambridge, MA

Recent technological advances in Next Generation Sequencing and Mass Spectrometry-based proteomics have been instrumental in facilitating Proteogenomics—the integrative analysis of genomic, transcriptomic, proteomic and post-translational modification (PTM) data. We have implemented PGDAC (ProteoGenomic Data Analysis in the Cloud) for applying state-of-the-art algorithms and transforming proteogenomic data into biologically meaningful and interpretable results. PGDAC provides a simple interface to deploy and run a range of algorithms by leveraging Terra—a cloud-based platform developed at the Broad Institute for extreme-scale genomic analysis and data sharing—to implement proteogenomic pipelines. The platform is designed to be flexible, automated, reproducible, scalable, secure, and sharable. The cloud-based architecture is inherently scalable and supports parallel execution, thereby reducing processing time.

A wide range of proteogenomic analyses, many based on recently published landmark studies, have been implemented in PGDAC including:

- Basic data pre-processing such as normalization, filtering and sample;
- Correlation analysis of RNA, CNA and proteomics data;
- Identification of candidate driver genes using the concordance of cis- and trans-regulated protein expression with gene knockdown mRNA profiles in the Connectivity Map (CMAP);
- Association analysis for identification of markers for subgroups of interest, with gene set enrichment analysis (GSEA) to characterize these markers using pathway enrichment;
- Consensus and multiomic clustering for unsupervised identification of subgroups; and
- PTM signature enrichment analysis using PTM site-specific pathway databases.

Modules developed by collaborators, such as outlier analysis which identifies samples with aberrantly expressed kinases, can be easily added. In addition, G-TEX's RNAseq characterization pipeline, and MS philosopher for processing raw LC-MS/MS data are being currently integrated to provide raw-data analysis capabilities.

PGDAC has been used to analyze proteogenomic profiles from multiple cancers including breast cancer (BRCA), medulloblastoma, uterine cancer (UCEC), kidney cancer (CCRCC), and lung adenocarcinoma (LUAD). PGDAC will ultimately be used for integrative pan-cancer analysis across cancer cohorts.

National Institute of Standards and Technology

15 In Depth Analysis of Peptide Modifications - Estimating Modification-specific FDR and Variation Across Samples

Meghan C. Burke, Guanghui Wang, Stephen E. Stein

Mass Spectrometry Data Center, National Institute of Standards and Technology, Gaithersburg, MD

We present a method for false discovery rate (FDR) estimation of identifications made by the hybrid mass spectral library search (Burke et al, JPR, 2017), comparable to a “blind” or “open” modification mass spectral library search, based on an extension of the target-decoy approach. Estimation of the FDR for hybrid search results allows identifications to be compared to or used in conjunction with alternative methods for peptide identification. Results obtained from using 100 concatenated random decoy mass spectral libraries illustrate that a small number of preferred decoy DeltaMass, or mass difference between a search and library peptide, values exist and that FDR can be estimated for any DeltaMass value.

NIST MS Metrics has also been extended to (1) identify single amino acid substitutions and (2) channel-specific variation due to sample processing. The amino acid substitutions, identified based on a precisely localized modification and DeltaMass value corresponding to a substitution, have been analyzed to identify potential sources of false positive identifications, e.g., deamidation and artifactual methylation, as well as biologically relevant genetically variant peptides. Furthermore, the distribution of amino acid substitutions can be compared between analyses. The second recent advancement is computation of variation of individual channels for select modifications based on results obtained from the hybrid search. These results have identified semi-tryptic peptide formation as a significant source of sample-specific variation. Potential sources of semi-tryptic peptide formation, included trypsin mis-cleavage and protein fragmentation, have been further investigated.

Leidos

16 The Comprehensive Data Resource (CDR): An Informatics System for Complex, Multi-Omics Research Studies

Liqun Qi

National Cancer Institute, National Institutes of Health, Bethesda, MD

The CPTAC project collects biospecimens from 20 different human tumor types, including matched non-malignant tissue, and germline blood for comprehensive characterization by genomics, proteomics, and imaging technologies to better understand the molecular basis of cancer and help accelerate precision oncology. To facilitate CPTAC biospecimen collection and processing, the Comprehensive Data Resource (CDR), a web-based informatics system, was developed and initially released in 2016. The CDR streamlines the data entry workflow for candidate enrollment, specimen procurement, specimen evaluation, clinical data collection and candidate follow-up. The forms, data workflows, analytics, custom reports and tools implemented in the CDR are fully functional for 20 tumor types. The system is used by several CPTAC components including 32 Tissue Source Sites (TSS) spanning 13 countries, the Biospecimen Core Resource (BCR), the Pathology Resource Center (PRC), Data Managers (DM), Quality Managers (QM), Technical Project Managers (TPM), Data Coordinating Center (DCC) and NCI program managers. The automatic generation of entities from the CDR simplifies the process of uploading case and clinical data to the Genomic Data Commons (GDC). RESTful services available from the CDR enable the upload of case data in JSON format to the DCC.

The Cancer Imaging Archive

17 The Cancer Imaging Archive (TCIA): CPTAC Radiology and Pathology Imaging Research Efforts

Brenda Fevrier-Sullivan, Justin Kirby, Carl Jaffe, Kirk Smith, Tracy Nolan, Bill Bennett, Natasha Honomichl, Ashish Sharma, Ryan Birmingham, Fred Prior, John Freymann

National Cancer Institute, National Institutes of Health, Bethesda, MD

The Cancer Imaging Archive (TCIA) is the National Cancer Institute's (NCI) repository of publicly accessible cancer images and supporting metadata. TCIA supports the Clinical Proteomic Tumor Analysis Consortium (CPTAC) by hosting de-identified clinical radiological images, pathology images and follow-up clinical imaging to augment the consortium's proteogenomic analyses with imaging phenotypes. The data are organized into collections by cancer type. Each collection includes a descriptive summary, licensing details, multiple search and download interfaces, citations, version control, and links to supporting data. A help desk provides email and phone support along with user manuals and FAQs.

PCC: PNNL

18 Evaluation of Differential Peptide Loading on TMT-Based Proteomic and Phosphoproteomic Data Quality

Paul D. Piehowski, Yang Wang, Vladislav A. Petyuk, Joshua R. Hansen, Marina A. Gritsenko, Karl K. Weitz, Cristina Tognon, Wei-Jun Qian, Tao Liu, Brian J. Druker, Karin D. Rodland

Pacific Northwest National Laboratory, Richland, WA

Global and phosphoproteome profiling have demonstrated great potential utility for the analysis of clinical specimens. One important challenge facing the analyst carrying out these studies, is the large amount of sample required. Although dependent on tissue type, on the order of 25 mg wet tissue weight is currently required to obtain in-depth, global and phosphoproteome coverage. In the case of acute myeloid leukemia (AML) as studied here, the sample requirement is in excess of 10M white blood cells. Throughout the course of a prospective study, this requirement will certainly exceed what is obtainable from many of the individual patients/timepoints. For this reason, we were interested in examining the impact of differential peptide loading on proteomic data quality. To achieve this, we tested a range of channel loading amounts (20, 40, 100, 200, and 400 μ g; or approximately 5E5, 1E6, 2.5E6, 5E6, and 1E7 AML patient cells) to assess proteome coverage, quantification reproducibility and accuracy. As expected, we found that fewer missing values are observed in TMT channels with higher loading amounts compared to those with a lower loading amount. Moreover, channels with lower loading amounts have greater quantitative variability than channels with higher loading amounts. Statistical analysis among the five loading groups showed that the 20 and 40 μ g loading groups, were significantly different from the 400 μ g loading groups. However, no significant differences were detected among the 100, 200 and 400 μ g loading groups after normalization. These analyses provide the basis for designing the optimal clinical proteomics study when specimen quantities are limited.

DCC: ESAC/Georgetown

19 Proteomic Resources for the Cancer Research Community: The CPTAC Assay Portal, Data Portal, and Clinical Dashboard

Toan Le¹, Shuang Cai¹, Yin Lu¹, Kristen Nyce¹, Marcin J. Domagalski¹, Alicia Francis¹, Ratna R. Thangudu¹, Nathan J. Edwards², Simina M. Boca², Peter B. McGarvey², Maya Zuhl¹, Padmini Chilappagari¹, Lei Ma¹, William FitzHugh¹, Michael Holck¹, Jeffrey R. Whiteaker³, Amanda G. Paulovich³, Mehdi Mesri⁴, Christopher R. Kinsinger⁴, Karen A. Ketchum¹

¹ESAC, Inc. Rockville, MD; ²Georgetown University, Washington, DC; ³Fred Hutchinson Cancer Research Center, Seattle, WA; ⁴National Cancer Institute, National Institutes of Health, Bethesda, MD

Proteogenomic data from the CPTAC program are available through a collection of integrated resources for the cancer research community. Here we describe 3 websites from the Data Coordinating Center (DCC) that provide investigators with access to (1) targeted mass spectrometry-based peptide assays, (2) global proteomic, phosphoproteomic and glycoproteomic tumor profiles, and (3) clinical descriptions of CPTAC cancer cohorts.

