Urinary metabolic phenotyping of mucopolysaccharidosis type I combining untargeted and targeted strategies with data modeling

Abdellah Tebani, Isabelle Schmitz-Afonso, Lénaïg Abily-Donval, Bénédicte Héron, Monique Piraud, Jérome Ausseil, Anaïs Brassier, Pascale de Lonlay, Farid Zerimech, Frederic Vaz, et al.

Clinica Chimica Acta 475 (2017) 7–14

Contents lists available at ScienceDirect

Clinica Chimica Acta

journal homepage: www.elsevier.com/locate/cca

Urinary metabolic phenotyping of mucopolysaccharidosis type I combining untargeted and targeted strategies with data modeling

Abdellah Tebani[a,b,c], Isabelle Schmitz-Afonso[c], Lenaig Abily-Donval[b,d], Bénédicte Héron[e], Monique Piraud[f], Jérôme Ausseil[g], Anais Brassier[h], Pascale De Lonlay[h], Farid Zerimech[i], Frédéric M. Vaz[j], Bruno J. Gonzalez[b], Stephane Marret[b,d], Carlos Afonso[c], Soumeya Bekri[a,b,*]

[a] Department of Metabolic Biochemistry, Rouen University Hospital, Rouen 76000, France
[b] Normandie Univ, UNIROUEN, CHU Rouen, INSERM U1245, 76000 Rouen, France
[c] Normandie Univ, UNIROUEN, INSA Rouen, CNRS, COBRA, 76000 Rouen, France
[d] Department of Neonatal Pediatrics and Intensive Care, Rouen University Hospital, Rouen 76031, France
[e] Departement of Pediatric Neurology, Reference Center of Lysosomal Diseases, Trousseau Hospital, APHP, GRC ConCer-LD, Sorbonne Universities, UPMC University 06, Paris, France
[f] Service de Biochimie et Biologie Moléculaire Grand Est, Unité des Maladies Héréditaires du Métabolisme et Dépistage Néonatal, Centre de Biologie et de Pathologie Est CHU de Lyon, Lyon, France
[g] INSERM U1088, Laboratoire de Biochimie Métabolique, Centre de Biologie Humaine, CHU Sud, 80054 Amiens Cedex, France
[h] Reference Center of Inherited Metabolic Diseases, Imagine Institute, Hospital Necker Enfants Malades, APHP, University Paris Descartes, Paris, France
[i] Laboratoire de Biochimie et Biologie Moléculaire, Université de Lille et Pôle de Biologie Pathologie Génétique du CHRU de Lille, 59000 Lille, France
[j] Laboratory of Genetic Metabolic Diseases, Department of Clinical Chemistry and Pediatrics, Academic Medical Center, Amsterdam, The Netherlands

ARTICLE INFO

Keywords:
Metabolomics
Inborn errors of metabolism
Mucopolysaccharidosis type I
Lysosomal storage diseases
Mass spectrometry
Ion mobility

ABSTRACT

Background: Application of metabolic phenotyping could expand the pathophysiological knowledge of mucopolysaccharidoses (MPS) and may reveal the comprehensive metabolic impairments in MPS. However, few studies applied this approach to MPS.

Methods: We applied targeted and untargeted metabolic profiling in urine samples obtained from a French cohort comprising 19 MPS I and 15 MPS I treated patients along with 66 controls. For that purpose, we used ultra-high-performance liquid chromatography combined with ion mobility and high-resolution mass spectrometry following a protocol designed for large-scale metabolomics studies regarding robustness and reproducibility. Furthermore, 24 amino acids have been quantified using liquid chromatography coupled to tandem mass spectrometry (LC-MS/MS). Keratan sulfate, Heparan sulfate and Dermatan sulfate concentrations have also been measured using an LC-MS/MS method. Univariate and multivariate data analyses have been used to select discriminant metabolites. The mummichog algorithm has been used for pathway analysis.

