Complementary ERLIC and RPLC Online Separations Significantly Expand Sequence Coverage in MS-based Proteomic and Proteogenomic Studies

C. R. Guerrero1, P. Jagtap2, J. Johnson1, J. Yong1, T. J. Griffin1,2
1University of Minnesota Minneapolis, MN/US, 2Center for Mass Spectrometry and Proteomics, St. Paul, MN

Introduction

- Traditional “bottom-up” MS-based proteomics uses RPLC-MS separation to resolve complex tryptic protein digests
- Although RPLC-MS is the “Gold Standard,” it lacks detection of adequate protein sequence coverage across a majority of proteins
- “Bottom-up” methods for improving protein sequence coverage are particularly important for PTM, isoform detection, and proteogenomic assessments
- Proteogenomics requires widespread coverage for the identification of sequence regions that may carry disease-associated variants
- We demonstrate that protein sequence coverage can be substantially increased when executing two modes of liquid chromatography prior to MS acquisition

Galaxy-P Framework

- Galaxy-P is a web-based community developed bioinformatics framework/platform/workbench
- Originally designed to address issues in genomic informatics including:
  - Software accessibility and usability
  - Analytical transparency
  - Reproducibility
  - Scalability
  - Share-ability

ERLIC

- Electrostatic repulsion hydrophilic interaction chromatography (ERLIC) utilizes two dimensions of separation
  - Hydrophilic Interaction
  - pI – Isoelectric Point
- ERLIC offers a complementary approach to RPLC due to its affinity to retain more polar peptides, as well as, longer acidic peptide sequences
- The combination of ERLIC-MS and RPLC-MS offers huge potential to provide the protein sequence coverage for proteogenomic purposes

Methods

- 40 μL of MCF-7 protein lysate was resolved using SDS-PAGE gel separation, of which, five major bands were excised and subsequently digested by trypsin
- ERLIC samples were loaded with a buffer containing 85% ACN, 0.1% acetic acid, and 10 mM ammonium acetate
- The analytical column (PolySAX LP) was pre-equilibrated with 80 μL of 90% ACN with 0.1% acetic acid
- Each sample (~200 ng) was analyzed by both RPLC-MS/MS and ERLIC-MS/MS using “Top Speed” data dependent acquisition on a Thermo Orbitrap Fusion
- Raw data files were analyzed using Galaxy-based proteomic and proteogenomic workflaws (Jagtap et al. 2014)

Proteomic Results

- Peptide and Protein IDs from RPLC-MS/MS and complementary ERLIC-MS/MS for individual MCF-7 gel bands using ProteinPilot™ v5.0, (5% PSM local FDR)
- 44% Peptide and Protein IDs from RPLC-MS/MS and complementary ERLIC-MS/MS overlap
- 38% Peptide and Protein IDs from ERLIC-MS/MS and complementary RPLC-MS/MS overlap

Results

- ERLIC with RPLC-MS drastically increased protein sequence coverage for a majority of proteins in MCF-7 cells
- Galaxy-P’s proteogenomic workflow yielded 456 and 397 novel peptide sequences within MCF-7 cell lines for ERLIC and RPLC, respectively
- Using RPLC with complementary ERLIC can provide improved sequence coverage for improved proteogenomic outcomes which could improve our understanding of disease-associated variants.

Conclusions and Future Directions

- NSF Grant 1147079
- Center for Mass Spectrometry and Proteomics, University of Minnesota

Acknowledgements