1. The CPTAC Assay Portal is a centralized public repository of “fit-for-purpose,” multiplexed proteomic targeted assays. Each of the 2375 assays includes experimental data showing a response curve to the peptide analyte and the repeatability of the protocol. All data is reviewed and must meet the quality standards documented by the CPTAC program (Whiteaker, 2014). Assays have been submitted by laboratories in the United States, Canada, and Asia. Presently, these data cover over 2200 unique peptides corresponding to more than 1300 proteins from human and mouse.

Website: <https://assays.cancer.gov/>

2. The Proteomic Data Portal provides access to global proteomic analyses of tumors originating from breast, ovarian, colon, brain, lung, and other cancers (Edwards, 2015). This data warehouse has over 20 TB of mass spectrometry files along with peptide-spectrum matches and protein summary reports, enabling investigators to obtain data at any level that is useful for their cancer research. Links from the Data Portal to the NCI Genomic Data Commons (GDC) and The Cancer Imaging Archive (TCIA) are provided for easy navigation to complementary genomic and image information.

Website: <https://proteomics.cancer.gov/data-portal>

3. Our Clinical Dashboard and API, two resources under development, describe the tumor collections in the CPTAC program and provide methods for users to query cancer cohorts based on the clinical attributes. The API endpoints are documented on a swagger board and facilitate automated connections, while the graphical dashboard enables browsing and stratification through an interactive web interface. Both resources connect tumors to their associated slides at TCIA.

API: <https://clinicalapi-cptac.esacinc.com/api/tcia/>

Clinical Dashboard: <https://clinicaldashboard.esacinc.com>

20 Engaging the Protein Science Community to Expand Protein Literature Representation and Annotations in UniProt

Peter McGarvey^{2,3}, Cecilia Arighi^{1,3,*}, Hongzhan Huang^{1,3}, Yongxing Chen^{1,3}, Qinghua Wang^{1,3}, Cathy Wu^{1,2,3}, UniProt Consortium³

¹Center for Bioinformatics and Computational Biology, University of Delaware, Newark, NJ; ²Department of Biochemistry and Molecular and Cellular Biology, Georgetown University Medical Center, Washington, DC; ³EBI-EMBL, UK; SIB, CH and PIR

* Correspondence should be addressed. Tel: +1 302-831-3444; Fax: +1 302-831-4841; Email: arighi@dbi.udel.edu

UniProt Knowledgebase is a publicly available database with access to a vast amount of protein sequence and function information. Expert curation at UniProt includes a critical review of experimental data from the literature and predicted data from a range of sequence analysis tools. A representative set of literature articles is selected for annotation. Thus, many articles with potentially

useful content may not be included. Also, UniProt expert curation focuses on selected species with proteins from many organisms not being actively annotated.

To facilitate access to more comprehensive literature related to entries, UniProt compiles and organizes publications from external biological databases and text mining results. For a better user experience, this bibliography is classified, via a neural network-based method or based on the source databases, into the different topics in the entry, similar to the curated references. These publications are available under “Computationally mapped” in the publication section of the protein entry.

Still, many experts request articles and annotations to be added to protein entries. To respond to this need, we are developing a “Community” section where researchers are able to add directly the articles that they deem relevant to an entry, along with performing optional annotation tasks. ORCIDs will be used to validate and give credit to the contributors.

With the community expert contributions, UniProt will enable access to a more comprehensive set of articles and annotations, enabling discovery and benefitting the wider protein science community.

Head & Neck Cancer

PGDAC: Baylor College of Medicine

21 **Proteogenomic Characterization of HPV-negative Head and Neck Squamous Cell Carcinoma** *CPTAC HNSCC Working Group* *Baylor College of Medicine, Houston, TX*

Currently, patients with head and neck squamous cell carcinoma (HNSCC) are treated with surgery, radiation, and chemotherapy. Few targeted therapies exist. Compared to human papillomavirus associated (HPV-positive) HNSCCs, conventional (HPV-negative) HNSCCs are characterized by worse treatment response and prognosis. Identifying new targets for HPV-negative HNSCCs represents an unmet need. We prospectively collected tumor specimens, matched normal adjacent tissues (NATs), and blood samples from 110 HNSCC patients for deep proteogenomic characterization. This cohort is dominated by tumors from oral cavity (45, 41.0%) and larynx (49, 44.5%), with only six cases (5.5%) from oropharyngeal sites, the most common sites of HPV-positive HNSCC. Accordingly, analysis of RNA-Seq data identified only one HPV-positive case among all tumors. Thus, the cohort provides a unique opportunity for a focused analysis of HPV-negative HNSCC. We have generated somatic mutation, somatic copy number alteration (SCNA), RNA expression, and data-independent acquisition (DIA) proteomics data. Tandem mass tag (TMT)-based global proteomics and phosphoproteomics data are being generated. Mutation signature analysis identified 67 tumors harboring the smoking signature, which was largely consistent with self-reported smoking status. Somatic mutation analysis recapitulated frequently mutated genes in HPV-negative HNSCC, including *TP53*, *NOTCH1*, *FAT1*, *KMT2D*, and *NSD1*. SCNA analysis identified amplifications in 3q, 5p and 8q and deletions in 3p, 5q, 8p and 11q. Unsupervised principal component analysis of RNA-Seq and DIA proteomics data clearly distinguished tumors from NATs, as well as tumors from oral cavity and larynx. Transcriptomic subtyping identified all four previously reported subtypes (atypical, classical, basal, and mesenchymal), with expected subtype-specific molecular and histological characteristics. We plan to perform proteogenomic integration when all data are available to prioritize genomic drivers and investigate their impact on the proteome and phosphoproteome, to functionally characterize different etiological, pathological, and molecular subtypes, and to identify new biomarkers and potential targets for precision treatment of HPV-negative HNSCC.

Kidney Cancer

PCC: Johns Hopkins University

22 Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma

David J. Clark^{1,32}, Saravana M. Dhanasekaran^{2,32}, Francesca Petralia^{3,32}, Jianbo Pan^{1,32}, Xiaoyu Song^{4,5,32}, Yingwei Hu^{1,32}, Felipe da Veiga Leprevost^{2,32}, Boris Reva^{3,32}, Tung-Shing M. Lih^{1,32}, Hui-Yin Chang², Weiping Ma³, Chen Huang⁶, Christopher J. Ricketts⁷, Lijun Chen¹, Azra Krek³, Yize Li⁸, Dmitry Rykunov³, Qing Kay Li¹, Lin S. Chen⁹, Umut Ozbek^{4,5}, Suhas Vasaikar¹⁰, Yige Wu⁸, Seungyeul Yoo³, Shrabanti Chowdhury³, Matthew A. Wyczalkowski⁸, Jiayi Ji³, Michael Schnaubelt¹, Andy Kong², Sunantha Sethuraman⁸, Dmitry M. Avtonomov², Minghui Ao¹, Antonio Colaprico¹¹, Song Cao⁸, Kyung-Cho Cho¹, Selim Kalayci³, Shiyong Ma¹, Wenke Liu¹², Kelly Ruggles¹³, Anna Calinawan³, Zeynep H. Gümüş³, Daniel Geizler¹⁴, Emily Kawaler¹², Guo Ci Teo², Bo Wen⁶, Yuping Zhang², Sarah Keegan¹², Kai Li⁶, Feng Chen¹⁵, Nathan Edwards¹⁶, Phillip M. Pierorazio¹⁷, Xi Steven Chen^{11,18}, Christian P. Pavlovich¹⁷, A. Ari Hakimi¹⁹, Gabriel Brominski²⁰, James J. Hsieh²¹, Andrzej Antczak²⁰, Tatiana Omelchenko²², Jan Lubinski²³, Maciej Wiznerowicz^{24,25}, W. Marston Linehan⁷, Christopher R. Kinsinger²⁶, Mathangi Thiagarajan²⁷, Emily S. Boja²⁶, Mehdi Mesri²⁶, Tara Hiltke²⁶, Ana I. Robles²⁶, Henry Rodriguez²⁶, Jiang Qian²⁸, David Fenyö¹², Bing Zhang^{6,29}, Li Ding⁸, Eric Schadt^{3,30}, Arul M. Chinnaiyan², Zhen Zhang¹, Gilbert S. Omenn^{14,31}, Marcin Cieslik², Daniel W. Chan¹, Alexey I. Nesvizhskii², Pei Wang³, Hui Zhang¹, Clinical Proteomic Tumor Analysis Consortium

