Development of a human plasma and serum metabolomic feature database and application in the study of type 1 diabetes mellitus

Thomas O. Metz[1], Christina M. Sorensen[1], Navdeep Jaitly[1], Stephen J. Callister[1], Matthew E. Monroe[1], Jennifer S.D. Zimmer[1], Ronald J. Moore[1], Patricia W. Mueller[2], and Richard D. Smith[1]

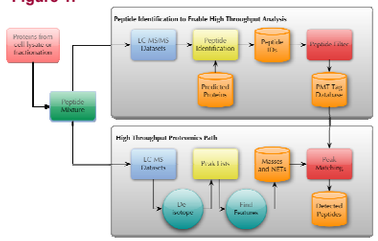
[1]Pacific Northwest National Laboratory, Richland, WA 99352

[2]U.S. Centers for Disease Control and Prevention, Atlanta, Georgia 30341

Overview

The mass range typically scanned in LC-MS-based metabolomics measurements (100 – 1,000 *m/z*) leads to a large number of detected features that include both real analytes and chemical noise. While the contribution of chemical noise can be reduced through the use of intensity thresholding or application of matched filtration[1,2] during feature identification, an unknown number of chemical noise features are often included in downstream data processing and analysis. Higher confidence can be placed in metabolomic data if the identified features are observed in multiple replicates of the same sample. Therefore, a modified accurate mass and time (AMT) tag approach[3] was used to catalogue those metabolomic features reproducibly observed in human plasma and serum samples, which were subsequently used in comparative analyses of control and disease samples.

Figure 1.



The AMT tag approach for high-throughput proteomics

Introduction

A comparative study using capillary LC-FTICR was performed to both assess the effect of fasting and non-fasting on metabolomic profiles of plasma and serum and to develop a feature database for future comparative analyses. Samples collected from the same ten individuals were pooled accordingly to create the following four uniform samples:

- fasting serum
- non-fasting serum
- fasting plasma
- non-fasting plasma

Subsequently, ten healthy and ten newly diagnosed type 1 diabetic patient samples from the Diabetes Autoantibody Standardization Program (DASP) were analyzed by capillary LC-FTICR to identify potential novel biomarkers of type 1 diabetes mellitus.

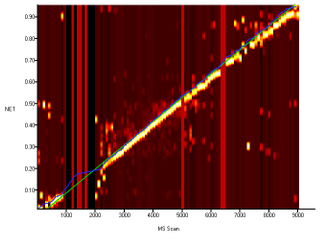
Methods

- Samples prepared in triplicate by chloroform/methanol extraction
- Aqueous fractions analyzed by reversed-phase capillary LC-FTICR
- Raw data deconvoluted and clustered into mass, time, and intensity features
- Resulting datasets aligned by either manual method or automated approach

Manual alignment was accomplished using a two-step iterative process:

- Initial alignment of common features using wide LC normalized elution time (NET) window (± 0.075 NET) and mass window representative of instrument mass measurement accuracy (± 3 ppm) to produce composite reference dataset
- Re-alignment of original datasets to reference dataset, which allowed use of tighter LC NET window (± 0.02 NET) during second comparative iteration.
- Automated alignment accomplished using MultiAlign software (developed in house), which employs MSWARP[4] algorithm
- After alignment of different datasets (± 3 ppm and ± 0.02 NET), a master list of common features is generated.

Figure 2.



The retention times of all datasets were transformed to a common scale by correlating elution time subsections of different datasets to each other and finding the transformation function of elution time that allows the best overlap of continuous subsections from different datasets.

- Metabolomic feature databases populated with monoisotopic masses and LC NETs for features reported in both manual alignment and automated dataset alignment outputs
- Comparative datasets from individual control and type 1 diabetic patient samples were then matched against the feature database using VIPER[3].

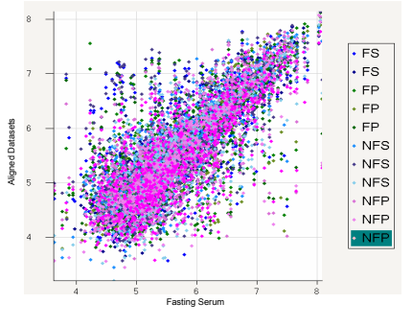
Results

Table 1.

	Manual method	Automated method
# Features available for alignment (n = 12)	3,334 ± 244	3,334 ± 244
# Database entries	1,973	2,216
# Features from control/patient datasets matching database (n = 52)	264 ± 98	410 ± 135

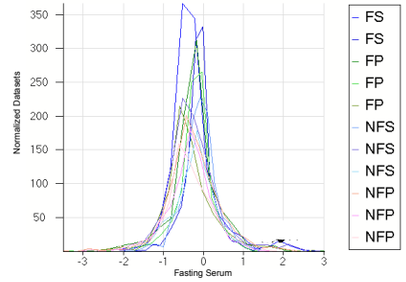
Comparison of plasma/serum metabolomic databases built from manually aligned datasets versus those datasets aligned using the MultiAlign software. The automated method reproducibly aligned a larger fraction of the data.