Results: The studied groups yielded distinct biochemical phenotypes using multivariate data analysis. Univariate statistics also revealed metabolites that differentiated the groups. Specifically, metabolites related to the amino acid metabolism. Pathway analysis revealed that several major amino acid pathways were dysregulated in MPS. Comparison of targeted and untargeted metabolomics data with in silico results yielded arginine, proline and glutathione metabolisms being the most affected.

Conclusion: This study is one of the first metabolic phenotyping studies of MPS I. The findings might help to generate new hypotheses about MPS pathophysiology and to develop further targeted studies of a smaller number of potentially key metabolites.

Abbreviations: IEM, inborn errors of metabolism; LSD, lysosomal storage diseases; MPS, mucopolysaccharidoses; GAGs, glycosaminoglycans; MPS I, mucopolysaccharidosis type I; MPS IT, treated mucopolysaccharidosis type I; UPLC-IM-MS, ultraperformance liquid chromatography-ion mobility mass spectrometry; CCS, collision cross section; ERT, enzyme replacement therapy; QC, quality control; HS, heparan sulfate; DS, dermatan sulfate; KS, keratan sulfate; ROC, receiver operating characteristic; FDR, false discovery rate; PCA, principal component analysis; OPLS-DA, orthogonal partial least-squares-discriminant analysis; VIP, variable influence in projection; AUC, area under curve; mTORC1, mammalian target of rapamycin complex 1

* Corresponding author at: Department of Metabolic Biochemistry, Rouen University Hospital, 76031 Rouen Cedex, France.
E-mail address: soumeya.bekri@chu-rouen.fr (S. Bekri).

http://dx.doi.org/10.1016/j.cca.2017.09.024
Received 12 September 2017; Received in revised form 29 September 2017; Accepted 30 September 2017
Available online 02 October 2017

1. Introduction

Inborn errors of metabolism (IEM) represent a group of about 500 rare diseases with an overall estimated incidence of 1/2500. The diversity of involved metabolisms explains the difficulties in establishing their diagnosis. Optimal management of these patients requires then improved speed of biochemical investigations to allow early diagnosis and better monitoring. The rise of "omic" approaches offered a growing hope to provide new effective tools for screening, diagnosis and monitoring of these diseases. Unlike the conventional medical biology practice based on the sequential study of genes, proteins and metabolites, the great challenge of modern biology is to understand disease as a complex, integrated and dynamic network [1]. The concept of "metabolome" refers to the comprehensive complement of all metabolites present in a given biological system, fluid, cell or tissue [2]. So, metabolomics is one of the "omic" technologies based on biochemical characterizations of the metabolome and its changes related to genetic and environmental factors. Metabolomics allows evaluating the biochemical mechanisms involved cell or tissue changes in a systematic fashion [3,4]. Given the strong link between IEM and metabolism, metabolomics is very appealing to explore these diseases [5]. For years, mass spectrometry has been used to assess IEM [6–9]. However, few metabolomic research has been published in lysosomal storage diseases (LSD) field. LSDs represent a group of about 50 inherited disorders due to lysosomal proteins deficiencies which lead to a progressive accumulation of compounds within the lysosome. This metabolite storage causes various organ failures and premature death [10]. Mucopolysaccharidoses (MPS) belong to the LSD group. They are caused by impaired catabolism of glycosaminoglycans (GAGs), leading to their accumulation in lysosomes and extracellular matrix [11]. Accumulated GAGs cause progressively multiple tissues and organ damages [12]. There are 11 known enzyme deficiencies, resulting in seven distinct forms of MPS [10]. The overall incidence is > 1 in 30,000 live births [13]. Most MPS patients are asymptomatic after birth, however, prenatal symptoms may be observed in MPS I, MPS IVA and more frequently in MPS VII. MPS symptoms and severity vary with patients and MPS subtypes. Several MPS treatments are in clinical use or being investigated under clinical trials for patients [14]. MPS I is a rare autosomal recessive disorder caused by α-L-Iduronidase (IDUA, EC 3.2.1.76) deficiency. IDUA degrades complex polysaccharides by removing a single α-L-iduronyl residue from heparan sulfate and dermatan sulfate. The symptoms range from the severe Hurler form [MPS IH - OMIM #67014] to the more attenuated Hurler–Scheie (MPS IH/S - OMIM #607015) and Scheie (MPS IS - OMIM #67016) phenotypes. The classification is mainly based on the age at first symptoms and the presence or not of mental retardation [15]. The average survival age is of 28 years which imply patient's shift from pediatrics to adults [16]. Two specific treatments are available: hematopoietic stem cell transplantations (from bone marrow or blood cord donors) since the 1980s, and enzyme replacement therapy (ERT) (Laronidase, ALDURAZYME) since the 2000s. The aim of this study is to apply both targeted and untargeted metabolic profiling on MPS I patients compared to controls and to treated MPS I patients (MPS IT) to assess metabolic changes in this condition.