¹Department of Pathology, ¹⁷Brady Urological Institute and Department of Urology, ²⁸Department of Ophthalmology, Johns Hopkins University, Baltimore, MD; ²Department of Pathology, ¹⁴Department of Computational Medicine and Bioinformatics, ³¹Department of Internal Medicine, Human Genetics, and School of Public Health, University of Michigan, Ann Arbor, MI; ³Department of Genetics and Genomic Sciences and Icahn Institute for Data Science and Genomic Technology, ⁴Department of Population Health Science and Policy, ⁵Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY; ⁶Lester and Sue Smith Breast Center, ²⁹Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; ⁷Urologic Oncology Branch, Center for Cancer Research, ²⁶Office of Cancer Clinical Proteomics Research, National Cancer Institute, National Institutes of Health, Bethesda, MD; ⁹Department of Public Health Sciences, University of Chicago, Chicago, IL; ¹⁰Department of Translational Molecular Pathology, MD Anderson Cancer Center, Houston, TX; ¹¹Department of Public Health Sciences, ¹⁸Sylvester Comprehensive Cancer Center, University of Miami Miller School of Medicine, Miami, FL; ¹²Institute for Systems Genetics and Department of Biochemistry and Molecular Pharmacology, ¹³Department of Medicine, New York University School of Medicine, New York, NY; ¹⁵Departments of Medicine and Cell Biology and Physiology, ²¹Department of Medicine, ⁸Washington University School of Medicine, St. Louis, MO; ¹⁶Department of Biochemistry and Cellular Biology, Georgetown University, Washington, DC; ¹⁹Department of Surgery, Urology Service, ²²Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY; ²⁰Department of Urology, ²⁵Poznań University for Medical Sciences, Poznań, Poland; ²³Department of Genetics and Pathomorphology, Pomeranian University of Medicine, Szczecin, Poland; ²⁴International Institute for Molecular Oncology, Poznań, Poland; ²⁷Frederick National Laboratory for Cancer Research, Frederick, MD; ³⁰Sema4, Stamford CT

*These authors contributed equally

Background: Clear cell renal cell carcinoma (ccRCC) is the most predominant histology of renal cancer, representing 75% of all cases and accounting for the majority of associated deaths. To gain insight into the impact of genomic alterations on the functional modules that drive ccRCC tumorigenesis, we leveraged comprehensive proteogenomic characterization of 110 treatment-naïve renal cell carcinoma (RCC) and 84 paired-matched normal adjacent tissue (NAT) samples.

Methods: We utilized an integrated proteogenomic approach, performing whole genome sequencing (WGS), whole exome sequencing (WES), DNA methylation profiling, for all tumors, while RNA-seq, proteomic, and phosphoproteomic characterization was performed for all samples.

Results: WGS analysis revealed arm-level loss of chromosome 3p as a frequent event in ccRCC, with 61% of tumors showing evidence of 3p chromosomal translocation events. Comparative profiling of ccRCC and NATs samples identified pathways associated with immune response, epithelial mesenchymal transition, and glycolysis to be up-regulated in ccRCC, while TCA cycle,

fatty acid metabolism, and oxidative phosphorylation (OXPHOS) were down-regulated. Examination of mRNA-protein correlation revealed a non-linear relationship in cellular processes including Warburg Effect-related metabolism, as well as the tumor-specific trend of higher sample-wise correlation associating with prognostically-defined aggressive features of ccRCC. Analysis of differential phosphosite occupancy between tumors and NAT showed MAPK/ERK signalling and G2/M stalling to be up-regulated across the majority of ccRCC cases. We deconvoluted immune and stromal cell gene signatures in the tumor microenvironment (TME), with consensus clustering of the TME compositions identifying four immune-based subtypes: Inflamed CD8+, Inflamed CD8-, VEGF Immune Desert, and Metabolic Immune Desert. Integrated transcriptomic and proteomic profiling of the ccRCC subtypes revealed unique, discriminatory signalling pathways associated with immune exhaustion, cancer-associated fibroblast-related signalling, angiogenesis, and metabolic activity.

Conclusions: Our results link the functional impact of genomic alterations at the protein level, and provides evidence for rational treatment selection stemming from proteomic, phosphoproteomic, and tumor microenvironment signatures.

Endometrial Cancer

PGDAC: Baylor College of Medicine

23 Proteogenomic Characterization of Endometrial Carcinoma

Yongchao Dou^{1, 2, 3*}, Emily A. Kawaler^{4, 5*}, Daniel Cui Zhou^{6, 7*}, Marina A. Gritsenko^{8*}, Chen Huang^{1, 2, 3}, Lili Blumenberg⁹, Alla Karpova^{6, 7}, Vladislav A. Petyuk⁸, Sara R. Savage^{1, 2, 3}, Shankha Satpathy¹⁰, Wenke Liu^{4, 5}, Yige Wu^{6, 7}, Chia-Feng Tsai⁸, Bo Wen^{1, 2, 3}, Zhi Li^{4, 5}, Song Cao^{6, 7}, Jamie Moon⁸, Zhiao Shi^{1, 2, 3}, MacIntosh Cornwell^{4, 5}, Matthew Wyczalkowski^{6, 7}, Rosalie K. Chu⁸, Suhas Vasaikar¹¹, Hua Zhou^{4, 5}, Qingsong Gao^{6, 7}, Ronald J. Moore⁸, Kai Li^{1, 2, 3}, Sunantha Sethuraman^{6, 7}, Matthew E. Monroe⁸, Rui Zhao⁸, David Heiman¹⁰, Karsten Krug¹⁰, Karl Clauser¹⁰, Ramani Kothadia¹⁰, Yosef Maruvka¹⁰, Alex Pico¹², Amanda E. Oliphant¹³, Emily L. Hoskins¹³, Samuel Pugh¹³, Sean J. I. Beecroft¹³, David W. Adams¹³, Jonathan C. Jarman¹³, Andy Kong¹⁴, Hui-Yin Chang¹⁴, Boris Reva¹⁵, Dmitry Rykunov¹⁵, Antonio Colaprico^{16, 17}, Steven Chen^{16, 17}, Tara Hiltke¹⁸, Christopher R. Kinsinger¹⁸, Mehdi Mesri¹⁸, Ana I. Robles¹⁸, Henry Rodriguez¹⁸, David Mutch¹⁹, Katherine Fuh¹⁹, Matthew Ellis^{1, 2, 3}, Deborah DeLair²⁰, Mathangi Thiagarajan²¹, D R Mani¹⁰, Gad Getz¹⁰, Michael Noble¹⁰, Alexey Nesvizhskii¹⁴, Pei Wang¹⁵, Matthew L. Anderson²², Douglas A. Levine²³, Richard D. Smith⁸, Samuel H. Payne¹³, Kelly V. Ruggles⁹, Karin D. Rodland^{8, 24#}, Li Ding^{6, 7#}, Bing Zhang^{1, 2, 3#}, Tao Liu^{8#}, David Fenyo^{4, 5#¶}, Clinical Proteomic Tumor Analysis Consortium

¹Lester and Sue Smith Breast Center, Baylor College of Medicine, Houston, TX; ²Department of Molecular and Human Genetics, ³Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX; ⁴Institute for Systems Genetics, ⁵Department of Biochemistry and Molecular Pharmacology, ⁹Department of Medicine, NYU School of Medicine, New York, NY; ⁶Department of Medicine and Genetics, Siteman Cancer Center, ⁷McDonnell Genome Institute, ¹⁹Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Washington University in St. Louis, St. Louis, MO; ⁸Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA; ¹⁰Broad Institute, Cambridge, MA; ¹¹Department of Translational Molecular Pathology, MD Anderson Cancer Center, Houston, TX; ¹²Institute of Data Science and Biotechnology, Gladstone Institutes, San Francisco, CA; ¹³Department of Biology, Brigham Young University, Provo, UT; ¹⁴Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI; ¹⁵Department of Genetics and Genomic Sciences, Icahn Institute of Genomics and Multiscale Biology, Icahn School of Medicine at Mount Sinai New York, NY; ¹⁶Sylvester Comprehensive Cancer Center, Miami, FL; ¹⁷Division of Biostatistics, Department of Public Health Science, University of Miami Miller School of Medicine, Miami, FL; ¹⁸Office of Cancer Clinical Proteomics Research, National Cancer Institute, Bethesda, MD; ²⁰Department of Pathology, ²³Gynecologic Oncology, Laura and Isaac Perlmutter Cancer Center, NYU Langone Health, New York, NY; ²¹Leidos Biomedical Research Inc., Frederick National Laboratory for Cancer Research, Frederick, MD; ²²College of Medicine Obstetrics & Gynecology, University of South Florida Health, Tampa, FL; ²⁴Department of Cell, Developmental, and Cancer Biology, Oregon Health & Science University, Portland, OR

*The authors contributed equally; #Correspondence should be addressed to: karin.rodland@pnpl.gov (K.R.), lding@wustl.edu (L.D.), bing.zhang@bcm.edu (B.Z.), tao.liu@pnpl.gov (T.L.), david@fenyolab.org (D.F.); ¶ Lead contact

