Metabolic perturbations in classic galactosemia beyond the Leloir pathway: insights from an untargeted metabolomic study

S. Taylor Fischer¹, Allison B. Frederick¹,* ViLinh Tran², Shuzhao Li², Dean P. Jones², and Judith L. Fridovich-Keil¹,+  
¹Department of Human Genetics, Emory University School of Medicine, Atlanta, GA, USA  
²Clinical Biomarkers Laboratory, Division of Pulmonary, Allergy and Critical Care Medicine, Department of Medicine, Emory University School of Medicine, Atlanta, GA USA

Summary

Classic galactosemia (CG) is an autosomal recessive disorder that impacts close to 1/50,000 live births in the United States, with varying prevalence in other countries. Following exposure to milk, which contains high levels of galactose, affected infants may experience rapid onset and progression of potentially lethal symptoms. With the benefit of early diagnosis, generally by newborn screening, and immediate and lifelong dietary restriction of galactose, the acute sequelae of disease can be prevented or resolved. However, long-term complications are common, and despite many decades of research the bases of these complications remain unexplained. As a step toward defining the underlying pathophysiology of long-term outcomes in CG, we applied an untargeted metabolomic approach with mass spectrometry and dual liquid chromatography, comparing thousands of small molecules in plasma samples from 183 patients and 31 controls.

*Correspondence to: Judith L. Fridovich-Keil, Department of Human Genetics, Emory University School of Medicine, Rm. 325.2 Whitehead Bldg., 615 Michael St, Atlanta, GA 30322 TEL 404-727-3924, FAX 404-727-3949, jfridov@emory.edu.  
*Present address: University of South Carolina School of Medicine Greenville, Greenville, SC, USA  
Contributions of individual authors:  
Taylor Fischer performed all of the statistical analyses, generated all tables and figures, and contributed to writing and editing of the manuscript.  
Allison Frederick helped assemble the plasma samples for mass spectrometry and the spreadsheet of outcome and covariate data for analysis, and contributed to editing of the manuscript.  
ViLinh Tran conducted the mass spectrometry on all samples and contributed to editing of the manuscript.  
Shuzhao Li contributed to study design, data analysis, and editing of the manuscript.  
Dean Jones contributed to study design, oversaw all aspects of the mass spectrometry, and contributed to editing of the manuscript.  
Judith Fridovich-Keil designed the overall project, contributed to data collection, coordinated the activities of the other authors, and wrote most of the manuscript. Guarantor.

Competing interest statement:  
Taylor Fischer declares that she has no conflict of interest.  
Allison Frederick declares that she has no conflict of interest.  
ViLinh Tran declares that she has no conflict of interest.  
Shuzhao Li declares that he has no conflict of interest.  
Dean Jones declares that he has no conflict of interest.  
Judith Fridovich-Keil declares that she has no conflict of interest.

Ethics approval: The work was conducted with approval of the Emory University Institutional Review Board (Protocol 00024933, PI: JL Fridovich-Keil). Further, all procedures followed were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1975, as revised in 2000. Informed consent was obtained from all patients for being included in the study.

Patient consent statement: All participants in this study were consented to protocol Emory IRB 00024933 in accordance with Emory IRB policy.

Animal Rights (IACUC): This article does not contain any studies with animal subjects performed by the any of the authors.
patients were on galactose-restricted diets. Using both univariate and multivariate statistical methods, we identified 252 differentially abundant features from anion exchange chromatography and 167 differentially abundant features from C18 chromatography. Mapping these discriminatory features to putative metabolites and biochemical pathways revealed 14 significantly perturbed pathways; these included multiple redox, amino acid, and mitochondrial pathways, among others. Finally, we tested whether any discriminatory features also distinguished cases with mild versus more severe long-term outcomes and found multiple candidates, of which one achieved FDR-adjusted q<0.1. These results extend substantially from prior targeted studies of metabolic perturbation in CG and offer a new approach to identifying candidate modifiers and targets for intervention.

1 sentence summary: Plasma samples from patients with classic galactosemia on dietary restriction of galactose demonstrate hundreds of metabolic features with differential abundance between cases and controls.

