

## EXPANDED METABOLOMICS METHODS AND OTHER TECHNICAL DETAILS

In keeping with consensus reporting standards for metabolomics experiments,<sup>1</sup> we provide in-depth metabolomics details as follows. In addition, we provide expanded details on statistical analyses and other technical details:

### Sample Collection and Processing:

Blood samples were collected immediately before transplant into K<sub>2</sub>EDTA plasma preparation tubes. Once samples were collected from patients, they were immediately placed on ice. The samples remained on ice and were centrifuged at 2500 x g for 15 minutes at 4°C within 5 hours of collection (the mean time from collection to processing for all samples = 1.1 hours). Following centrifugation, three aliquots were made from the supernatant, which were stored at -80°C shortly after.

### Metabolomics Profiling with Liquid Chromatography-Mass Spectrometry (LC-MS)

High-resolution metabolomics was performed in the Emory Clinical Biomarkers Laboratory as previously described.<sup>2</sup> Acetonitrile containing internal standards was added to thawed plasma samples at a 2:1 ratio. Samples were then vortexed and allowed to incubate on ice for 30 minutes. Following incubation, samples were vortexed and then centrifuged 20,817 x g for 20 minutes. Supernatants were then transferred into new microcentrifuge vials. An aliquot of these samples was then diluted 8-fold in acetonitrile into autosampler vials. These were loaded onto autosampler plates and were kept at -20°C until analysis. Pooled control plasma (Equitech-Bio, SHP45), National Institute of Standards and Technology (NIST) reference plasma, and plasma from individual control donors were prepared and analyzed in tandem.

Metabolomics analysis was conducted with a liquid chromatography-mass spectrometry system (LC-MS) consisting of a Vanquish Duo UHPLC coupled to an Orbitrap ID-X Tribrid Mass Spectrometer (Thermo Scientific). Autosampler plates containing samples were maintained at 4°C in the autosampler of the LC-MS throughout the analysis. Samples were analyzed in triplicate on a 5-minute method using C18 chromatography coupled to negative electrospray ionization (ESI) (C18-) and hydrophilic interaction liquid chromatography coupled to positive ESI (HILIC+). Analyte separation for HILIC was performed with a Waters Acquity BEH Amide HILIC column (2.1 mm x 100 mm, 1.7 µm particle size) and gradient elution with LCMS-grade solvents and additives. The HILIC mobile phases included (Buffer A) water with 1 mM ammonium acetate and 0.1% formic acid and (Buffer B) 95% acetonitrile with 1 mM ammonium acetate and 0.1% formic acid. For the HILIC gradient, an initial 0.5 min hold at 90% B was followed by a linear decrease to 20% B from 0.6 to 2.55 min, a 2 min hold, and a 5-minute re-equilibration period. C18 chromatography was performed on a Thermo Hypersil Gold C18 column (2.1 mm x 100 mm, 1.9 µm particle size). The C18 mobile phases included (Buffer A) water with 1 mM ammonium acetate and (Buffer B) 99% acetonitrile with 1 mM ammonium acetate. For the C18 gradient, an initial 0.5 min hold at 1% B was followed by a 0.75 min linear increase to 99% B, held for 3.75 min, and a 5-minute re-equilibration period. The flow rate for both methods was 0.3 mL/min, and the column compartment was heated to 45 °C. The mass spectrometer was operated at 120k resolution and MS<sup>1</sup> scans were collected for *m/z* 85-1,275. Tune parameters consisted of sheath gas at 50, auxiliary gas at 10, and sweep gas at 1. The spray voltage was set to 3.50 kV for ESI+ and -2.75 kV for ESI-.

### Feature Extraction, Annotation, and Identification

ProteoWizard v3<sup>3</sup> was used to convert raw spectral files to .mzXML files. Untargeted feature tables were then generated from extracted data using established laboratory workflows and softwares<sup>4-6</sup> (namely, apLCMS and xMSAnalyzer). These feature tables comprised all HILIC+ and C18- features observed with a unique mass to charge ratio (*m/z*) and corresponding retention time (RT). Targeted feature tables were generated using A) an in-house annotation algorithm, MSMICA (see <https://github.com/jamesjiadazhan> for R package releases within the year), which