We collected tumor and blood samples from 95 endometrial cancer (EC) patients, 49 of which have matched adjacent non-tumor tissues. These samples were analyzed using eight omics platforms: whole-genome sequencing, whole-exome sequencing, RNA-Seq, miRNA-Seq, MethylationEPIC array, and isobaric tandem mass tag (TMT)-labeled proteome, phosphoproteome, and acetylome quantification, the latter marking the first genome-wide acetylation survey of EC. When comparing serous to endometrioid tumors, we noted the highly elevated phosphorylation of TP53BP1-S1763 and CHEK2-S163 in serous tumors; elevated phosphorylation of these sites is associated with the DNA damage response (DDR). We further identified six hyperphosphorylated proteins that can be targeted by FDA-approved drugs in DDR-high samples. Integrative analysis of the proteome, phosphoproteome, acetylome, and somatic mutation data highlighted the *cis*-acting and *trans*-acting effects of *CTNNB1* and *TP53* hotspot mutations on protein abundance and of post-translational modifications (PTMs) on their protein complex partners and regulators. Moreover, we observed significant positive associations between *BRD3* and several histone H2B N-terminal acetylation sites. Integrating proteomics and somatic copy number alteration data suggests that genes encoded on chromosome 1q may inhibit p53 activity by repressing p53 pathway components and/or downstream target genes. Proteo-transcriptomic integration revealed a putative positive feedback loop between *QKI*, circRNAs, and miRNAs to promote epithelial-to-mesenchymal transition in EC. Proteogenomic investigation of the immune landscape identified eight recurrently overexpressed cancer/testis antigens; it also found putative neoantigens in 49.3% of tumors. We identified proteogenomic alterations in EC that may drive antigen processing and presentation defects, which are a key immune evasion mechanism in microsatellite instability (MSI)-high tumors. In summary, our study provides insights into EC-associated proteins and PTMs, new regulatory mechanisms and drug targets. Both raw and processed data have been made publicly accessible to the scientific community in order to facilitate new biological and therapeutic discoveries.

Pancreatic Cancer

PCC: Johns Hopkins University

24 Integrated Proteogenomic Characterization of Pancreatic Ductal Adenocarcinoma

Investigators from the Sol Goldman Pancreatic Cancer Research Center, Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD; Investigators from Lester and Sue Smith Breast Center and Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; Investigators from Department of Medicine and Genetics, McDonnell Genome Institute, Washington University School of Medicine, St. Louis, MO; CPTAC Investigators

Pancreatic Ductal Adenocarcinoma is a highly aggressive cancer with poor survival rates. Towards understanding the underlying molecular alterations that drive the PDAC oncogenesis as part of The Clinical Proteomic Tumor Analysis Consortium, we conducted the first comprehensive proteogenomic analysis of 76 pancreatic tumors and 44 adjacent normal tissues (NATs). Proteomic analysis identified and quantified a total of 6,770 proteins. In addition, whole genome sequencing, whole exome sequencing, methylation, and RNA-seq were performed on the same tumors and NATs to facilitate an integrated proteogenomic analysis and determine the impact of genomic alterations on protein expression. Genomic analysis showed that 89% of tumor samples have a detectable *KRAS* hotspot mutation, which is close to the mutation rate reported by TCGA (93%), suggesting adequate tumor purity percentages in most of our samples. Interestingly, we found that mutations did not have a strong *cis* impact on the overall proteome level. However, we observed significant increases in BAX and DNMT1 protein levels associated with *TP53* mutations. Tumor-normal analyses revealed tumor-specific gene and protein expression, including EMT, cell adhesion, and keratin genes. Using cell surface marker gene expression, we identified and classified our samples into four immune subtypes: NK- excluded, NK+ excluded, immune desert, and inflamed. PLEK, WAS, CD48, ALDOA, and ASPH had highly differential protein levels in inflamed tumors, among others. We further identified that protein levels of INTS14, SULF2, and PTPMT1, among others, were associated with patient survival, demonstrating the prognostic value of proteomics over RNA alone. Additionally, we are characterizing an additional 81 pancreatic tumors and 58 NATs to the cohort using our deep proteogenomic analysis workflow, aiming to verify these preliminary findings and discover new biology. This integrated proteogenomic characterization of PDAC will be a valuable resource for the community, paving the way for the discovery of novel therapeutic targets.

Acute Myeloid Leukemia

PTRC: Pacific Northwest National Laboratory/Oregon Health & Science University

25 Global and Phosphoproteomic Analysis of AML Cell Line Response to Kinase Inhibitor Treatment

Paul D. Piehowski¹, Kevin Watanabe-Smith², Ozgun Babar⁴, Jamie Moon¹, Joshua R. Hansen¹, Samantha L. Savage², Cristina E. Tognon², Anupriya Agarwal², Jeffrey W. Tyner², Brian J. Druker^{2,3}, Karin D. Rodland¹

¹Integrative Omics, Environmental and Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA; ²Hematology and Medical Oncology, Knight Cancer Institute, ⁴Computational Biology, Oregon Health & Science University, Portland, OR; ³Howard Hughes Medical Institute, Portland, OR

In the United States, approximately 21,000 people are diagnosed with AML each year and over 10,000 AML related deaths are reported. Genomic analysis has revealed at least 11 genetic classes of AML. Recent results from the Beat AML clinical trial revealed that response to drugs is associated with mutational status, including instances where drug sensitivity is specific to combinatorial mutational events. However, attempts to use genomic heterogeneity to customize drug therapies have been complicated by the complex pattern of mutations and a lack of pharmacological agents for most mutational events. To better understand the molecular mechanisms underpinning drug response, we performed global and phosphoproteomic analyses on 4 AML cell lines that carry mutations associated with AML: K562 with a BCR-ABL fusion, MOLM14 with a FLT3 mutation, HL60 with a RAS mutation, and CMK with a JAK3 mutation. These cell lines were each treated with a different kinase inhibitor drug for which they have a known sensitivity. The results of these experiments indicated that:

- Proteomics and phosphoproteomics is capable of differentiating between leukemia cell lines based on the driving mutation
- Substrate phosphorylation can be used to infer upstream kinase activity
- Proteomics and phosphoproteomics can be used to track the dynamic response of leukemia cell lines to drug treatment
 - Drug responses can be observed as early as 30 minutes after drug treatment
- Early time points appear to be dominated by down-regulation of pathways downstream of the drug target
- Later time points often reflect up-regulation of apoptosis-associated processes

In the remainder of our PTRC funding period, we will apply the same approach to patient samples in order to identify mechanisms associated with drug sensitivity.

Ovarian Cancer

PTRC: Fred Hutchinson Cancer Center/University of Alabama at Birmingham

26 Proteogenomic Approach to Identify Mechanisms of Platinum Refractoriness in High-Grade Serous Ovarian Cancers

Shrabanti Chowdhury*, Sara Savage*, Jacob J. Kennedy, Xiaonan Hou, Catherine J. Huntoon, Richard G. Ivey, Qing Yu, ChenWei Lin, Dongqing Huang, Lei Zhao, Julia Voytovich, Regine Schoenherr, Zahra Shire, Aura Burian, Steven Skates, Jeffrey Whiteaker, Andy Hoofnagle, Travis Lorentzen, Bing Zhang, Larry M. Karnitz, Saravut J. Werooha, Steven Gygi, Scott H. Kaufmann, Pei Wang, Michael J. Birrer, Amanda G. Paulovich

Fred Hutchinson Cancer Center, Seattle, WA; University of Alabama at Birmingham, Birmingham, AL; Icahn School of Medicine at Mount Sinai, New York, NY; Harvard Medical School, Cambridge, MA; University of Washington Seattle, Seattle, WA; Mayo Clinic, Rochester, MN; Baylor College of Medicine, Houston, TX; Massachusetts General Hospital, Boston, MA

E-mails: pei.wang@mssm.edu, mbirrer@uab.edu, apaulovi@fredhutch.org; *Equal contributor

High grade serous ovarian cancer (HGSOC) is frequently diagnosed at an advanced stage and is the most lethal gynecological malignancy. Although response rates to platinum-based chemotherapy is ~85%, unfortunately, 80-90% of initially responsive patients develop drug-resistant disease. Moreover, ~15% of patients have platinum-refractory disease at diagnosis. Unfortunately, there is a dearth of predictive biomarkers to identify the refractory patients up front, and hence, there are crucial clinical needs for methods to predict platinum refractoriness of HGSOC, and for novel treatments that can be used either alone or in combination with platinum compounds to overcome this resistance. Our NCI Clinical Proteomic Tumor Analysis Consortium (CPTAC)-funded approach combines genomic and proteomic ("proteogenomic") analyses of both preclinical cell line and PDX models (0, 8- & 24-hours post-platinum exposure) and treatment-naïve human tumors to understand the mechanisms and to identify the potential predictors of platinum resistance. Leveraging well characterized HGSOC intra-patient cell line pairs and PDX models before and after development of platinum resistance, along with treatment-naïve human tumors, we performed proteogenomic analysis to characterize molecular signatures of platinum response. In parallel, we curated candidate platinum response biomarkers from the literature and integrated these with our empirical findings from the proteogenomic datasets to identify a candidate signature for detecting platinum-refractory disease prior to the treatment. We also performed gene-regulatory network analysis to identify potential biomarkers driving the response to chemotherapy.

