Aiding diagnosis of rare disease: applications of mass spectrometry-based metabolomics in the Undiagnosed Diseases Network

Thomas O. Metz1, Jennifer E. Kyle1, Kelly G. Stratton2, Jean-Philippe Gourdine3, Erika M. Zink1, Young-Mo Kim1, Heino M. Heyman1, Lindsey N. Anderson1, Jeremy D. Zucker1, Bobbie-Jo M. Webb-Robertson2, Katrina M. Waters1, Melissa A. Haende1, David M. Koeller4

1Biological Sciences Division, Pacific Northwest National Laboratory, Richland, WA, USA, 2Comprehensive & Statistical Analysis Division, Pacific Northwest National Laboratory, Richland, WA, USA, 3Department of Medical Informatics and Clinical Epidemiology and OHSU Library, Oregon Health & Science University, Portland, OR, USA, 4Department of Molecular & Medical Genomics, Oregon Health & Science University, Portland, OR, USA

Introduction

• U.S. law defines a rare disease as one that affects <200,000 individuals
• To address undiagnosed disease, the NIH has established the Undiagnosed Diseases Network (UDN): https://undiagnosed.hms.harvard.edu/

Overview

• 6% of the U.S. population suffers from a rare disorder that has evaded diagnosis
• We are performing mass spectrometry (MS)-based metabolomics analyses of plasma, cerebrospinal fluid (CSF), and urine from patients, as well as for Drosophila and zebrafish disease models
• To date, we have performed >2000 analyses of samples from healthy control individuals, patients and their first degree relatives

Methods

Design

• Patients accepted to the UDN have exhausted all possibilities for a diagnosis
• Once accepted, patients visit a clinical site for a 1-week evaluation that includes collection of blood, urine, and, if relevant, CSF samples

Approach

• Metabolites, proteins, and lipids are simultaneously extracted from plasma and CSF using the MFLPC protocol1
• Metabolites and proteins are extracted from urine using methanol
• Metabolites are chemically derivatized, analyzed using GC-MS and identified based on match of retention indices and fragmentation spectra to an augmented version of FiehnLib3
• Lipids are analyzed using LC-MS/MS and identified based on their fragmentation spectra using the in-house tool LIQUID4
• Patient profiles are compared to reference datasets to identify outlier metabolites or lipids

Results

Reference Datasets

Reference sample and UDN participant demographics

Analysis of samples and creation of reference data sets

Linking metabolomics and lipidomics results to patient genomes

Summary

• >2000 metabolomics and lipidomics analyses performed, including of 391 individuals with no known metabolic disease and of 83 UDN patients and first degree relatives
• Reference data sets for plasma, urine, and CSF were generated
• Data from UDN patients and first degree relatives were normalized against and compared to the appropriate reference data sets in order to identify outlier metabolites or lipids
• A human metabolic knockout model was created to predict downstream metabolic effects of gene variants
• Work continues to integrate patient genome effects with metabolome/lipidome data

Acknowledgements

This work was funded by NIH CCSC grant 1U18DK097222. Samples were collected as part of the Integrative Clinical and Genomic Evaluation in the Undiagnosed Patient (ICGEP) initiative at PNNL. We acknowledge support from the Department of Energy (DOE) Office of Biological and Environmental Research (BER) Plant and Microbial Science Program at PNNL. Metabolomics and lipidomics data were generated using University of Washington national scientific user facility on the UMNigma, a PNNL facility located at Battelle for Pacific Northwest National Laboratory. The authors would like to thank Mr. Matt Amsden for graphics support.

References

2. Burnum-Johnson et al., MPLeC: a method for simultaneous pathogen contamination and extraction of samples for multivariate mS. Anal Chem. 2017
5. Dunn et al., Procedures for high-scale metabolic profiling of liver, heart and brain using gas chromatography and high-resolution mass spectra. Nat Protoc. 2011