Keywords
galactosemia; metabolomics; metabolic features; plasma; metabolic pathways

Introduction

Patients with classic galactosemia (CG) are at dramatically increased risk for a constellation of long-term complications despite neonatal detection and careful lifelong dietary restriction of galactose (Berry 2014). These include cognitive, speech, movement, and behavioral disabilities in many patients, and premature ovarian insufficiency in >80% of young women (Berry 2014). The mechanisms underlying these complications remain unclear (Segal 1995; Fridovich-Keil 2006) hampering efforts to develop more effective interventions. Prior studies have addressed the question of mechanism by exploring metabolites in patient samples (e.g. (Yager et al 2003; Berry 2011)) and model systems (e.g. (Ning et al 2001; Kushner et al 2010; Tang et al 2014)) looking for perturbations that associate with disease outcome. The results of these studies have been limited, however, as many have looked at perturbations only following galactose exposure, and most have focused almost entirely on metabolites directly linked to the Leloir pathway (Figure 1).

The association of elevated levels of metabolites such as galactose, galactose-1-phosphate (gal1P), galactitol, and galactonate with acute galactose exposure in CG are clear, but which, if any, of these metabolites associate with long-term outcome severity remains unknown. For example, a striking elevation of gal1P in red blood cells (RBC) is almost universally seen in affected infants drinking milk (Berry 2014) but, following dietary restriction of galactose, gal1P drops to near-normal levels, and association of elevated RBC gal1P with long-term outcomes is controversial. Specifically, Waggoner et al (1990), Schweitzer et al (1993), and Hughes et al (2009) all found no significant association of outcome severity with gal1P in their cohorts of patients, while Guerrero et al (2000) and Yuzyuk et al (2018) both did. The explanation for these apparent conflicts remains unclear but may reflect, at least in part, the relatively small sizes of some of the cohorts.
Further, if the goal is to understand the disease process, then whether or not gal1P or another galactose metabolite associates with negative long-term outcomes in CG still skirts the question of mechanism. Presumably, perturbations in the Leloir pathway lead to disturbances in other pathways that lead to disturbances in yet other pathways, and one or more of these direct or indirect consequences of galactose-1-phosphate uridylyltransferase (GALT) deficiency causes the long-term complications associated with CG. Mechanistic understanding therefore requires delineation of these other pathways. As a first step toward that goal, we conducted an untargeted high-resolution metabolomic (HRM) analysis of plasma samples from 183 patients with CG, all on long-term dietary restriction of galactose, and 31 controls. The results, presented here, serve as a proof of principle that untargeted metabolomic studies can give insight into the broad metabolic consequences of treated GALT-deficiency and offer previously unseen clues to mechanism, modifiers, and novel candidate targets for intervention.

Materials and Methods

Study volunteers:

Cases for this study were selected from among volunteers already enrolled in our longitudinal protocol “Bases of Pathophysiology and Modifiers of Outcome in Galactosemia” (Emory IRB00024933; PI: JL Fridovich-Keil), which has been continuously approved by the Emory Institutional Review Board, or its predecessor, since 1992. Children and adults with CG already enrolled in the longitudinal study were selected for inclusion here based on availability of frozen plasma samples. Controls were adults recruited from the Emory University community to match, as closely as possible, the demographic makeup of cases in the study (Table 1). Controls were unrelated to cases. All plasma samples used in this study derived from blood collected by venipuncture into sodium-heparin vacutainer tubes. We analyzed a single sample from each volunteer in the study.

Outcomes and covariates:

Outcomes considered in the current study included whether subjects, ages 6 years and older, received special educational services or speech therapy in elementary school and, for girls and women ages 2–33 years, whether plasma anti-Müllerian hormone (AMH) was detected at ≥0.1 or <0.1 ng/mL. Covariates collected for this study included subject gender, age at blood draw, race and ethnicity, and years the plasma sample was in frozen storage (at −80°C) before being thawed for this study. Those covariates demonstrating a statistically significant difference in distribution (explained below) between outcome groups, i.e. cases and controls, were adjusted for in subsequent linear models.

High-resolution liquid chromatography-mass spectrometry

Samples were analyzed as previously described (Soltow et al 2013; Jones et al 2016). Specifically, each plasma sample was analyzed in three technical replicates on a Thermo Scientific LTQ Velos Orbitrap mass spectrometer, coupled with dual liquid chromatography, alternating data collection between AE and C18 columns. Analyses were performed with positive electrospray ionization mode, an injection volume of 10 μL, mass-to-charge ratio.
(m/z) scan range of 85 to 2,000, and resolution of 60,000 (full width at half maximum [FWHM]).

Samples were run in batches of 20, randomized for case/control status, with pooled reference plasma (Q-Standard) samples included prior to and following each batch of 20 to enable quality control and normalization among batches, as described previously (Go et al 2015). Data extraction was performed using apLCMS Version 6.1.3 (Yu et al 2009) and xMSanalyzer Version 2.0.7 (Uppal et al 2013). We performed principal component analysis (PCA) to evaluate potential batch effects (Yang et al 2008) and corrected for these effects where necessary using ComBat (Johnson et al 2007) and xMSanalyzer.