generates annotations based on precursor-product correlations (note: this version of the algorithm chose a single best adduct for the feature table based on the strength of the precursor-product correlation alone), or B) by matching observed  $m/z$ 's and retention times to the  $m/z$ 's and retention times of a library of authentic standards previously run on the same mass spectrometry instrument. Level 1<sup>7</sup> identification of metabolites was obtained by ion dissociation mass spectrometry relative to authentic standards (mass error =  $\pm 5$  ppm, retention time threshold =  $\pm 30$  seconds), and other metabolites identified by MSMICA were considered Level 3<sup>7</sup> identifications. For algorithmically identified features (Level 3) and features identified based on standards (Level 1), metabolites that were dually identified in HILIC+ and C18- modes were both retained for statistical analysis and treated as independent observations. If following statistical analysis (descriptive statistics or t-tests in **Figure 3**), an identification that was significantly different between groups had an  $m/z$  and retention time that mapped to multiple possible metabolites and/or adducts, this was indicated or accounted for on the graph as appropriate.

#### Data Structuring and Transformation

HILIC+ and C18- untargeted feature tables were merged, with features being labeled with their column (HILIC or C18),  $m/z$ , and retention time, delimited by underscores (i.e., HILIC\_285.8142245\_24.93741273 or C18\_335.1135917\_151.5102926). The full untargeted feature table with all HILIC+ and C18- features can be found on the accompanying Metabolomics Workbench. All '0' values (indicating non-detection by the instrument) were replaced with  $\frac{1}{2}$  the minimum value for that feature detected across all samples.<sup>8</sup> Following this, all spectral intensity values were  $\log_2$  transformed prior to further analysis. For the full targeted feature tables and associated metadata, see **Supplementary Material 3**.

#### Metabolomics Statistical Analysis and Data Visualization

All data were analyzed and visualized using R (version 4.4.0), MetaboAnalyst (version 6.0),<sup>9</sup> or GraphPad Prism (version 10.4.0) for Mac (GraphPad Software, Boston, Massachusetts). All code was written, compiled, and run in Visual Studio Code (Microsoft). For the full source code, see **Supplementary Material 4**. For a full list of all R packages used, see the 'Dependencies' header in the source code. Microsoft Excel was used for compiling and organizing both raw and structured data. For all analyses, statistical significance was set at  $\alpha = 0.05$ .

All partial least squares discriminant analysis (PLS-DA) was performed using the R package mixOmics<sup>10</sup> and visualized using ggplot2.<sup>11</sup> All heatmaps were created based on untargeted feature tables using MetaboAnalyst's<sup>9</sup> 'Statistical Analysis [one factor]' function. Prior to the creation of the heatmap, a variance filter based on interquartile range was applied to eliminate non-informative features (40% of features filtered out); however, no abundance or reliability filters were applied. No further transformation was performed as the data had already been  $\log_2$  transformed; however, all features were scaled with z-scores for purposes of visualization on the heatmap. The top 250 features most differentiating features (based on t-test) between groups were displayed on the heatmap. Both samples and features underwent hierarchical clustering using a Ward method and Euclidean distance. Finally, the volcano plot in Figure 1 was created by calculating p-values (based on t-tests on transformed data) to compare the means of each feature between groups, while fold changes were determined using untransformed data. All p-value calculations were performed using R's *stats* package. For purposes of visualization, p-values were then  $-\log_{10}$  transformed, and fold changes were  $\log_2$  transformed.

All pathway enrichment analysis was performed using Mummichog.<sup>12</sup> A full list of p-values comparing means of all features between groups was first generated in R using the *stats* package via t-tests. These features, including the  $m/z$ , retention time, and ESI mode, along with their corresponding p-values, were then analyzed using MetaboAnalyst's<sup>9</sup> 'Functional Analysis' tool for LC-MS, which utilizes Mummichog<sup>12</sup> as its core algorithm. Notably, this tool allows for the usage of Mummichog v2.0, which takes into account retention time and ESI mode, whereas the original

algorithm simply used  $m/z$ . The parameters for the pathway enrichment included a mass tolerance of 10 ppm, mixed mode, retention time presence (seconds), enforcement of primary ions, and a p-value cutoff of 0.1. The Kyoto Encyclopedia of Genes and Genomes (KEGG) *homo sapiens* metabolic database was chosen as the pathway library. All results were then downloaded, and the enrichment factor for each pathway was then calculated manually by dividing the number of significant hits by the number of expected hits. A bubble plot was then constructed based on the results using the p-value (gamma) and the enrichment factor.