Pediatric Brain Cancers

Tissue Source Site: Biomatrix/China

27 Urinary Protein Biomarker for Pediatric Medulloblastoma

Xiaolei Hao^{1,#}, Zhengguang Guo^{2,#}, Haidan Sun², Xiaoyan Liu², Yang Zhang¹, Liwei Zhang¹, Yongji Tian^{1*}, Wei Sun^{2,*}

¹Department of Neurosurgery, China National Clinical Research Center for Neurological Diseases, Beijing Tiantan Hospital, Capital Medical University, Beijing, China; ²Core Facility of Instrument, Institute of Basic Medical Sciences, Chinese Academy of Medical Sciences, School of Basic Medicine, Peking Union Medical College, Beijing, China

contributed equally to the work; *Corresponding authors: Prof. Yongji Tian (ttvysw1@163.com) or Prof. Wei Sun (sunwei1018@sin.com)

Note: Beijing Tiantan Hospital is one of the sub-site of BioMatrix in China, where glioblastoma was collected in the current CPTAC PHASE III.

Objective: Medulloblastoma (MB) is the most common type of brain tumor in children. Magnetic resonance imaging, the major means of diagnosis, often lead to misdiagnosis or missed diagnosis in clinics. Up to now there is generally no reliable biomarker for clinical screening or monitoring of the disease post-surgically. In this study proteomic approach was used to discover the urinary protein biomarkers for MB diagnosis and monitoring using urine specimen.

Methods: Tandem Mass Tags (TMT)-labeled quantitative proteomics was performed to uncover proteomic patterns in discovery groups (9 pair of MB pre-/post- surgery and 9 healthy controls). The differential urinary proteins were validated by Parallel reaction monitoring (PRM) approach in validation group (29 pair of MB pre-/post- surgery, 26 healthy control and 29 disease control with non-tumor benign nervous system diseases). The receiver operating characteristic (ROC) curves were used to evaluate sensitivity and specificity of potential biomarkers.

Results: A total of 114 differential proteins were identified in the discovery group. The functional annotation showed they were related to cancer characteristics, including cell proliferation, cell movement, apoptosis etc. The function related 18 differential proteins were validated in validation group. The ROC value of urinary protein panel for diagnosis and monitoring was 0.974.

Conclusion: The findings of our study may shed light on urinary proteomics as an effective way for detecting and monitoring of MB in clinic especially post-surgically.

PGDAC: Mount Sinai

28 Integrated Proteogenomic Characterization of Pediatric Brain Tumors

Nicole Tignor^{*,1}, Francesca Petralia^{*}, Pichai Raman², Dmitry Rykunov¹, Shrabanti Chowdhury¹, Azra Krek¹, Xiaoyu Song¹, Jiayi Ji¹, Anna Calinawan¹, Gonzalo Lopez Garcia¹, Richard G. Ivey³, Jeffrey Whiteaker³, Antonio Colaprico⁴, Seungyeul Yoo¹, Yuankun Zhu², Weiping Ma¹, Liang-bo Wang⁵, Elizabeth M. Appert², Ximena Cuellar², Jena Lilly², Javad Nazarian², Cheryl Irving², Ying Wang⁶, Joshua Wang⁶, Xi Steven Chan⁴, Jun Zhu¹, Eric Schadt¹, Tara Hiltke⁷, Henry Rodriguez⁷, Antonio Iavarone⁸, Maciej Wiznerowicz⁹, David Fenyo⁶, Li Ding⁵, Alexey Nesvizhskii¹⁰, Boris Reva¹, Steven Gygi¹¹, Amanda Paulovich³, Phillip B. Storm^{*,2}, Adam C Resnick^{*,2}, Brian Rood^{*,12}, Pei Wang^{*,1}

¹Icahn School of Medicine at Mount Sinai, New York, NY; ²Children's Hospital of Philadelphia, Philadelphia, PA; ³Fred Hutchinson Cancer Research Center, Seattle, WA; ⁴University of Miami, Miami, FL; ⁵Washington University in St. Louis, St. Louis, MO; ⁶New York University, New York, NY; ⁷National Institute of Health, Bethesda, MD; ⁸Columbia University, New York, NY; ⁹Poznan University of Medical Sciences, Poznań, Poland; ¹⁰University of Michigan, Ann Arbor, MI; ¹¹Harvard Medical School, Cambridge, MA; ¹²Children's National Health System, Washington, DC

^{*}Presenter; Equal contributors. +Co-correspond (Contact: storm@email.chop.edu, RESNICK@email.chop.edu, BROOD@childrensnational.org, pei.wang@mssm.edu)

Genomic and transcriptomic characterization has allowed for subtyping tumors based upon mutational profiles and transcript levels. However, owing to the many layers of regulation between transcript and the active post-translationally modified protein, the functional gene product, it has been challenging to better understand tumor biology from these transcriptional differences. We hypothesized that a comparative analysis of the proteome and phosphoproteome across 7 childhood brain tumors would yield a deeper understanding of the differences in their functional biology. We performed isobaric labeling and mass spectrometric profiling of 219 fresh frozen tumor samples collected in a single institution representing the histologic diagnosis of high grade astrocytoma (26), low grade astrocytoma (93), ganglioglioma (18), ependymoma (32), medulloblastoma (22), atypical teratoid rhabdoid tumor (12), and craniopharyngioma (16). In total, we quantified 9155 proteins and 13632 phosphorylation sites across all the samples. Unsupervised clustering using the proteomic data reveals eight clusters with distinct protein and pathway activities. While some clusters coincide with histologic diagnosis subtypes, other clusters are a mixture of different diagnoses, including one cluster consisting of "aggressive" tumors characterized by poor survival, high stemness and proliferation scores. Immune landscape characterization reveals distinct regulation of immune related pathways across different diagnosis and protein clusters. Kinase-substrate association analysis identified the activation of key kinases in particular proteomics clusters. Leveraging the WGS data, we characterized the impact of BRAF and H3 mutations in low- and high-grade gliomas respectively, and of CNVs in multiple diagnoses. The incorporation of the proteomic dimension into large-scale efforts at tumor characterization adds functional insight that can help drive translational efforts.

Cancer Immunotherapy

PGDAC: Baylor College of Medicine

29 Cancer Neoantigen Prioritization Through Sensitive and Reliable Proteogenomics Analysis

Bo Wen^{1,2}, Kai Li^{1,2}, Yun Zhang^{1,2}, Bing Zhang^{1,2,#}

¹Lester and Sue Smith Breast Center, ²Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX - # Correspondence should be addressed to Bing Zhang (bing.zhang@bcm.edu)

Neoantigens encoded by tumor-specific mutations are ideal targets for T cell-based immunotherapy. Neoantigen discovery relies primarily on genomics but can be enhanced by proteomic evidence from proteogenomics. Variant peptide identification in proteogenomics has prompted the development of multiple quality control methods with varied sensitivity and reliability. Without an appropriate metric for systematic evaluation, there remains a lack of consensus on method performance. We propose to use the difference between accurately predicted and observed retention times for each peptide as an evaluation metric. To this end, we have developed a deep learning algorithm with high accuracy in retention time prediction based on peptide sequence. Analysis of three large scale cancer datasets with a total of 287 tumor samples using different quality control strategies resulted in substantially different numbers of identified variant peptides and putative neoantigens. Our systematic evaluation using the proposed metric provided novel insights and clear guidance on the selection of quality control strategies. We implemented the recommended strategy, which produced up to 3 times more potential treatment opportunities, in a computational workflow to support proteogenomics-based neoantigen prioritization. Source code for the workflow, named NeoFlow, can be downloaded at <https://github.com/bzhanglab/neoflow>.

PGDAC: Mount Sinai

30 Proteogenomics Characterization of Immune Infiltration and Immune Evasion in Tumor Tissue

Boris Reva*, Francesca Petralia*, Azra Krek, Dmitry Rykunov, Eric Schadt, Pei Wang*

Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mt Sinai, New York, NY

*Presenter; Equal contributors; +Correspond (contact: pei.wang@mssm.edu)

Cancers accumulate mutations that eventually make them very different from normal tissues and, hence, targetable by immune cells. Each tumor develops a particular way to evade immune cell surveillance, typically using in built biological mechanisms of evading or suppressing the immune response. For example, some cancer cells block immune response by increasing the production of checkpoint proteins, such as PD-1/PD-L1. The checkpoint therapy unblocks the suppressed immune system by inhibiting checkpoint proteins from interactions with receptors of immune cells. However, due to multiple immune evasion mechanisms simultaneously occurring in one tumor, only ~10-20% of tumors are truly sensitive to immune checkpoint therapies. To enhance efficiency of immune therapy, combination therapies able to suppress more than one immune evasion routes are often necessary. The recent CPTAC proteo-genomic cancer studies provide a unique opportunity to gain insights in immune evasion mechanisms in tumor tissues based on rich proteogenomic data sets. Characterization of the landscape of different types of immune and stromal cells in tumors resulted in identification of distinct immune subtypes in CCRCC and UCEC cancers. These immune subtypes were associated with specific immune evasion mechanisms according to RNA expression, proteome and phospho-proteome data. Additionally, we assessed activation of IFN- γ signaling pathway, a key pathway for initiating immune response, and identified genes and molecular pathways associated with activation or downregulation of IFN- γ signaling. Based on this analysis, we nominated genes and pathways involved in immune evasion, proposed targets for combination therapy and biomarkers for clinical diagnostics of cancer immune subtypes.