The resulting data tables contained individual features defined by accurate mass m/z, retention time (RT), and ion intensity. We averaged the non-zero intensities of technical replicates and performed log₂ transformation to normalize ion intensities. Confirmed metabolite identities were based on accurate mass m/z, co-elution with authentic standards, and MS/MS criteria (Go et al 2015; Go et al 2015; Jones et al 2016).

**Statistical analysis**

First, we performed a PCA to visualize the variation in feature intensity patterns across all plasma samples (Jolliffe 2002) (Supplemental Figure 1). Six samples, all from CG cases, fell outside ellipses representing 95% group-wise confidence intervals for data from both columns and were therefore identified as outliers and excluded from further analysis; this yielded a final study sample of 214 subjects (n=183 cases, n=31 controls). We summarized and compared patient characteristics between CG and control study volunteers using the non-parametric Fisher’s exact test for categorical variables and the Mann-Whitney U test for continuous variables (Table 1).

We filtered metabolomic features from both the C₁₈ and AE columns to retain only those with at least 80% non-missing values across samples of either group. We then imputed missing data, i.e. ion intensities falling below the limit of detection of the instrument, to half of the lowest intensity registered for a given feature across all biological and quality control samples (Xia and Wishart 2011), and log₂-transformed feature intensities. We used histograms and kernel density plots to visually confirm that the overall intensity distribution for both the AE and C₁₈ datasets showed the characteristic bell curve of a normal distribution following preprocessing (Xia and Wishart 2011). Furthermore, we used boxplots to confirm that the intensity distribution within each of the 214 plasma samples after missing value imputation and log₂-transformation were similar in terms of median, interquartile range, and overall range. Following logarithmic transformation and imputation of missing data the feature intensity data were approximately normally distributed.

We used a combination of univariate and multivariate statistical methods to identify features with differential abundance between cases and controls (Uppal et al 2017). First, for each individual feature, we fit a linear regression model of log₂-transformed intensities that included CG status, the potential confounders identified by tests described above (gender, age, storage time), and batch number as predictors. We adjusted the resulting p-values for significance of the CG status term for multiple comparisons using the Benjamini-Hochberg
method (Benjamini and Hochberg 1995). Here, the number of comparisons was equal to the number of features that remained after filtering for group-wise presence of ≥80%. As an alternate feature selection approach, we performed partial least squares discriminant analysis (PLSDA), a supervised multivariate statistical technique (Wold et al 2001). Feature intensities were standardized to mean zero and unit variance prior to PLSDA analysis. The features that (1) achieved false discovery rate (FDR) q<0.2 in the univariate approach and also (2) achieved variable importance in projection (VIP) scores >2 in the PLSDA model were considered discriminatory. Of note, we applied a Benjamini-Hochberg cut-off of FDR<0.2 to balance the very stringent VIP threshold used for multivariate analysis with PLSDA. We performed two-way hierarchical cluster analysis (HCA) of selected discriminatory features using R version 3.4.0 (Team 2017).

Among CG subjects only, we further tested each discriminatory feature for association with binary AMH (≥0.1 or <0.1 ng/mL) using the Mann-Whitney U test. We used the Mann-Whitney U test in this instance because the smaller size of this subset of samples left open the possibility that some feature distributions might fail to meet the parametric assumptions of the t-test. Finally, we also tested each discriminatory feature for association with receipt of special educational services/speech therapy in elementary school, also using the Mann-Whitney U test for the same reasons described above. For these two analyses, we adjusted the resulting p-values for multiple comparisons by the Benjamini-Hochberg FDR method, whereby the number of comparisons performed equaled the number of discriminatory features selected above by both linear regression and PLSDA approaches.

