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

Thomas O. Metz[1], Jennifer E. Kyle[1], Kelly G. Stratton[2], Jean-Philippe Gourdine[3], Erika M. Zink[1], Young-Mo Kim[1], Heino M. Heyman[1], Lindsey N. Anderson[1], Jeremy D. Zucker[1], Bobbie-Jo M. Webb-Robertson[2], Katrina M. Waters[1], Melissa A. Haendel[3], David M. Koeller[4]

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

Pacific Northwest NATIONAL LABORATORY

*Proudly Operated by* **Battelle** *Since 1965*

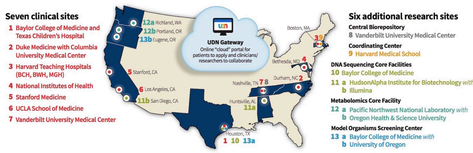
## Overview

- 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

## 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/)

**The Undiagnosed Diseases Network (UDN)**




- The UDN is comprised of a Coordinating Center, a Central Biorepository, 7 Clinical Sites, 2 Sequencing Cores, a Model Organisms Screening Center, and a Metabolomics Core (UDNMC) located at Pacific Northwest National Laboratory and Oregon Health & Science University (OHSU)
- The UDNMC is utilizing gas-chromatography (GC) and liquid chromatography (LC) coupled with mass spectrometry (MS) to perform untargeted metabolomics, lipidomics, and glycoproteomics analyses of samples from patients and their first degree relatives in the search for biomarkers

## Methods

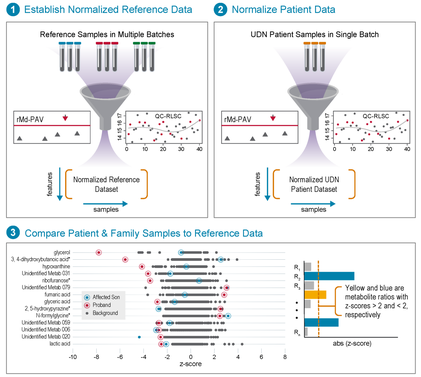
### Design

- 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
- To allow for proper statistical analysis of UDN patient data, we generated reference datasets from >391 individuals with no known metabolic disease and representative of UDN patient demographics

Dr. Stanley Nelson of the UCLA Clinical Site performs various tests on a UDN patient during the 1-week evaluation. Photo source: Kaiser Health News

### Approach

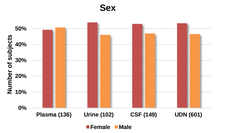
- Metabolites, proteins, and lipids are simultaneously extracted from plasma and CSF using the MPLEx protocol[1,2]. Metabolites and proteins are extracted from urine using methanol
- Metabolites are chemically derivatized, analyzed using GC-MS and identified based on match of retention indices and fragmentation spectra to an augmented version of FiehnLib[3]
- Lipids are analyzed using LC-MS/MS and identified based on their fragmentation spectra using the in-house tool LIQUID[4]
- Patient profiles are compared to reference datasets to identify outlier molecules using the workflow shown below[5]

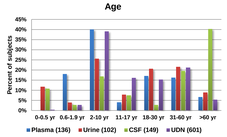



## Results

### Reference Datasets

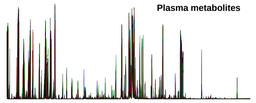
Reference sample and UDN participant demographics

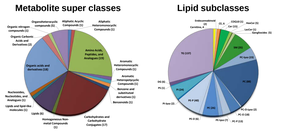


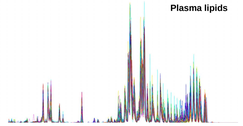

Reference samples from individuals with no known metabolic disease were received from the Vanderbilt University Metabolic Screening Laboratory, the Mayo Clinic Biochemical Genetics Laboratory, the Oregon Clinical & Translational Research Institute Biolibrary, and the OHSU Layton Aging and Alzheimer's Research Center.

Analysis of samples and creation of reference data sets



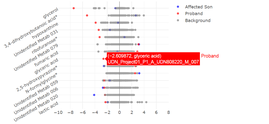




Over 300 unique identified and unidentified metabolites and >500 unique identified lipids were measured in plasma, urine, and CSF samples from 391 reference individuals. Shown are the distributions of identified plasma metabolites and lipids among superclasses and subclasses, respectively.