Pancancer Studies

Tissue Source Site: Poland

31 Cancer Stemness Defined by CPTAC Proteogenomics Guides Identification of Protein Targets Towards Personalized Therapies for Cancer Patients

Antonio Colaprico¹, Daniel Cui Zhou², Eric Storrs², Francesca Petralia³, Weiping Ma³, Bo Wen⁴, Giacomo Lanzoni⁵, Olivier Geveart⁶, Alexander J. Lazar⁷, Xi Chen¹, Antonio Iavarone⁸, Richard Luczak⁹, Larry Lambe¹⁰, Sylwia Mazurek¹², Iga Kołodziejczak¹², Henry Rodriguez¹¹, Mehdi Mesri¹¹, Chris Kinsinger¹¹, Ana I. Robles¹¹, Li Ding^{2,*}, Maciej Wiznerowicz^{12,*}

¹Department of Public Health Sciences, ⁵Department of Biochemistry and Molecular Biology, University of Miami - Miller School of Medicine, Miami, FL; ²Division of Oncology, Department of Medicine, Washington University in St. Louis, St. Louis, MO; ³Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY; ⁴Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX; ⁶Stanford University, Palo Alto, CA;

⁷Departments of Pathology and Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX; ⁸Institute for Cancer Genetics, Columbia University Medical Center, New York, NY; ⁹RL Aerodynamics, Dallas, TX; ¹⁰Multidisciplinary Software Systems Research Corporation, Bloomingdale, IL; ¹¹Office of Cancer Clinical Proteomics Research, National Cancer Institute, Bethesda, MD; ¹²International Institute for Molecular Oncology, Poznan University of Medical Sciences, Clinical Hospital of Lord's Transfiguration, Poznań, Poland

*senior co-authors

Cancer progression involves the gradual loss of a differentiated phenotype and acquisition of progenitor and stem cell-like features. Here, we provide new stemness indices for assessing the degree of oncogenic dedifferentiation. We used machine learning to extract epigenetic, transcriptomic, and proteomic features from human stem cells and applied the computed result to index the CPTAC tumor samples by their proteogenomic hallmarks of stemness.

Leveraging the resource generated by Human Induced Pluripotent Stem Cells Consortium, we extracted stemness signatures from the coherent epigenetic, transcriptomic, and proteomic datasets, using one-class logistic regression machine learning algorithm. The obtained stemness scores based on the DNA methylation and gene expression are significantly more robust compared to those published in our original work based on the dataset obtained from a smaller number of samples and analyzed by older genomics technologies. The stemness score computed by a mean of machine learning algorithms using proteomic dataset is novel and original. The obtained proteomic score clearly identifies the stem cells from the non-stem cell classes. Indexing of CPTAC tumors with proteomic stemness score brought us with previously unappreciated findings.

We have used the stemness scores computed using mDNA, mRNA, and proteins to interrogate the coherent proteogenomic CPTAC datasets. The initial analysis of over 1000 cancer samples obtained from primary carcinomas of breast, lung, kidney, uterus, brain, head and neck, and pancreas has confirmed our previously published results. Importantly, our original proteomic-based scored brought the analyses to novel dimension, far beyond previously described results. The initial findings of our work identified proteins driving aggressiveness of the primary tumors, thus causing resistance to existing therapies and worse survival of cancer patients. Targeting these proteins with existing or novel drugs may limit the progression of lethal neoplasms, thus bringing effective cures for cancer patients.

PGDAC: New York University/Washington University in St. Louis/ Brigham Young University

32 Mutational Signatures and Environmental Factor Influence Across Cancer Types

Yize Li^{1,2}, Michael Wendl^{1,2}, Song Cao^{1,2}, Li Ding^{1,2}

¹Department of Medicine and Genetics, Siteman Cancer Center, ²McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO

Cancers are caused by somatic mutations. Different mutational processes often generate different combinations of mutation types, termed “signatures” [PMID: 23945592]. In the CPTAC LUAD cohort, we identified signatures that recapitulated the well-known characteristics of cancer mutations, including C>A transversions in smoking signature [PMID: 30371878]. We found 385 and 197 DEGs with an absolute median protein expression change >1 and FDR < 0.05 in tumor and normal adjacent tissue between the samples with and without a clear smoking signature. Oncogenes CDK1 and cyclin B1 (CCNB1) are highly expressed in tumors only with smoking signatures, suggesting potential smoking-related etiology in the initialization and progression of lung cancers. Furthermore, we found that the phosphorylation level of other cyclin-dependent kinase CDK12 and CCNB2 are upregulated in tumors with smoking signature.

We further investigated links to environmental agents based on regression of the 96 possible trinucleotide mutation combinations between the samples in CPTAC LUAD and the environmental signatures reported by Kucab et al. [30982602]. Correlations were calculated by least-squares fit, with significance assessed by T-testing and subsequent multiple test correction via FDR. We found strong correlations (>0.75) of many samples with many of the signatures of polycyclic aromatic hydrocarbons (PAHs) known to be present in cigarette smoke.

With using the same strategy, we investigated the mutational signatures and environmental factor influence across different cancer types in CPTAC, TCGA, ICGC, and MMRF. Some signatures (e.g. APOBEC signature) were commonly found across cancer types while some were identified only in specific cancers such as ultraviolet light signature in skin cancers and in cancers of the lip categorized as head and neck cancers. Moreover, among the environmental agents, we found distinct Simulated Solar Radiation (SSR) in SKCM, smoking-related agents in lung cancer, and some in HNSC, ESCA, LIHC. Drug therapy signatures in a small set of STAD, COAD, READ.

33 Aberrations in Acetylation of Metabolic Proteins and Histone Complexes Across Cancer Types

Alla Karpova^{1,2,3}, Daniel Cui Zhou^{1,2}, Li Ding^{1,2}

¹Department of Medicine and Genetics, Siteman Cancer Center, ²McDonnell Genome Institute, ³Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO

Acetylation is one of the most abundant post-translational modifications in eukaryotic cells. This modification occurs on histones and is crucial for epigenetic regulation of gene expression. Acetylation also targets non-histone proteins involved in major vital cellular processes. Cancer cells frequently acquire aberrant acetylation level of both histone and non-histone protein types, as well as abnormal expression of histone acetyltransferases (HATs) and histone deacetylases (HDACs), contributing to cancer hallmarks manifestation. Here we investigated aberrations in acetylation of over 3000 proteins across 4 cancer types: breast (BRCA), endometrial (UCEC), lung adenocarcinoma (LUAD), and glioblastoma (GBM). In breast cancer, we extensively characterized acetylation profiles of metabolic protein in PAM50 subtypes, and discovered glycolytic enzymes are hypoacetylated in Basal subtype, while mitochondrial enzymes, including TCA cycle proteins, are hyperacetylated in this subtype. Acetylation of the majority of metabolic proteins has been shown to inhibit their enzymatic activity, suggesting that mitochondrial activity is suppressed in Basal subtype. In UCEC we showed association between hot-spot mutations in CTNNB1 and elevated acetylation of histone H2B N-terminal sites. Finally, we applied investigated mutational impact of significantly mutated genes on acetylation and characterized acetylation differences between cancer type subtypes and tumor and normal tissue.

34 DNA Methylation Instability and Its Functional Impacts in Human Cancers

Wen-Wei Liang^{1,2}, Hui Shen³, Li Ding^{1,2}

¹Department of Medicine and Genetics, Siteman Cancer Center, ²McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO; ³Center for Epigenetics, Van Andel Research Institute, Grand Rapids, MI

Deregulated DNA methylation are present in human cancers and cooperate with genetic alterations to drive the cancer phenotypes, such as MLH1 promoter hypermethylation leading to defective DNA mismatch repair. However, the nuances of DNA methylation instability and its translational impact in tumor development has not been systematically studied. Here we investigated the genetic, DNA methylation, RNA sequencing, and quantitative proteomics of more than 600 tumors to obtain a comprehensive molecular portraits of methylome. We modeled the distributions of DNA methylation at specific loci in adjacent normal samples, and evaluated the methylation status of corresponding tumor samples. The associated cis-transcriptional or translational changes were quantified, applying a generalized linear model for copy number variation and somatic mutations as covariates. In total, we identified 245 silenced genes and 43 epigenetically enhanced genes from RNA-seq data (FDR < 0.1), among which 40.2% exhibited corresponding protein changes in tumors, as well, suggesting epigenetic driver events execute functional roles in tumorigenesis. Of the genes showing cis-effects of methylation only at RNA but not at protein, their interacting partners and downstream players were further examined. Our results suggest a portion of methylation trans-effects are oncogenic. Under this framework, we are able to discover oncogenic processes that have been regulated by methylation and reveal novel and robustly molecular players with translational potential.