Pathway analysis and annotation of discriminatory features

Features identified from the PLSDA and linear regression models as discriminating between cases and controls with both VIP>2 and FDR<0.2, respectively, were selected for pathway enrichment analysis using Mummichog version 1.0.5 (Li et al 2013). We considered pathways to differ between groups when p<0.05. We further annotated discriminatory features using xMSannotator version 1.3.2 (Uppal et al 2017), first identifying accurate-mass matches to known metabolites in the Human Metabolome Database (HMDB) (Wishart et al 2007) on the basis of monoisotopic mass m/z with a mass error threshold of ±10 ppm (Level 5 identification by the criteria of (Schymanski et al 2014)). We considered multiple adducts of each metabolite (e.g., M+H, M+2H, M+H+NH₄⁺, M+ACN+2H, M+2ACN+2H, M+NH₄⁺, M+Na, M+ACN+H, M+ACN+Na, M+2ACN+H, 2M+H, 2M+Na, 2M+ACN+H, M+2Na-H, M+H-H₂O, M+H-2H₂O) when first searching HMDB prior to confidence assignment. To assign confidence levels to top annotations, we used xMSannotator’s multilevel scoring algorithm by which candidate matches are graded on the basis of multiple parameters (e.g., presence of multiple adducts, retention time similarity across adducts, intensity correlations between samples, and biochemical pathway knowledge). We accepted only levels 2 (medium confidence) and 3 (high confidence). Identities of selected metabolites (e.g., amino acids, creatine, acylcarnitine) in pathways were confirmed by co-elution with authentic standards and Level 1 (confirmed structure) or Level 2 (probable structure) MS/MS criteria as defined by (Schymanski et al 2014)).
Results

Identifying plasma features that distinguish cases from controls

Of the 11,234 features detected by HRM and AE chromatography of our 214 plasma samples, 7,711 features met the filtering protocols explained in Materials and Methods. Similarly, 7,458 of the original 11,215 features resolved by C18 chromatography remained after filtering. To be clear, these sets are expected to overlap, and though extensive, are not expected to be comprehensive representations of the plasma metabolome.

To select plasma features that distinguished cases from controls, we applied both a univariate linear regression approach, adjusting for covariates, and a multivariate PLSDA approach. In the univariate analysis, we found 742 AE and 254 C18 features that met the less stringent threshold of FDR<0.2. Given our group sizes were unbalanced, we assessed the 10-fold cross-validation accuracy of these features using balanced classification rate (BCR). The AE features achieved a mean BCR of 0.91, and the C18 features achieved a mean BCR of 0.97. In the multivariate analysis, we found 343 AE features (mean BCR=0.98) and 320 C18 features (mean BCR=1) that met the more stringent threshold of VIP>2 for PLSDA. Overall, we found that 252 AE features and 167 C18 features were differentially abundant between cases and controls by both approaches. This result alone confirms that while >95% of metabolic features detected between these two columns did not differ substantially between cases and controls, the differences that were detected extended well beyond the Leloir pathway, which includes only a handful of metabolites (Figure 1). Additionally, the direction and magnitude of covariate effects on metabolite abundance was varied (Supplemental Figure 2) as shown by point estimates of the regression coefficients for covariate terms relative to the CG status term in the linear models.

To visualize the abundance patterns of discriminatory features between cases and controls, we performed two-way HCA in which (1) features were clustered by how similar their intensity levels were among the different plasma samples, and (2) subjects were clustered by how similar their intensity levels were for the different features. Heat maps from both the AE (Supplemental Figure 3) and C18 (Supplemental Figure 4) data sets showed strong clustering of cases separately from controls, demonstrating reproducibility of the patterns within each cohort.

Next, we visualized the distribution of discriminatory features from both columns with regard to m/z and retention time using Manhattan plots (Figure 2 and Supplemental Figure 5). The dashed line in each panel indicates the threshold of FDR=0.2. Features meeting the FDR threshold (FDR<0.2) in each panel that also demonstrated VIP>2 are colored blue; features that demonstrated FDR<0.2 but failed to meet the VIP cut-off are colored red. As expected, assuming discriminatory features share general chemical properties with other molecules in the sample, reflected as m/z and retention time, the distributions of discriminatory (blue/ red) and non-discriminatory (grey) features in each graph appear similar by visual inspection.
**Identifying metabolic pathways that distinguish cases from controls**

While we successfully identified some metabolites by comparing accurate mass \( m/z \) and RT to our confirmed reference library, we also sought to assign high-confidence annotations to the remaining discriminatory features. To do this, we used xMSannotator (Uppal et al 2017) to first map each discriminatory feature to candidate matches from the Human Metabolome Database (HMDB) (Wishart et al 2007) on the basis of accurate mass \( m/z \) and then scored the strength of each candidate match, as explained in Materials and Methods. These annotations served to supplement the initial annotation results from Mummichog.

As illustrated by coloration of the heat maps in Supplemental Figures 3 and 4, some discriminatory features from both the AE and C\(_{18}\) columns were found at elevated levels in cases, while others were found at diminished levels in cases. Examples of metabolites mapped to features that were detected at elevated levels in cases included glutamic acid and creatine; examples of metabolites mapped to features detected at diminished levels in cases included arginine and tetradecanoylcarnitine (Figure 3). A list of all discriminatory features detected in both columns is presented in Supplemental Table 1, and box plots of differential abundance between cases and controls for each of these features are provided in Supplemental File 1.