Overlaid chromatograms from 300 GC-MS-based metabolomics analyses (top frame) and 297 LC-MS-based lipidomics analyses (bottom frame) of reference and QC plasma samples. The chromatograms are shown prior to alignment.

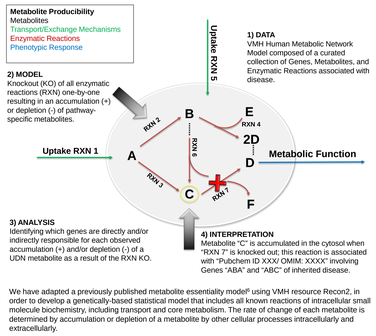
Comparison of UDN proband to reference data set




Data from the reference samples are normalized and used to create reference data sets, against which data from UDN probands, and any relatives, are normalized and compared. Shown is a screenshot from a representative interactive z-score plot of plasma metabolomics data. The user can hover over individual points to display additional information.

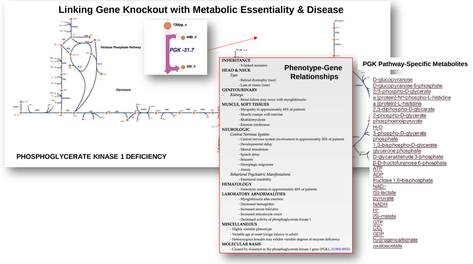
### Linking metabolomics and lipidomics results to patient genomes

A human metabolic knockout model has been constructed to predict downstream effects




We have adapted a previously published metabolite essentiality model[6] using VMH resource Recon2, in order to develop a genetically-based statistical model that includes all known reactions of intracellular small molecule biochemistry, including transport and core metabolism. The rate of change of each metabolite is determined by accumulation or depletion of a metabolite by other cellular processes intracellularly and extracellularly.

Phosphoglycerate kinase 1 deficiency




In this hypothetical example, the downstream metabolic effects of a deficiency in phosphoglycerate kinase 1 are predicted, as well as the likely associated patient phenotype.

## 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 and metabolome/lipidome data

## Acknowledgements

This work was funded by NIH NCATS grant U01 TR001395. Samples were analyzed using capabilities developed in part by support of the NIH National Institute of General Medical Sciences (8 P41 GM103493-10) and the U.S. Department of Energy (DOE) Office of Biological and Environmental Research (OBER) Pan-omics program at PNNL. Metabolomics and lipidomics data were generated in the Environmental Molecular Sciences Laboratory, a DOE OBER national scientific user facility on the PNNL campus. PNNL is a multi-program national laboratory operated by Battelle for the DOE. The authors would like to thank Mr. Nathan Johnson for graphics support.

## References

1. Nakayasu *et al.*, MPLEx: a Robust and Universal Protocol for Single-Sample Integrative Proteomic, Metabolomic, and Lipidomic Analyses. *mSystems*. 2016
2. Burnum-Johnson *et al.*, MPLEx: a method for simultaneous pathogen inactivation and extraction of samples for multi-omics profiling. *Analyst.* 2017
3. Kind *et al.*, FiehnLib: mass spectral and retention index libraries for metabolomics based on quadrupole and time-of-flight gas chromatography/mass spectrometry. *Anal Chem.* 2009
4. Kyle *et al.*, LIQUID: an-open source software for identifying lipids in LC-MS/MS-based lipidomics data. *Bioinformatics*. 2017
5. Dunn *et al.*, Procedures for large-scale metabolic profiling of serum and plasma using gas chromatography and liquid chromatography coupled to mass spectrometry." *Nat Prot.* 2011
6. Imieliński *et al.*, Investigating metabolite essentiality through genome-scale analysis of Escherichia coli production capabilities. *Bioinformatics.* 2005

**Career Opportunities:** For potential openings in the Integrative Omics Group at PNNL please visit http://omics.pnl.gov/careers

**CONTACT: Thomas O. Metz, Ph.D.**
Biological Sciences Division
Pacific Northwest National Laboratory
**E-mail: thomas.metz@pnnl.gov**

**omics.pnl.gov**