35 LINE-1 Activity in Cancer is Correlated Across Life-Cycle Stages and with Interferon Signatures

Wilson McKerrow, Xuya Wang, David Fenyo

New York University School of Medicine, New York, NY

Retrotransposons are genomic elements that are capable of copying themselves to a new locus via an RNA intermediate. LINE-1 elements are the only retrotransposon that is autonomous and active in the human genome. LINE-1 elements contain an internal promotor, two open reading frames and a poly A 3' UTR. LINE-1 elements are expressed during early development but become repressed through a combination of DNA and histone methylation as cells differentiate into somatic tissues. However, in many cancers, LINE-1 elements are dramatically de-repressed, possibly shaping tumor evolution. With support from Seven Bridges Genomics and the Cancer Genome Cloud, we are reanalyzing CPTAC data to measure LINE-1 activity at RNA, protein and insertion. Our early focus has been on endometrial cancer as we observe robust LINE-1 expression in this cancer and have high quality data. We see good correlation between LINE-1 RNA and protein expression and between LINE-1 expression and the number of identified somatic LINE-1 insertions. Additionally, after correcting for cancer purity, we see a correlation between LINE-1 expression and interferon signatures, reflecting in vitro experiments showing that LINE-1 expression leads to an induction of type 1 interferons.

Targeted Proteomics

PTRC: Fred Hutchinson Cancer Center/University of Alabama at Birmingham

36 Multiplexed Quantification Of Phosphorylation Signaling Dynamics in the DNA Damage Response Using Peptide Immunoaffinity Enrichment Coupled to Quantitative Mass Spectrometry

Jeff Whiteaker, Lei Zhao, Richard Ivey, Julia Voytovich, Regine Schoenherr, Jake Kennedy, Chenwei Lin, Amanda Paulovich
Fred Hutchinson Cancer Research Center, Seattle, WA

The DNA-damage response (DDR) is a critical network for maintaining genomic integrity, and mutations in the DDR are among the most frequently identified in tumors. Furthermore, quantifying the signaling dynamics of the DNA damage response pathway may provide predictive markers of response to therapy in ovarian cancer. We sought to expand our capability to use high throughput, mass spectrometry-based assays to monitor phosphopeptides by targeting components of the DNA damage response (DDR) network. Phosphoproteomics profiling was used to empirically identify cellular signaling components responsive to DNA damage (ionizing radiation) and dependent on the ATM kinase. Anti-peptide antibodies to the proteins and phosphopeptides identified as targets were generated and used to develop a 60-plex (36 phosphopeptides) targeted immuno-MRM assay. Assay characterization included response curves to determine limits of detection and the linear range of response and repeatability was determined by analyzing three validation levels using independent preparations on five separate days. The characterized assays were applied to lymphoblast cell lines and peripheral blood mononuclear cells (PBMCs) exposed to ionizing radiation to monitor the response to DNA damage. Kinase inhibitors were also applied to demonstrate the utility of the assay for profiling pharmacodynamics. The results show key changes in the temporal response to DNA damage, demonstrating the use of the multiplexed immuno-MRM assay to measure phosphopeptides and characterizing phosphorylation dynamics and pharmacodynamics in cells.

Technology Development

PGDAC: New York University/Washington University in St. Louis/ Brigham Young University

37 Single-Cell Omics and Proteomic Approaches to Study Pancreatic Ductal Adenocarcinoma Heterogeneity and Treatment Response

Daniel Cui Zhou^{1,2}, Reyka Jayasinghe^{1,2}, Wagma Caravan^{1,2}, Ryan Fields¹, Li Ding^{1,2}

¹Department of Medicine and Genetics, Siteman Cancer Center, ²McDonnell Genome Institute, Washington University in St. Louis, St. Louis, MO

Pancreatic Ductal Adenocarcinoma (PDAC) is a highly aggressive cancer with poor survival rates. In this study, we conduct a vast exploration of tumor heterogeneity in PDAC focusing on the interconnection with treatment regimens and patient response. We collected 30 PDAC tumors which were split into the following groups: treatment-naïve, neoadjuvant-FOLFOX, and neoadjuvant-gemcitabine/abraxane. Each tumor biopsy was sampled 2-6 times to address tumor heterogeneity via extensive omics, including bulk DNA, RNA, and proteomics, single-cell RNA-Sequencing (scRNA-Seq), and cellular imaging, all from the same sample. Preliminary analysis revealed that samples from the same tumor tend to have similar overall single cell profiles but with a variable degree of tumor purity and immune/stroma cell types. We further observe distinguishable ductal cell subpopulations consistent across samples, which may be related to treatment response. We envision great synergy with current CPTAC proteomics work on PDAC.

CPTAC has generated genomic, transcriptomic, and proteomic data for 76 pancreatic tumors and adjacent normal tissues so far, with a goal of 157 cases. The purity of these tumors, estimated via *KRAS* hotspot mutation variant allele fraction or ESTIMATE, varies greatly. Gene expression of cell surface markers and immune checkpoint genes revealed 4 immune subtypes in PDAC with unique proteomics markers: T cell-inflamed, NK+ excluded, NK- excluded, and immune desert. For instance, T cell-inflamed samples have upregulated hemostasis proteins, which seem to be anticorrelated with immune desert samples. Interestingly, CD38, a B cell marker, is significantly upregulated in the inflamed subtype only at the protein level. The relationship between proteomic markers and subtypes with single-cell-derived findings will provide great insight into the tumor microenvironment in PDAC and its relationship and impact on treatment response.

Tissue Source Site: Biomatrix/China

38 Microfluidic System of Zebrafish PDX Models Facilitates Fast-Track Regimen Decision for Pancreatic Cancer

Lei Wang^{*,1}, Huan Chen^{*,2,3}, Xianfeng He^{2,3}, Bo Yu^{2,3}, Shirley X. Tsang[#], Xu Wang¹, Jiang Long^{2,4,5,6}

¹Department of Biochemistry and Molecular Biology, Key Laboratory of Metabolism and Molecular Medicine, Ministry of Education, School of Basic Medical Sciences, ⁴Department of Pancreatic Surgery, ³Fudan University Shanghai Cancer Center, ⁵Department of Oncology, Shanghai Medical College, ⁶Pancreatic Cancer Institute, Fudan University, Shanghai, China; ²National Human Genetic Resources Sharing Service Platform (2005DKA21300); [#] BioMatrix, Maryland, USA

Precision oncology aims to find the most beneficial therapeutic strategies for individual patients, and PDX models generated by xenografting of human primary cancer tissues/cells into immune-compromised hosts are promising tools for personalized drug assessment in clinic. Previously, we reported an optimized approach of using zebrafish larvae instead of mice as hosts to generate heterogeneous PDX model of pancreatic cancer for comparative drug assessment (PMID: 31107449). However, the successful rate for isolating culturable tumor cells from primary tissues is low (<20%), while the isolation of culturable non-tumor stromal cells, which form the majority of the tumor microenvironment in pancreatic cancer, achieves a successful rate close to 100%. To investigate whether the drug responses in individual pancreatic cancer cases can be predicted by assessing the behaviors of those primary stromal cells with or without primary tumor cells, we first harvested and expanded 12 cases of tumor stromal cells, which are transiently immortalized by introducing BCL2L1, and mixed with cancer cell lines MiaPaCa-2 and MIN6 separately. The cell mixtures were then injected into the yolk sac of transgenic zebrafish larvae conditionally expressing two human growth factors that are extensively used in non-serum *in vitro* culture, followed by alignment of the zPDX models in a three-dimensional printing-based microfluidic system, which partially immobilizes the models for high-throughput scan. After that, 7 different chemotherapy regimens and DMSO were administrated via internal cycling tunnels on the microfluidic chip. Both the number and the composition of the stromal-tumor mixtures in zPDX models were quantitated, and the results indicated that the expression differences of drug metabolism-relevant genes in stromal cells are significantly associated with the survival and drug sensitivity of their chaperoned tumor cells. Lastly, we also harvested the PBMCs from the circulating blood, performed transient immortalization and/or genetic enhancement, followed by injection into the circulation of zPDX models. We observed that the modified PBMCs robustly bind and attack the xenografted tumor cells in zPDX models, and therefore the zPDX model may serve as a fast-track platform for assessing a variety of strategies of both chemo- and immuno- therapy *in vivo*.