To help put the many discriminatory metabolites detected in cases and controls into context, we next used Mummichog (Li et al 2013) to map these metabolites to biochemical pathways. We considered pathways to differ between groups when \( p \leq 0.05 \). Table 2 presents the 14 pathways that met this threshold and also included at least four overlap metabolites. Pathway size represents the total number of metabolites detected for a given pathway, and overlap size represents the number of these metabolites that were found to differ significantly between CG and control subjects. A list of all pathways identified by Mummichog with \( p \leq 0.05 \) is presented in Supplemental Table 2.

**Identifying features that distinguish among cases by long-term outcome severity**

Finally, we tested whether any of the metabolic features that discriminated between cases and controls in our study also distinguished among cases by outcome severity. We focused on two outcomes: (1) for all cases age 6 years and older, whether the participant received any special educational services, including speech therapy, in elementary school, and (2) for girls and women only, whether plasma AMH, a marker of ovarian status, was <0.1ng/mL or \( \geq 0.1\)ng/mL (Frederick et al 2018). Of note, we restricted the first outcome marker to volunteers who had received special services in elementary school to minimize potential complication from families who enrolled their pre-school child in preventative intervention.

In the first comparison, which included 114 cases for whom we had data on receipt of special educational services/speech therapy (n=65 no, n=49 yes), we found eight features that nominally associated with the outcome \( (p<0.05) \) using the Mann-Whitney U test, but none survived FDR adjustment \( (q>0.2) \). In the second comparison, which included 91 cases for whom we had binary plasma AMH data (n=20 with AMH \( \geq 0.1\)ng/mL, n=71 with AMH <0.1ng/mL), we found 16 features that nominally associated with the outcome \( (p<0.05) \) and one of these that also achieved FDR<0.2 after correction for multiple testing. A box and
whisker plot illustrating the intensity levels for this feature found in plasma samples from controls, cases with AMH ≥0.1ng/mL, and cases with AMH≤0.1ng/mL is presented in Figure 4. As illustrated, the abundance of this metabolite in cases with higher levels of AMH, indicative of more normal ovarian function, closely matched that seen in controls. However, the abundance in cases with low to undetectable AMH, indicative of premature ovarian insufficiency (Visser et al 2006; Sanders et al 2009; La Marca et al 2010; Spencer et al 2013; Frederick et al 2018), was significantly higher compared to those with AMH ≥0.1ng/mL (FDR=0.0734).

This feature had m/z=514.0603 and 116 s retention time on the AE column. With the mass tolerance of 10 ppm, this indicates that the exact mass for an H⁺ adduct would be between 514.0552 and 514.0654. From the retention time for elution, the molecule has the characteristics of an anion with relatively weak binding to the anion exchange column. If this were an H⁺ adduct (H⁺ mass equals 1.007825), the monoisotopic mass would be approximately 513.0525. Prediction of elemental compositions for this mass using ChemCalc (http://www.chemcalc.org/mf_finder/mfFinder_em_new) and the assumption that the molecule contained C, H, O, N, and S, showed seven possible compositions within the accuracy of measurement (10 ppm) ranging from 12 to 17 C, 19 to 25 H, 2 to 3 N, 8 to 18 O and 1 to 4 S. Unfortunately, we found no candidate matches for this feature in metabolomics databases, and initial attempts at identification by MS/MS were confounded by weak signal and strong interference, so its identity remains unknown.

Finally, we asked whether any of the features that discriminated between cases and controls might also discriminate among cases as a function of residual GALT activity predicted from GALT genotype (e.g. (Riehman et al 2001); Spencer et al 2013). Applying a Mann-Whitney U test, none of the features tested met the cut-off for significance after multiple test correction.

**Discussion**

The work presented here is important for two reasons. First, it documents that patients with classic galactosemia experience metabolic disturbance well beyond the Leloir pathway. While this has been assumed for decades, to our knowledge the results presented here offer the first real confirmation. Especially given the relatively large size of our participant cohort, this is a robust finding. Second, by defining specific biochemical pathways perturbed in treated patients with CG, we offer new candidate targets for intervention.

