Multiomics Sample Preparation Workflow for Simultaneous Cleanup of DNA, RNA, and Proteins using ProMTag

Stephanie Biedka[1], Svitlana Yablonska[1], Duah Alkam[2], Charity L. Washam[2], Xi Peng[3], Kevin Xiao[3], Stephanie D. Byrum[2], Jonathan S. Minden[1]

[1]*Impact Proteomics, Pittsburgh, PA,* [2]*UAMS, Little Rock, AR,* [3]*Center for Clinical Mass Spectrometry, AHN Cancer Institute, Pittsburgh, PA*

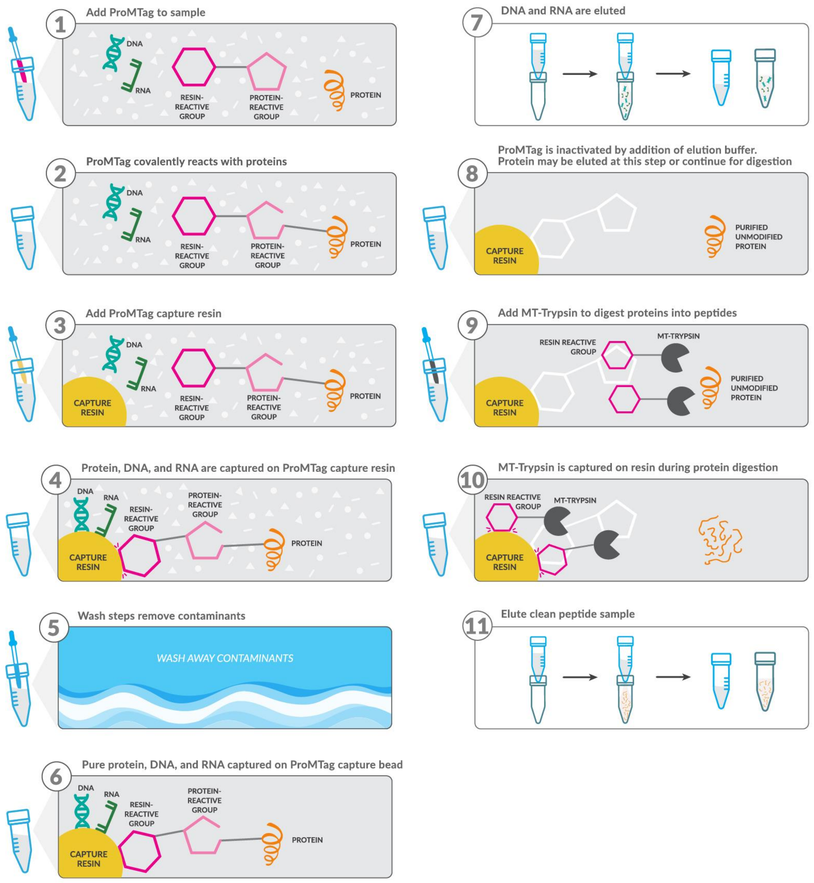
Abstract

Sample preparation is the critical first step in any -omics workflow. Removal of contaminants, such as salts, detergents, and other molecules that interfere with analysis, while maintaining high yields of the desired product(s) is key to obtaining reproducible, high-quality results. Combining multiple -omics analyses to gain deeper insights into biological processes is becoming increasingly common. However, multiomics sample preparation typically requires several different sample preparation kits or workflows.

We have developed a novel multiomics sample preparation workflow that starts with a single cell or tissue sample and yields clean preparations of DNA, RNA, and protein ready for downstream analysis. This workflow relies on Impact Proteomics' reversible click chemistry tag, ProMTag. ProMTag reversibly modifies primary amines on the surface of proteins, allowing tagging of all proteins in a sample and capture of those proteins on a complimentary ProMTag capture resin. We have modified our foundational Universal Protein Extraction and Cleanup Kit (UPECK) workflow to enable capture of DNA and RNA on this same ProMTag capture resin. The resin-bound protein and nucleic acids can then be thoroughly washed to remove contaminants. The nucleic acids are eluted first by a solubilizing buffer. Proteins are then released from the capture resin by reversal of the ProMTag and digested in a 1-hour digest with our modified MT-Trypsin. The entire workflow requires less than 3 hours to complete. Starting with one aliquot of a cell lysate containing 100 µg of protein, we were able to produce high-quality DNA (average yield 0.35 µg), RNA (average yield 2.4 µg), and peptides (average yield 10.4 µg), which we utilized for whole-genome sequencing, mRNA sequencing, and mass spectrometry, respectively. This novel, streamlined workflow establishes a new standard for high-yield and reproducible simultaneous preparation of DNA, RNA, and peptides, allowing for simplification of multi-omics analyses and more efficient multiomics data integration.