2. Materials and methods

2.1. Urine samples

Random urine samples were collected from MPS patients in whom the diagnosis had been confirmed by demonstrating marked enzyme deficiency in leucocytes and/or by molecular analysis. Pseudodeficiencies have been ruled out. Urine samples were collected within seven reference centers for inherited metabolic diseases in France. Nineteen untreated MPS I patients were evaluated: 18 males (age range from 1 to 43.6 years, mean age: 22 years) and 1 female (age 5.5 years). Control urine samples from 66 healthy subjects, 27 males and 39 females (age range from 5.5 to 70 years, mean age: 40.8 years). Fifteen samples from MPS IT with enzyme replacement therapy, 11 males and 4 females (ages range from 1.3 to 39.3 years, mean age: 11.5 years) were analyzed. This project was approved by the Research Ethics Board of Rouen University Hospital (CERNI E2016-21).

2.2. Metabolic phenotyping

2.2.1. Sample preparation

For untargeted metabolomics, urine samples were processed by transferring 200 μL of urines to 1.5 mL tubes and centrifuged at 4 °C for 10 min at 13,000*g* then 100 μL ultrapure water were added to 100 μL of supernatant and mixed. For amino acids and GAG analysis, detailed protocols are presented in Supporting information.

2.2.2. Untargeted analysis

Ultra-high-performance liquid chromatography-ion mobility mass spectrometry and data-independent MS acquisition with simultaneous analysis of molecular fragmentation (MS^E) were performed on Synapt G2 HDMS (Waters, Saint-Quentin-en-Yvelines, France) mass spectrometer as previously described [17]. Detailed protocol is presented in Supplementary material.

2.2.3. Raw data preprocessing

All LC-IM/MS raw data files, data processing, peak detection and peak matching across samples using retention time (t_R) correction and chromatographic alignment along with drift time and cross collision section (CCS) calculation were performed using Progenesis QI (Waters MS Technologies, Manchester, UK) to yield a data matrix containing retention times, accurate masses (*m/z*), CCS and peak intensities. The preprocessing step resulted in an X-matrix where t_R, CCS and *m/z* values were concatenated into "t_R_m/z_CCS" features (in columns) present in each sample (in rows) with corresponding peak areas.

2.2.4. Quality control

Ten microliters of each urine sample are mixed together to generate a pooled quality control sample (QCs). QCs and mobile phase blank samples were injected sequentially in-between the urine samples. In addition, a dilution series of QC samples (6%, 12.5%, 25%, 50% and 100% of original concentration) are used to assess the quality of the extracted features. More details are presented as Supporting information.

2.2.5. Targeted analysis

2.2.5.1. Amino acids quantification. The analysis of free amino acid profiles in urine was based on a liquid chromatography coupled to tandem mass spectrometry method and the aTRAQ reagent. The aTRAQ kit (Sciex, France) allows to quantify 24 proteinogenic and non-proteinogenic free amino acids, in a range of biological fluids. The detailed description of the applied LC-MS/MS methodology is presented in Supplementary material. The amino acids concentrations were normalized using creatininuria.