**Pathways perturbed in CG:**

We identified 14 biochemical pathways as notably perturbed in plasma samples from volunteers with CG at p<0.05 and at least four overlap metabolites (Table 2). This list is, by definition, incomplete, but provides a proof of principle and a first look at the diversity of pathways involved. While some pathways were identified only by results from one column (AE or C₁₈), four pathways were identified by results from both columns. Of note, galactose metabolism was, of course, also perturbed in cases (Figure 1) but failed to meet our significance cut-off (p=0.058), likely reflecting the fact that all of our cases were on galactose restricted diets and some key galactose metabolites that accumulate in...
galactosemic patients (e.g. gal1P) are strictly intracellular and therefore not detected well in plasma.

Perhaps the most impressive match that did make the list in Table 2 is glutathione metabolism, identified at p=0.0012 with four out of nine metabolites perturbed. This match is especially noteworthy because altered redox status, including perturbed GSH/GSSH ratio and striking over-expression of glutathione S-transferase genes, was demonstrated previously in a *Drosophila melanogaster* model of GALT-deficiency (Jumbo-Lucioni et al 2013) and also subsequently in a mouse model (Tang et al 2014), though in both model systems the altered redox status was described only following galactose exposure. That cysteine and vitamin E metabolism, also both related to redox, were also both perturbed in cases versus controls in the study reported here further reinforces the idea that patients with classic galactosemia may experience heightened oxidative stress and/or perturbed redox signaling despite dietary galactose restriction. Also evident in the list (Table 2) were multiple examples of amino acid metabolism and pathways involving mitochondria (e.g. carnitine shuttle, porphyrin metabolism, nicotinate (niacin) metabolism).

**Limitations:**

While informative, this study also had a number of limitations. For example, our cohort size of 183 cases, while large for a rare disease, was nonetheless limiting especially when comparing among cases by severity of outcomes with substantial missing data, or that were relevant only to a subset of cases, for example ovarian function (females only). Further, all cases in this study were on galactose restricted diets – and had been for years. The metabolomic perturbations we detected therefore are relevant to long-term outcomes in CG but are unlikely to shed light on acute outcomes. The absence of diet information on controls in this study is also a limitation; we do not know which, if any controls did versus did not consume milk. We therefore cannot formally exclude the possibility that at least some of the differences we detected between cases and controls might reflect differences stemming from diet rather than differences caused by GALT-deficiency. Our controls were also limiting, both in number and in age distribution; all controls were at least 18 years old while cases included both adults and children. While we did correct for age in our analyses, it is worth noting that in the linear regression models, only 19% of discriminatory AE features showed association with age at blood draw (raw p<0.05) and for discriminatory features from the C₁₈ column this number was only 16%.

As a further step to test whether the disparity in number and age range of cases and controls skewed our results we repeated the univariate and multivariate analyses comparing plasma features between the 31 controls in the study and a subset of 31 cases selected randomly from the full set of 183 with regard to gender, outcomes, and metabolic features, but intentionally with regard to age at blood draw, so that in this modified analysis cases and controls no longer differed by either age or cohort size. In short, 12 of the 14 pathways that were significantly perturbed in our original analysis were similarly dysregulated in this age-and cohort size-matched analysis (data not shown).

To address further how the disparity in sample size between cases and controls influenced our results, we generated 100 additional random subsets of 31 cases and again repeated our
feature selection approach. While these random case subgroups were not age-matched to controls, the discriminatory features identified in these additional analyses were consistent with those identified in our original analysis with all 183 cases. On average, 88% of the features showing differential abundance between controls and the size-matched random sample of cases matched those from our original analysis (SD = 7.7% for AE dataset, SD = 6.3% for \( C_{18} \) dataset).

It is also important to note that the high dimensionality of our metabolomic dataset posed a challenge to data preprocessing. The application of \( \log_2 \) transformation may not have been the optimal normalization method for each individual feature. While choosing a normalization method on a feature-by-feature basis guided by careful investigation of each individual feature’s unique intensity distribution might have achieved better normalization, \( \log_2 \) transformation did yield a normal overall intensity distribution and a linear combination of feature intensities may be considered approximately normal given our sample size and the Central Limit Theorem.

A further limitation is the variable we used as a binary proxy for cognitive outcome in this study: whether or not the volunteer received special educational services in elementary school. We chose this very imperfect measure because it was available for a large number of volunteers, thus increasing the power of the study. To test whether receipt of special educational services in elementary school correlated with a well-established indicator of cognitive outcome in our cohort, IQ, we selected those volunteers for whom we had both IQ and information on receipt of special educational services in elementary school (n=19). The average IQ of CG subjects in this group who did not receive special educational services (n=9) was 99.2 while the average IQ of CG subjects in this group who did receive special educational services (n=10) was 83.6. The difference between IQs for these 2 groups was statistically significant (p=0.035) confirming that receipt of special educational services in elementary school was a meaningful indicator of cognitive outcome in our cohort.

