CLUSTERING PRECURSOR AND ITS FRAGMENTATION DATA ACROSS MULTIPLE SAMPLES

Guo-Zhong Li, Marc V. Gorenstein, Dan Golick, Richard Denny, Craig A. Dorschel, Jeffrey C. Silva, Scott J. Geromanos

Waters Corporation, Milford, MA

INTRODUCTION

Biomarker discovery by proteomic profile based on LC/MS^E technology needs to track ions from different samples. Here we develop algorithm to use protein ID, peptide sequence and retention time to track ions. The tracking result from replicated injections of one sample can be used for more accurate intensity measurement because that they are replicated data. The tracking result from different samples can give the expression changes (intensity changes) of the samples and lead to biomarker discovery.

Purpose:

Track peptides and proteins identified from LC/MS^E across multiple samples to study what is changed, what is not changed, what is new and what is missing in the samples.

Keys of Technology:

1. Data is acquired by alternating MS (Waters QTof Premier)
2. Lock mass and < 10 ppm mass accuracy of QTof.
3. Low energy data (precursors) and elevated energy data (products) are linked together by *retention time in 0.05 minutes or less window*.
4. Protein fasta database are searched from list of precursor with products by Ion Accounting Database Search Engine.
5. Protein ID, peptide sequence and calibrated retention time of peptide are used to track peptides across injections.
6. Tracking results are studied statistically.

METHODS

LC Conditions

- ***A Waters* nano ACQUITY**
- **The column:** Atlantis dC18 300 μ m

MS Conditions

- ***A QTof mass spectrometer in Vmode*** (~12K FWHM)
- ***a nano-LockSpray ion source*** (~ 10ppm mass accuracy).
- The data was acquired ***(LE/EE Switch) i***n alternate scanning mode with a data acquisition time of 1.85 second for both the low (CE = 8) and high energy (CE = 28-35) channels.

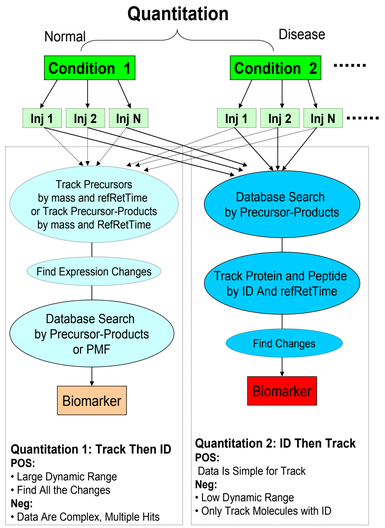
Sample Preparation

MassPrep Digestion Standards (**MPDS**):
Yeast Alcohol Dehydroenase (ADH)
Yeast Enolase (ENO)
Bovin Serum Albumin (BSA)
Rabbit Glycogen Phosphorylase (GPB)

Study 1
Sample 1 = EColi + MPDS Mix 1
Sample 2 = EColi + MPDS Mix 2

Study 2 (J. Vissers et al. WP662, 2006 ASMS)
Sample 1 = Depleted Human Serum Normal
Sample 2 = Depleted Human Serum Disease
Sample 3 = Depleted Human Serum Drug Treatment
Sample 4 = Purified Biomarker

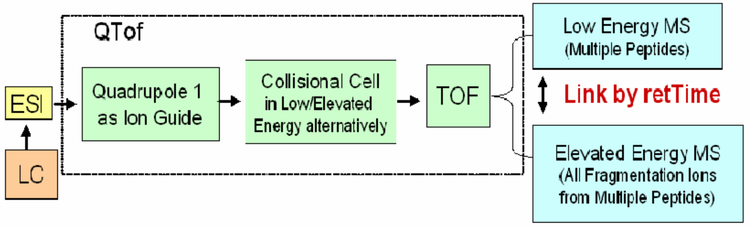
DATA FLOW FOR QUANTITATION



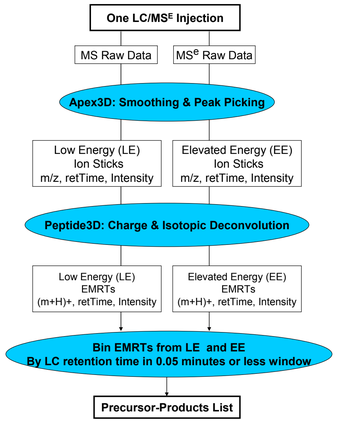
Quantitation 1: Track Then ID
POS:
• Large Dynamic Range
• Find All the Changes
Neg:
• Data Are Complex, Multiple Hits