2.2.5.2. Glycosaminoglycan quantification (HS, DS and KS). Total urinary GAGs were measured with the dimethyl methylene blue-binding assay [18]. Urinary GAG-derived disaccharides (heparan sulfate, dermatan sulfate and keratan sulfate) were analyzed using LC-MS/MS as described by Langereis et al. [19]. The detailed description of the protocol is presented in Supplementary material. The GAG concentrations were normalized on creatininuria.

2.3. Statistical analyses

A one-way analysis of variance (ANOVA) test was applied to each selected variable in order to confirm their actual difference between the

three groups. A *t*-test is used when binary comparison is applied. Furthermore, the Benjamini and Hochberg false discovery rate (FDR) method was used for calculating the false-positive rate associated with multiple comparisons, and provides corrected q-values with a 0.05 significance level (FDR 5%). A Receiver operating characteristic curve (ROC) has been used to assess the diagnostic performance of the chosen classifiers.

2.4. Data analysis and modeling

Support vector regression normalization method was applied using the MetNormalizer R package [20] before any data analysis, to remove the unwanted intra- and inter-batch measurement analytical variations. The normalized data matrix has been log-transformed and pareto-scaled. All data analyzes and modeling were done using SIMCA 14.0 (MKS DAS, Umeå, Sweden) and R software. First, hierarchical cluster analysis and PCA were used as exploratory unsupervised methods [21]. Orthogonal Partial Least-Squares-Discriminant Analysis (OPLS-DA) was used as a supervised method for predictive modeling purposes. Details regarding data modeling and validation results from all OPLS-DA models are provided in Supplementary material.

2.5. Feature selection and annotation

To select the most discriminant variables for the separation of groups, S-Plot was used. The S-plot combines the covariances and correlations between the **X** matrix and OPLS scores for a given model component. The covariance values give the magnitude of contribution of a variable while the correlation values reflect the effect and reliability of the variable for the model component scores. Variables with both very high correlation and covariance are important for the explanation power of the model. Selection of discriminant variables was achieved using the VIP (Variable Influence in Projection) score procedure for each validated OPLS-DA model [22]. Putative annotation of detected features was performed using accurate mass comparison using freely available metabolite databases HMDB, LipidBlast, KEGG, and Metlin. Furthermore, CCS values were also compared to the MetCCS database [23].

2.6. Pathway and network analysis

In order to provide a broader understanding of metabolic changes in MPS I, we explored the biochemical pathways [24]. A network analysis approach using the Mummichog software has been performed. This Python package highlights pathways that are significantly impacted in the studied groups. Significantly impacted biochemical pathways are those exhibiting an adjusted p-value < 0.05. For this comparison, we focused on features that significantly changed (q-values = 0.05 and FDR = 5%). Mummichog annotates metabolites based on accurate mass m/z and tests significant pathway enrichment within a reference metabolic network using a Fisher's exact test [25]. To protect against incorrect pathway selection, redundant pathways or those enriched by fewer than two metabolites were excluded. MetaboAnalyst [26] has been used for Metabolite Set Enrichment Analysis on the amino acid concentration matrix. The Fig. 1 presents an overview of the adopted metabolomics workflow.

3. Results

3.1. Untargeted analysis

The untargeted analysis of urine samples of control individuals, MPS I and MPS IT patients yielded 854 features. The analysis by independent ANOVA test resulted in 511 metabolites above the $p < 0.05$ cut-off (FDR 5%). A hierarchical clustering analysis was first applied to group samples with similar profiles of variable intensity. The heatmap in