Another limitation stems from the reality that not all metabolites eluted as quantifiable peaks from the columns and elution programs used here, and of those metabolites that were counted as features, not all could be identified, or identified uniquely. This means our data set analyzed for discriminatory features was at best a subset – it was not comprehensive. This limitation was especially notable in our search for features perturbed in association with AMH level in girls and women with CG, as the feature most strongly associated with that outcome did not map to a known metabolite, and attempted identification by MS/MS was unsuccessful. Future work will be required to reveal the identity this feature. The absence of clear metabolite assignments for all \( m/z \) features also means that we likely missed pathways that were perturbed because key metabolites were perturbed in anonymity.

Finally, perhaps the most significant limitation of our study is that our metabolomic analyses were conducted on plasma, which, by definition, lacks metabolites restricted to the intracellular space. Metabolites found in plasma therefore cannot reflect the full spectrum of metabolites found in tissues such as brain and ovary. Beyond the scope of the current study we plan to address this point using relevant tissues collected from animal model systems of GALT deficiency.
Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are especially grateful to the many individuals and families who participated in this study, and to the Galactosemia Foundation (www.galactosemia.org) through which most volunteers were recruited. Without them, none of this work would have been possible. We also thank Andrei Todor for assistance with some feature annotation and Dr. Yating Wang for MS/MS experiments. This research was funded in part by HERCULES, a P30 Core Center grant from the National Institute of Environmental Health Sciences (P30 ES019776) and in part by NIH grant R01 DK107900 (to JLFK). DPJ was supported in part by NIH grants ES023485 and ES025632, with instrumentation support by S10OD018006.

Funding: This work was funded in part by NIH grant DK107900 (to JLFK) and in part by a pilot award (to JLFK) from HERCULES, a P30 Core Center grant from the National Institute of Environmental Health Sciences (P30 ES019776). DPJ was supported in part by NIH grants ES023485 and ES025632, with instrumentation support by S10OD018006. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. The authors confirm independence from the sponsors; the content of this article was not influenced by the sponsors.

Abbreviations:

- AE: anion exchange
- BCR: balanced classification rate
- AMH: anti-Müllerian hormone
- CG: classic galactosemia
- FDR: false discovery rate
- FWHM: full width at half maximum
- gal1P: galactose-1-phosphate
- GALT: galactose-1-phosphate uridylyltransferase
- HCA: hierarchical cluster analysis
- HMDB: Human Metabolome Database
- HRM: high-resolution metabolomics
- IQR: interquartile range
- LC-MS: liquid chromatography mass spectrometry
- m/z: mass-to-charge ratio
- PCA: principal component analysis
- PLSDA: partial least squares discriminant analysis
- RBC: red blood cells
- RT: retention time
VIP variable importance in projection

References


Figure 1: Major pathways of galactose metabolism in humans.

(A) Enzymes of the Leloir pathway are presented in italic font. Classic galactosemia results from profound deficiency of the middle enzyme, galactose 1-phosphate uridyltransferase.
Figure 2: Metabolome-wide associations of AE plasma features with CG.
Manhattan plots show $-\log_{10}(p\text{-value})$ from linear regression models as a function of mass-to-charge ratio (A) and retention time (B). Dashed line indicates selection cutoff (FDR=0.2). Features above the cutoff that further achieved VIP>2 are shown in blue.
Figure 3: Example box and whisker plots showing four individual discriminatory features that mapped to known metabolites: two (panels A and B) were elevated in CG cases, and two (panels C and D) were lower in cases.

(A) Glutamic acid, \( m/z = 148.0605 \), RT (AE)=144s, (B) Creatine, \( m/z = 132.0768 \), RT (AE)=98s, (C) Arginine, \( m/z = 175.1190 \), RT (C\(_{18}\))=38s, (D) Tetradecanoylcarnitine, \( m/z = 372.3110 \), RT (C\(_{18}\))=407s.
*m/z:* 514.0603  
Retention time: 116s (AE)  
Annotation: unknown

Figure 4: Box and whisker plots showing observed intensity of a single m/z feature that associated with AMH, a marker of ovarian outcome, among cases. Although this feature was identified uniquely by m/z=514.0603 and retention time (AE column)=116s, it did not map to a recognized metabolite by our annotation protocols. Nonetheless, this feature was detected at notably higher levels in plasma from cases with AMH<0.1ng/mL, indicating premature ovarian insufficiency, than in cases with higher AMH, or controls. This difference was significant at raw p<0.05 and FDR=0.0734.
### Table 1.
Demographic and clinical characteristics of the study population.