ProMTag Multiomics protein, DNA, and RNA extraction and cleanup workflow

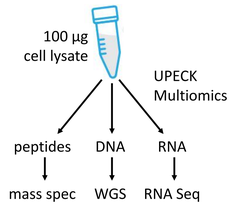
The Multiomics workflow relies on Impact Proteomics' unique tag, ProMTag. ProMTag is a reversible protein tag with a click chemistry moiety that allows for rapid, unbiased, and reproducible capture and release of proteins in the presence of harsh denaturants and contaminants without protein loss.



Acknowledgements

This work was supported by NSF SBIR Phase II Award 2036199 (Impact Proteomics), NIH grant P20GM121293 and the UAMS Winthrop P. Rockefeller Cancer Institute (Stephanie D. Byrum). We thank Uma Nagarajan, Archana Lovett, and Kyle Galford from Novogene for the WGS and RNA Sequencing analyses.

The ProMTag Multiomics workflow yields sufficient high-quality peptides, DNA, and RNA for mass spectrometry, WGS, and RNA Sequencing

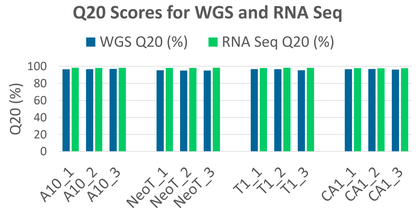


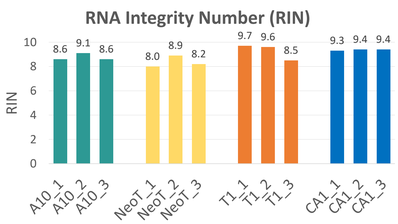
We prepared peptides, DNA, and RNA from a set of four genetically related human cell lines that recapitulate successive stages in breast cancer development.

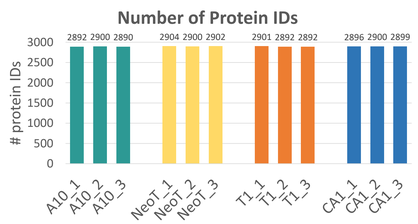
A10: Immortalized normal
NeoT: Benign hyperplasia
T1: Atypical hyperplasia
CA1: Invasive cancer

Sample	Peptide (ug)	DNA (ug)	RNA (ug)
A10	11.56	0.37	2.61
NeoT	11.54	0.72	2.29
T1	9.98	0.13	1.87
CA1	8.59	0.18	2.82

Average peptide, DNA, and RNA yield across three technical replicates.

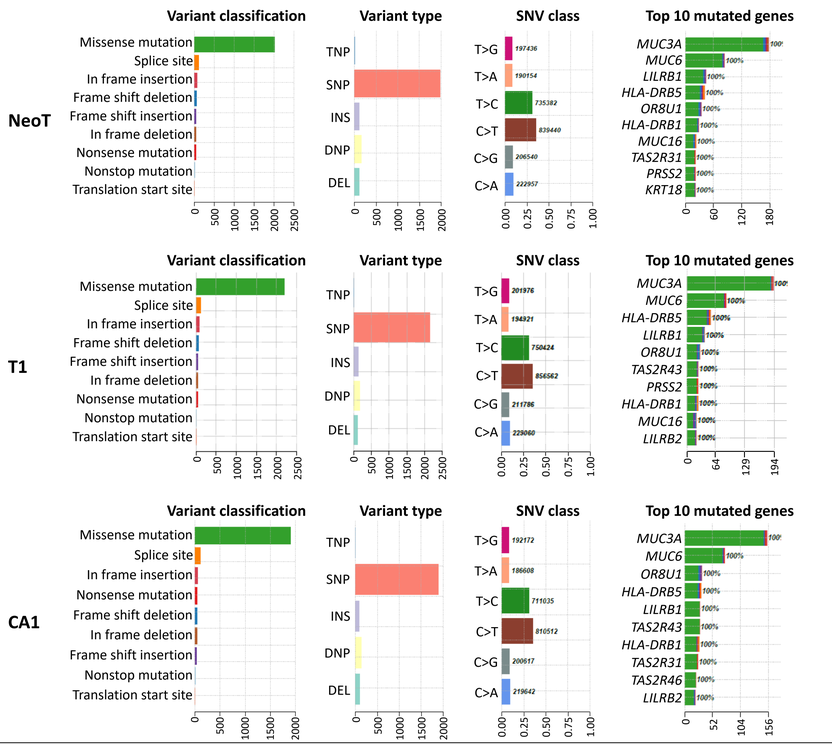




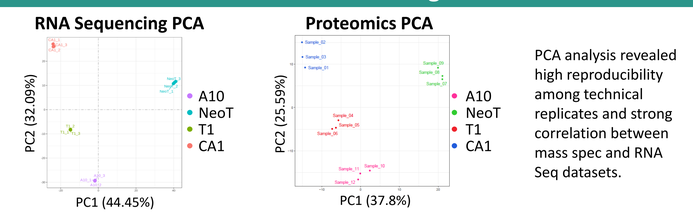


WGS data showed high consistency among mutation types and top mutated genes in NeoT, T1, and CA1