Quantitation 2: ID Then Track
POS:
Data Is Simple for Track
Neg:
• Low Dynamic Range
• Only Track Molecules with ID

Waters Protein Expression System Based on QTof

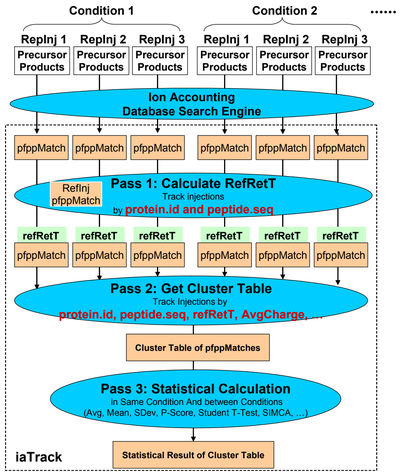


Alternating MS Data Processing Flow

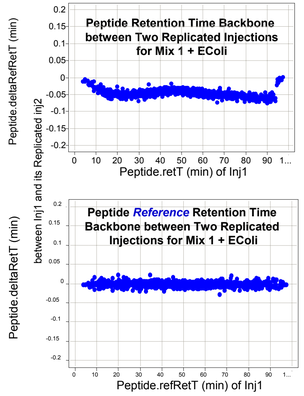


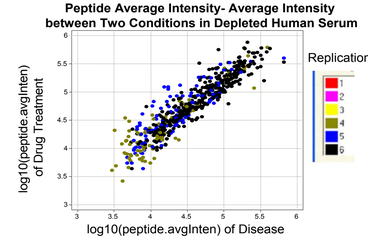
iaTrack Data Flow

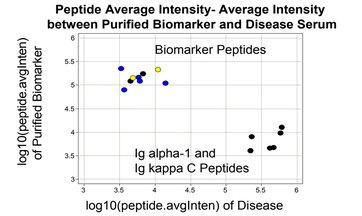
pfppMatch: Peptide-Fragments and Precursor-Products Match



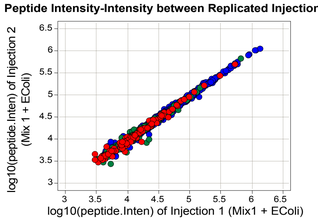
RESULTS

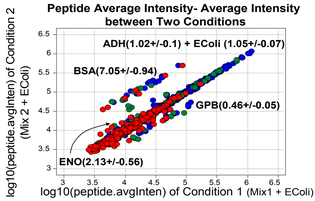






CONCLUSION





iaTrack Result for Mix 1 and 2 in EColi

Protein Name	Mix 1 (Amount)	Mix 2 (Amount)	iaTrack Result	sDev	CV(%)
Yeast Alcohol Dehydroenase (ADH)	1.0	1.0	**1.02**	0.1	9.45
Yeast Enolase (ENO)	1.0	2.0	**2.13**	0.56	26.53
Bovin Serum Albumin (BSA)	1.0	8.0	**7.05**	0.94	13.32
Rabbit Glycogen Phosphorylase (GPB)	1.0	0.5	**0.46**	0.05	11.14

Conclusion

1. Data is collected by Waters Protein Expression System (LC/MS^E).
2. Precursor ions (MS) and their Product ions (MS^E) are linked together by retention time.
3. Precursor-Products table are sent to Ion Accounting Database Search Engine to identify Peptides and Proteins.
4. Peptides and Proteins are tracked by Peptide Sequence, Protein ID and peptide's refRetTime.
5. After tracking, the intensity differences of Peptides from different injections of different sample are used for quantitation of peptides and proteins.
6. Statistical calculation are done in the conditions and between conditions.

Acknowledgements

J. Vissers (for Serum Raw Data, ref. WP662, 2006 ASMS)
M. Stapels (for Mix1 and Mix2 in EColi Raw Data)

720001723EN

TO DOWNLOAD A COPY OF THIS POSTER VISIT WWW.WATERS.COM/POSTERS