Figure 5.



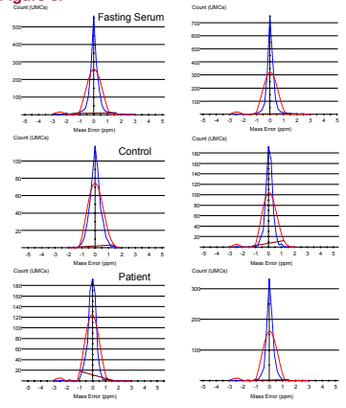
Alignment of fasting serum (FS), fasting plasma (FP), non-fasting serum (NFS), and non-fasting plasma (NFP) datasets using MultiAlign. Those features reproducibly observed in 3 of 3 replicates were added to the metabolomic plasma/serum database.

Figure 6.



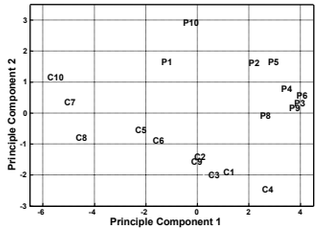
Intensity normalization of those datasets shown in Figure 3. Typical CVs for those features observed in 3 of 3 replicates were 25-35%.

Figure 3.



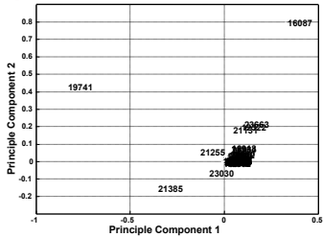
Mass error histograms for datasets constructed using the manual method (left) versus the database constructed using MultiAlign (right). The use of MultiAlign did not affect the quality of the data but resulted in a larger database available for matching to individual datasets.

Figure 7.



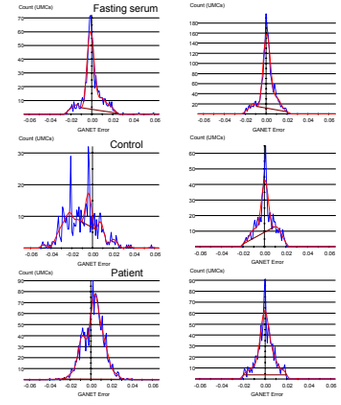
PCA scores plot showing segregation of control samples from type 1 diabetic samples. Only those features that were observed in all 10 control and/or all 10 patient datasets were used for the PCA.

Figure 8.



Corresponding PCA loads plot for the data shown in Figure 7. Eight features were found to be outliers in the loads plot.

Figure 4.



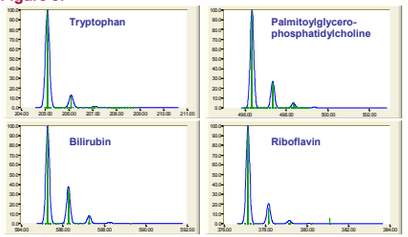
Corresponding NET error histograms for the datasets shown in Figure 1. The use of MultiAlign (right) improved the quality of data (NET error as metric) used to construct the plasma/serum metabolomic database.

Table 3.

Metabolite	Observed m/z	Predicted m/z	Mass error (ppm)
Tryptophan	205.0971	205.0977	2.9
Riboflavin	377.1456	377.1461	1.3
Myristoylglycerophosphatidylcholine	468.3090	468.3090	0.0
Pentadecanoylglycerophosphatidylcholine	482.3245	482.3246	-0.3
Palmitoleoylglycerophosphatidylcholine	494.3255	494.3246	-1.7
Palmitoylglycerophosphatidylcholine	496.3394	496.3403	1.8
Heptadecenoylglycerophosphatidylcholine	508.3414	508.3403	-2.2
Heptadecanoylglycerophosphatidylcholine	510.3558	510.3559	0.3
Linoleonylglycerophosphatidylcholine	518.3260	518.3246	-2.6
Linoleoylglycerophosphatidylcholine	520.3393	520.3403	1.9
Oleoylglycerophosphatidylcholine	522.3566	522.3559	-1.3
Stearoylglycerophosphatidylcholine	524.3711	524.3716	0.9
Arachidonoylglycerophosphatidylcholine	542.3246	542.3246	0.0
Eicosatrienoylglycerophosphatidylcholine	544.3387	544.3403	2.9
Eicosadienoylglycerophosphatidylcholine	546.3549	546.3559	1.9
Docosapentaenoylglycerophosphatidylcholine	568.3396	568.3403	1.2
Docosatetraenoylglycerophosphatidylcholine	570.3574	570.3559	-2.5
Bilirubin	585.2701	585.2713	2

Metabolomic features were tentatively identified based on accurate mass measurements, isotopic distribution (see Figure 9), targeted MS/MS experiments, and/or comparison to authentic standards.