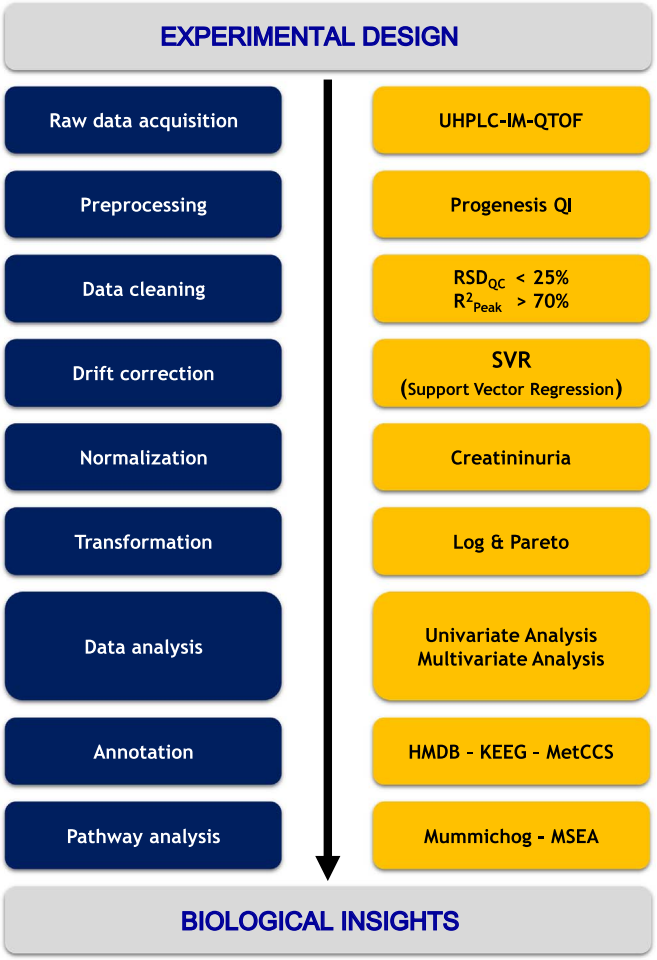

Fig. 1. Illustration of the untargeted metabolomics workflow spanning from experimental design to pathway analysis and biological interpretation. HMDB: Human Metabolome Database. KEGG: Kyoto Encyclopedia of Genes and Genomes. MetCCS: Metabolite CCS database. MSEA: Metabolite Set Enrichment Analysis. RSD: Relative Standard Deviation.

Fig. 2A represents the top 100 features ranked by ANOVA. The results show that all samples belonging to the same group were correctly clustered together. The dendrogram structure highlights two main clusters of variable intensities represented by its two longest branches (maximum dissimilarity according to the Euclidean distance). According to the color gradient, the intensity differences between groups are substantial. This first analysis allowed us to easily detect natural clusters in the data, although it did not facilitate extraction of discriminant variables among the dataset. To further explore natural separation between metabolic profiles and reduce data dimensionality, the dataset underwent a principal component analysis (PCA). The number of significant components was estimated using internal cross-validation with seven exclusion groups giving a three-component PCA model accounting for 21% of the total variance. The resulting scores plot was used to identify trends, groups and potential outliers within the data. Fig. 2B shows PCA scores plot. There is a clear separation between MPS I and MPS IT samples. However, there is an overlap with the control samples. Thus, to address our classification purposes, supervised methods are more suitable since they allow to accurately model the relationship between controls, MPS I and MPS IT samples. OPLS-DA classification was first applied to the dataset. Samples were labeled according to the corresponding groups, MPS I, MPS IT and control. A model was considered predictive if the Q2 (cross-validation measure of the predictive power) regression line intercept resulting from the permutation test was negative. This means that the random