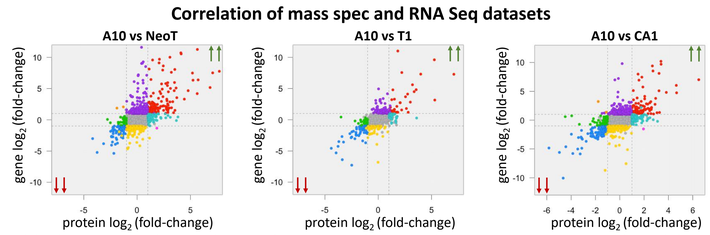
Genomic data for the NeoT (benign hyperplasia), T1 (atypical hyperplasia), and CA1 (invasive cancer) cell lines were compared to the A10 (immortalized normal) control.



Multiomics data integration

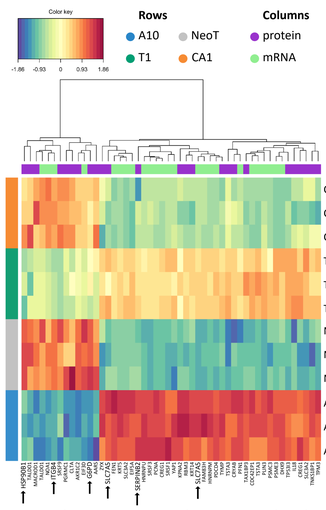


PCA analysis revealed high reproducibility among technical replicates and strong correlation between mass spec and RNA Seq datasets.



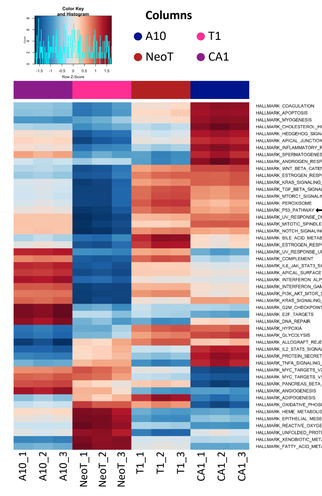
The $\log_2$ fold changes values for gene (RNA Seq) and protein (mass spec) expression are shown. The lower left and upper right quadrants indicate features with positive correlation where both the protein and gene are down- or up-regulated.

Top 50 features separating the groups

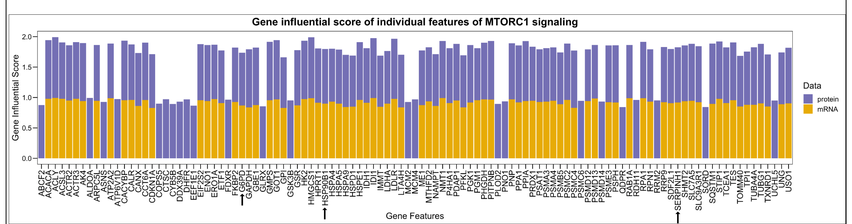


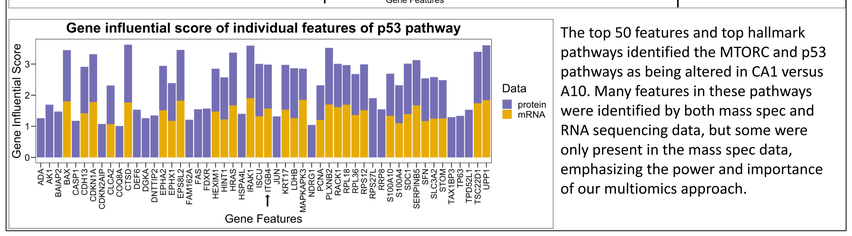
Protein and gene expression data were analyzed using MixOmics. This identified the top 50 features that distinguish the cancer cell lines from the A10 control samples.

Top Hallmark pathways



Protein and gene expression data were analyzed using multi-omics gene-set analysis (MOGSA) to identify the top hallmark pathways that distinguish the cancer cell lines from the A10 control samples.





The top 50 features and top hallmark pathways identified the MTORC and p53 pathways as being altered in CA1 versus A10. Many features in these pathways were identified by both mass spec and RNA sequencing data, but some were only present in the mass spec data, emphasizing the power and importance of our multiomics approach.