For targeted data, t-tests were employed to determine the p-values comparing means between each group. The top 5 upregulated and downregulated identifications (i.e., higher or lower in PGD) were determined based on the lowest p-values. All targeted features were classified via ClassyFire,<sup>13</sup> which generated kingdom, superclass, class, and subclass categorizations for each metabolite. Features that were significantly upregulated or downregulated were then counted within each superclass and class and graphed in a stacked bar graph. For the purposes of visualization, any class that only had one significantly different chemical in it was excluded from the graph.

### Sensitivity Analysis

Given the exploratory and hypothesis-generating nature of the untargeted metabolomics study performed herein, which measured over 20,000 individual metabolic features, a conventional power analysis is not applicable. However, we conducted a quantitative *post-hoc* sensitivity analysis (further details of which can be found in the GitHub repository) to determine the minimum detectable effect size based on the group sizes in our study. With 62 patients (8 with severe PGD and 54 without), we had ~80% power ( $\alpha=0.05$ ) to detect a minimum standardized mean difference of Cohen's  $d \approx 1.08$ . Based on observed within-group variance (calculated using the filtered, i.e. post-QC, untargeted feature table), the median minimum detectable fold-change was  $1.87\times$  (IQR  $1.46\text{--}2.99\times$ ) across all features. Considering this, this study was powered to detect very large metabolic differences between groups, which is consistent with the exploratory and hypothesis-generating design. More subtle differences would require larger cohorts moving forward.

## REFERENCES

1. Sumner LW, Amberg A, Barrett D, et al. Proposed minimum reporting standards for chemical analysis: Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics*. 2007;3(3):211-221. doi:10.1007/s11306-007-0082-2
2. Liu KH, Nellis M, Uppal K, et al. Reference Standardization for Quantification and Harmonization of Large-Scale Metabolomics. *Anal Chem*. 2020;92(13):8836-8844. doi:10.1021/acs.analchem.0c00338
3. Chambers MC, Maclean B, Burke R, et al. A cross-platform toolkit for mass spectrometry and proteomics. *Nat Biotechnol*. 2012;30(10):918-920. doi:10.1038/nbt.2377
4. Uppal K, Soltow QA, Strobel FH, et al. xMSanalyzer: automated pipeline for improved feature detection and downstream analysis of large-scale, non-targeted metabolomics data. *BMC Bioinformatics*. 2013;14(1):15. doi:10.1186/1471-2105-14-15
5. Jarrell ZR, Smith MR, Hu X, et al. Plasma acylcarnitine levels increase with healthy aging. *Aging*. 2020;12(13):13555-13570. doi:10.18632/aging.103462
6. Yu T, Park Y, Johnson JM, Jones DP. apLCMS—adaptive processing of high-resolution LC/MS data. *Bioinformatics*. 2009;25(15):1930-1936. doi:10.1093/bioinformatics/btp291
7. Schymanski EL, Jeon J, Gulde R, et al. Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence. *Environ Sci Technol*. 2014;48(4):2097-2098. doi:10.1021/es5002105
8. Wei R, Wang J, Su M, et al. Missing Value Imputation Approach for Mass Spectrometry-based Metabolomics Data. *Sci Rep*. 2018;8(1):663. doi:10.1038/s41598-017-19120-0
9. Pang Z, Lu Y, Zhou G, et al. MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation. *Nucleic Acids Res*. 2024;52(W1):W398-W406. doi:10.1093/nar/gkae253
10. Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. Schneidman D, ed. *PLoS Comput Biol*. 2017;13(11):e1005752. doi:10.1371/journal.pcbi.1005752
11. Wickham H. *Ggplot2: Elegant Graphics for Data Analysis*. Second edition. Springer; 2016.
12. Li S, Park Y, Duraisingham S, et al. Predicting Network Activity from High Throughput Metabolomics. Ouzounis CA, ed. *PLoS Comput Biol*. 2013;9(7):e1003123. doi:10.1371/journal.pcbi.1003123
13. Djoumbou Feunang Y, Eisner R, Knox C, et al. ClassyFire: automated chemical classification with a comprehensive, computable taxonomy. *J Cheminform*. 2016;8(1):61. doi:10.1186/s13321-016-0174-y