<table>
<thead>
<tr>
<th></th>
<th>All (n=214)</th>
<th>Control (n=31)</th>
<th>CG (n=183)</th>
<th>p-value&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td>0.049</td>
</tr>
<tr>
<td>Female [n (%)]</td>
<td>122 (57.0)</td>
<td>23 (74.2)</td>
<td>99 (54.1)</td>
<td></td>
</tr>
<tr>
<td>Male [n (%)]</td>
<td>92 (43.0)</td>
<td>8 (25.8)</td>
<td>84 (45.9)</td>
<td></td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td></td>
<td></td>
<td></td>
<td>0.192</td>
</tr>
<tr>
<td>White/not of Hispanic origin [n (%)]</td>
<td>172 (80.4)</td>
<td>29 (93.5)</td>
<td>143 (78.1)</td>
<td></td>
</tr>
<tr>
<td>White/of Hispanic origin [n (%)]</td>
<td>15 (7.0)</td>
<td>1 (3.2)</td>
<td>14 (7.7)</td>
<td></td>
</tr>
<tr>
<td>Other [n (%)]</td>
<td>9 (4.2)</td>
<td>1 (3.2)</td>
<td>8 (4.4)</td>
<td></td>
</tr>
<tr>
<td>Unknown [n (%)]</td>
<td>18 (8.4)</td>
<td>0 (0.0)</td>
<td>18 (9.8)</td>
<td></td>
</tr>
<tr>
<td>Age at blood draw [y, median (IQR)]</td>
<td>14.0 (16.0)</td>
<td>32.0 (19.5)</td>
<td>12.0 (12.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Sample storage time [y, median (IQR)]</td>
<td>8.3 (5.6)</td>
<td>10.0 (1.0)</td>
<td>7.3 (4.1)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

<sup>a</sup>Fisher’s exact test and Mann–Whitney U-test were performed for nominal and continuous variables, respectively.
Table 2.
Metabolic pathways perturbed in plasma of CG subjects with Mummichog \( p \)-value < 0.05 and at least four overlap metabolites.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Overlap size</th>
<th>Pathway size</th>
<th>( p )-value</th>
<th>Column</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glutathione metabolism</td>
<td>4</td>
<td>9</td>
<td>0.0012</td>
<td>AE</td>
</tr>
<tr>
<td>Aspartate and asparagine metabolism</td>
<td>11</td>
<td>56</td>
<td>0.0013</td>
<td>AE</td>
</tr>
<tr>
<td>Glycine, serine, alanine, and threonine metabolism</td>
<td>9</td>
<td>43</td>
<td>0.0014</td>
<td>AE</td>
</tr>
<tr>
<td>Alanine and aspartate metabolism</td>
<td>5</td>
<td>18</td>
<td>0.0017</td>
<td>AE</td>
</tr>
<tr>
<td>Porphyrin metabolism</td>
<td>6</td>
<td>20</td>
<td>0.0017</td>
<td>C(_{18})</td>
</tr>
<tr>
<td>Carnitine shuttle</td>
<td>7</td>
<td>29</td>
<td>0.0020</td>
<td>C(_{18})</td>
</tr>
<tr>
<td>Methionine and cysteine metabolism</td>
<td>8</td>
<td>41</td>
<td>0.0020</td>
<td>AE</td>
</tr>
<tr>
<td>Vitamin E metabolism</td>
<td>7</td>
<td>32</td>
<td>0.0023</td>
<td>C(_{18})</td>
</tr>
<tr>
<td>Arginine and proline metabolism</td>
<td>6</td>
<td>27</td>
<td>0.0027</td>
<td>C(_{18})</td>
</tr>
<tr>
<td>Urea cycle/amino group metabolism</td>
<td>6</td>
<td>34</td>
<td>0.0052</td>
<td>C(_{18})</td>
</tr>
<tr>
<td>Vitamin B3 (nicotinate and nicotinamide) metabolism</td>
<td>4</td>
<td>20</td>
<td>0.0071</td>
<td>AE</td>
</tr>
<tr>
<td>Pyrimidine metabolism</td>
<td>6</td>
<td>45</td>
<td>0.0219</td>
<td>AE</td>
</tr>
<tr>
<td>Purine metabolism</td>
<td>6</td>
<td>45</td>
<td>0.0219</td>
<td>AE</td>
</tr>
<tr>
<td>Glycerophospholipid metabolism</td>
<td>4</td>
<td>30</td>
<td>0.0257</td>
<td>C(_{18})</td>
</tr>
</tbody>
</table>