Figure 9.



Comparison of theoretical isotopic distributions (blue) to observed isotopic distributions (green) for select features tentatively identified in LC-FTICR analyses of human plasma.

Conclusions

- MultiAlign provides more effective and efficient alignment of datasets from multiple LC-MS experiments compared to manual methods
- Type 1 diabetic plasma samples were significantly segregated from healthy control samples based on metabolomic fingerprint
- Significant difference in phospholipid profiles were observed between control and type 1 diabetic samples

Table 4.

Cluster ID	Compound ID	P value	Median CV (%)	⇅
9589*	Unknown	N/A	82	↑ in Patient, ND in Control
9591*	Unknown	N/A	37	↑ in Patient, ND in Control
12120	Unknown	0.001	31	↑ in Patient
12322*	Unknown	N/A	28	↑ in Patient, ND in Control
12678	Unknown	0.002	26	↑ in Patient
13336	Unknown	0.01	27	↑ in Patient
15188	Unknown	0.00002	31	↑ in Patient
16066	Unknown	0.0002	38	↑ in Patient
16587*	Unknown	N/A	31	↑ in Patient, ND in Control
16618*	Unknown	N/A	28	↑ in Patient, ND in Control
18870	Pentadecanoylglycerophosphatidylcholine	N/A	60	↑ in Control, ND in Patient
18971	Pentadecanoylglycerophosphatidylcholine	0.02	34	↑ in Control
19585	Palmitoleoylglycerophosphatidylcholine	N/A	16	↑ in Patient, ND in Control
19741*	Palmitoylglycerophosphatidylcholine	0.05	33	↑ in Control
20407	Heptadecanoylglycerophosphatidylcholine	0.09	33	↑ in Control
20527	Heptadecanoylglycerophosphatidylcholine	0.003	47	↑ in Control
21131*	Linoleoylglycerophosphatidylcholine	N/A	14	↑ in Patient, ND in Control
21255*	Oleoylglycerophosphatidylcholine	0.03	34	↑ in Control
21385	Stearoylglycerophosphatidylcholine	0.003	63	↑ in Control
22463	Eicosatrienoylglycerophosphatidylcholine	N/A	23	↑ in Patient, ND in Control
22589	Eicosatrienoylglycerophosphatidylcholine	N/A	10	↑ in Patient, ND in Control
22888	Eicosenoylglycerophosphatidylcholine	N/A	51	↑ in Patient, ND in Control
23030*	Unknown	0.009	28	↑ in Control
23663*	Unknown	N/A	32	↑ in Patient, ND in Control
28954*	Unknown	N/A	56	↑ in Patient, ND in Control

Candidate biomarkers of type 1 diabetes mellitus identified on qualitative or quantitative level. Tentative phospholipid identifications made based on accurate mass measurements, isotopic distributions, and/or targeted MS/MS experiments. *Note: Also marked as outlier in PCA loads plot.

Acknowledgements

This research was supported by NIH/NIDDK grant DK070146. Experimental portions of this research were performed in the Environmental Molecular Sciences Laboratory, a U.S. Department of Energy (DOE) national scientific user facility located at the Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is a multiprogram national laboratory operated by Battelle for the DOE under Contract No. DE-AC05-76RLO 1830.

References

1. Andreev, V.P., Rejtar, T., Chen, H.S., Moskovets, E.V., Ivanov, A.R. Karger, B.L. A universal denoising and peak picking algorithm for LC-MS based on matched filtration in the chromatographic time domain. *Anal. Chem.* 2003, *75*, 6314-6326.
2. Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R., Siuzdak, G. XCMS: Processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal. Chem.* 2006, *78*, 779-787.
3. Zimmer, J.S., Monroe, M.E., Qian, W.J., Smith, R.D. Advances in proteomics data analysis and display using an accurate mass and time tag approach. *Mass Spectrom Rev.* 2006, *25*, 450-482.
4. Jaitly, N., Monroe, M.E., Petyuk, V.A., Clauss, T., Adkins, J.N., Smith, R.D. A robust alignment algorithm for aligning liquid chromatography-mass spectrometry analyses in the AMT tag pipeline. *Anal. Chem.* 2006, accepted.

Contact Information

Thomas O. Metz, Ph. D.
Biological Sciences Division, K8-98
Pacific Northwest National Laboratory
P.O. Box 999, Richland, WA 99352
e-mail: thomas.metz@pnl.gov
Phone: (509) 376-8333