PTRC: Broad Institute/Baylor College of Medicine

39 Microscaled Proteogenomic Methods for Precision Oncology

Shankha Satpathy^{1,}, Eric Jaehnig^{2,*}, Karsten Krug¹, Michael Gillette¹, Alexander Saltzman², Kimberly Holloway², Meenakshi Anurag², Chen Huang², Purba Singh², Beom-Jun Kim², Goerge Miles², Noel Namai², Anna Malovannaya², D.R. Mani¹, Chuck Perou³, Bing Zhang², Steven Carr¹, Matthew Ellis²*

¹Broad Institute, Cambridge, MA; ²Baylor College of Medicine, Houston, TX; ³University of North Carolina at Chapel Hill, Chapel Hill, NC

shankha@broadinstitute.org, scarr@broadinstitute.org, Matthew.Ellis@bcm.edu

Cancer proteogenomics combines genomics, transcriptomics and mass spectrometry-based proteomics to gain insights into cancer biology and treatment responsiveness. While proteogenomics analyses have shown great potential to deepen our understanding of cancer tissue complexity and signaling, how a patient's tumor changes upon treatment has largely been the province of genomics. This is due to technical difficulties associated with doing proteogenomic analysis on clinic-derived core-needle biopsies. To address this critical need, we have developed a "microscaled" proteogenomics approach for tumor-rich OCT-embedded core needle biopsies. Tissue-sparing specimen processing ("Biopsy Trifecta EXtraction," BioTExt) and microscaled proteomics (MiProt) methodologies allowed generation of deep-scale proteogenomics datasets, with copy number and transcript information for >20,000 genes and mass spectrometry-based identification and quantification of nearly all expressed proteins in a tumor (>10,000 proteins) and more than >20,000 phosphosites starting with just 25 micrograms of peptides per sample. In order to understand the capabilities and limitations of our approach relative to conventional deepscale proteomics requiring >10X more starting material, we compared preclinical patient derived xenograft (PDX) models at conventional scale with data obtained by core-needle biopsy of the same tissues. Comparable depth and biological insights were obtained from the cores relative to surgically resected tumors. As a proof-of-concept for implementation in clinical trials, we applied microscaled proteogenomic methods to a small-scale clinical study where biopsies were accrued from patients with ERBB2+ advanced breast cancer before and 48 to 72 hours after the first dose of neoadjuvant Trastuzumab-based chemotherapy. Multi-omics comparisons were conducted between samples associated with residual disease versus samples associated with complete pathological response. Integrative proteogenomic analyses efficiently diagnosed the molecular bases of diverse candidate treatment resistance mechanisms including: 1) absence of ERBB2 amplification (false-ERBB2+); 2) insufficient ERBB2 activity for therapeutic sensitivity despite ERBB2 amplification (pseudo-ERBB2+); 3) resistance features in true-ERBB2+ cases including androgen receptor signaling, mucin expression and an inactive immune microenvironment; 4) lack of acute phospho-ERBB2 down-regulation in non-pCR cases. In summary, we have developed a proteogenomics pipeline well suited for large-scale cancer clinical studies to identify potential resistance mechanism in patients. We conclude that microscaled cancer proteogenomics could improve diagnostic precision in the clinical setting.

PCC: Pacific Northwest National Laboratory

40 Comprehensive Nanoproteomics Analysis Enabled by an Intelligent Boosting to Amplify Signal with Isobaric Labeling (iBASIL) Strategy

Chia-Feng Tsai¹, Rui Zhao², Richard D. Smith¹, Karin D. Rodland¹, Tujin Shi¹, Ying Zhu², Tao Liu¹

¹Biological Sciences Division, ²Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA

Mass spectrometry (MS)-based proteomics has tremendous potential to overcome limitations of the popular flow cytometry and mass cytometry methods and achieve antibody-independent, comprehensive, and quantitative proteomics analysis of single cells. Such potential has yet to be realized, however, due mainly to ineffective sample processing and the inability to maintain or increase sample signal for MS analysis. To tackle this challenge, we recently developed a boosting to amplify signal with isobaric labeling (BASIL) strategy which uses one of the isobaric tandem mass tag (TMT) channels as the boosting channel (i.e., with much higher sample input) to enhance the detection of low abundance peptides or phosphopeptides at MS1 level. After MS/MS fragmentation of peptides backbone, the peptide sequence information is obtained with the major contribution from

the boosting sample, while the quantitation signal is recorded from the TMT reporter ion intensities in the individual sample channels. Using the BASIL strategy, we were able to quantify over 20,000 phosphorylation sites from human pancreatic islet. Herein, we set out to systematically evaluate the effects of boosting settings on the identification and quantitation performance of the method. We found that increasing boosting ratio increases signals variation and decreases the signal to noise ratio of lower abundance peptides due to insufficient ion injection time and the limited ion capacity of Orbitrap. Therefore, we developed an intelligent BASIL (iBASIL) strategy through optimization of both the boosting ratio and the MS data acquisition parameters for significantly improved quantitation precision while achieving comprehensive proteome coverage, for samples as small as single cells. With iBASIL, a boosting ratio of 1,000 can be used to provide ~1,500 protein identifications without compromising the ability to robustly quantify proteins in the single cells of three different acute myeloid leukemia cell lines. In addition, we further integrated the iBASIL strategy with the in-house tip-based immobilized metal affinity chromatography (IMAC) approach for ultrasensitive phosphoproteomics analysis. Approximately 700 phosphorylation sites can be quantified from as little as 20 ng tryptic peptides (equivalent to 200 cells). We believe this iBASIL platform can be broadly applied to systems biology and biomedical research in small numbers of mammalian cells as well as precious mass-limited clinical specimens, with the potential of moving toward single-cell proteomics applications.

41 Carrier-Assisted Single-Tube Processing Approach for Targeted Proteomics Analysis of Small Number of Cells

Pengfei Zhang¹, Matthew J. Gaffrey¹, Ying Zhu², William B. Chrisler¹, Thomas L. Fillmore², Carrie D. Nicora¹, Richard D. Smith¹, Karin D. Rodland¹, Tao Liu¹, Tujin Shi¹

¹Biological Sciences Division, ²Environmental Molecular Sciences Laboratory, Pacific Northwest National Laboratory, Richland, WA

Antibody-based flow cytometry and mass cytometry are predominant technologies for targeted proteomics analysis of single cells. However, they share the common shortcomings with other antibody-based methods (e.g., low-multiplex, the need of high-quality antibodies, and unavailability of antibodies for new proteins). Furthermore, they do not provide accurate protein concentrations. Mass spectrometry (MS)-based targeted proteomics such as selected reaction monitoring (SRM) has emerged as a promising alternative for antibody-independent, precise, and highly multiplexed quantification of target proteins. However, it has not been suitable for the analysis of single cells due to ineffective processing of such small samples for MS analysis. To tackle this issue, we recently developed a new carrier-assisted single-tube preparation method, namely carrier-assisted liquid chromatography coupled to SRM (cLC-SRM), for ultrasensitive targeted MS analysis of small number of cells. Herein, we combined a new single-tube digestion protocol to process low numbers of cells at low sample volume ($\leq 15 \mu\text{L}$) with minimal loss for protein quantification. This single-tube protocol builds upon trifluoroethanol digestion and further minimizes sample losses by tube pretreatment and the addition of carrier proteins. We also optimized the denaturing temperature and trypsin concentration to significantly improve digestion efficiency. The cLC-SRM method was demonstrated to have sufficient sensitivity for reproducible detection of most EGFR pathway proteins expressed at levels of $\geq 30,000$ and $\geq 3,000$ copies per cell for 10 and 100 mammalian cells, respectively. Thus, cLC-SRM enables reliable quantification of low to moderately abundant proteins in less than 100 cells, and could be broadly useful for multiplexed quantification of important proteins in small populations of cells or in size-limited clinical samples. We anticipate that further improvements of this method will enable targeted single-cell proteomics using either SRM or potentially other emerging ultrasensitive MS detection approaches. Furthermore, this method can be easily implemented in any proteomics laboratories at no additional cost for instrument or reagents. This approach can be broadly applied to biomedical research and systems biology for absolute quantification in small numbers of cells.

42 **An Integrative Analysis of Tumor Proteomic and Phosphoproteomic Profiles to Examine the Relationships Between Kinase Activity and Phosphorylation**

Osama A. Arshad, Vincent Danna, Vladislav Petyuk, Paul D. Piehowski, Tao Liu, Karin D. Rodland, Jason E. McDermott

Pacific Northwest National Laboratory, Richland, WA

Phosphorylation of proteins is a key way cells regulate function, both at the individual protein level and at the level of signaling pathways. Kinases are responsible for phosphorylation of substrates, generally on serine, threonine, or tyrosine residues. Though particular sequence patterns can be identified that dictate whether a residue will be phosphorylated by a specific kinase, these patterns are not highly predictive of phosphorylation. The availability of large scale proteomic and phosphoproteomic datasets generated using mass-spectrometry-based approaches provides an opportunity to study the important relationship between kinase activity, substrate specificity, and phosphorylation. In this study, we analyze relationships between protein abundance and phosphopeptide abundance across more than 150 tumor samples and show that phosphorylation at specific phosphosites is not well correlated with overall kinase abundance. However, individual kinases show a clear and statistically significant difference in correlation between known phosphosite targets for that kinase and randomly selected phosphosites. We further investigate relationships between phosphorylation of known activating or inhibitory sites on kinases and phosphorylation of their target phosphosites. Combined with motif-based analysis, this approach can predict novel kinase targets and show which subsets of a kinase's target repertoire are specifically active in one condition versus another.