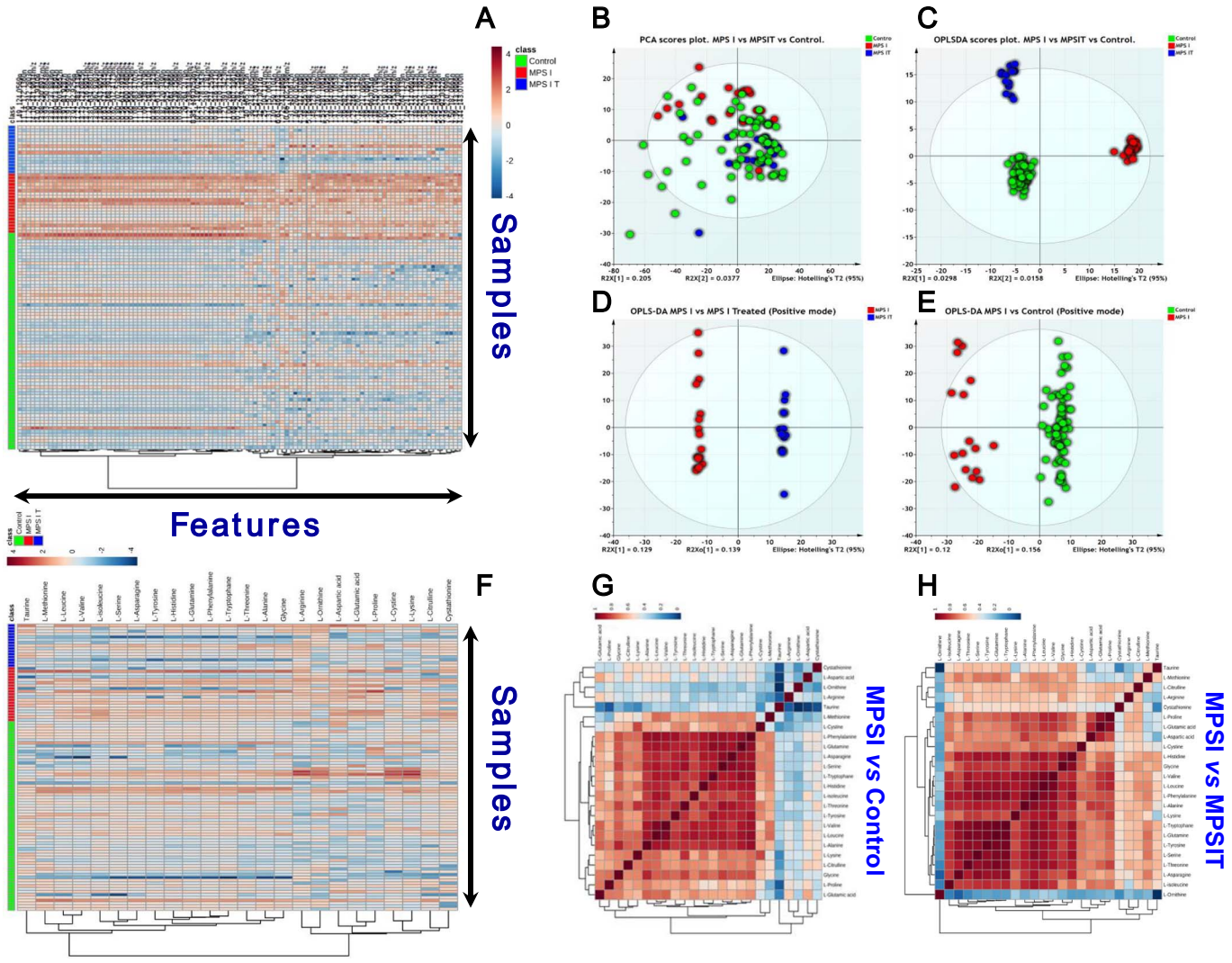

Fig. 2. 2A) Hierarchical cluster analysis and heat map visualization of top 100 variables (y-axis) ranked by ANOVA. The urine sample classes are represented along the x-axis. The color code was used to represent log-scaled intensities of features between − 4 (blue) and + 4 (brown), showing the features relative abundance according to the groups. 2B) PCA scores plot of the normalized dataset. The three groups are represented by different colors. MPS I and MPS IT samples are well separated on PC1 according to their class membership. However, control samples show an overlap. 2C) OPLSDA scores plot (R2 = 0.96, Q2 = 0.54) shows a clear separation between the different groups. PC 1 separates the MPS I samples from the controls. However, PC2 separates treated MPS I from control samples. 2D) Clear separation between treated MPS I and MPS IT samples is observed (R2 = 0.97, Q2 = 0.63). 2E) Clear separation between MPS I samples from the controls is observed (R2 = 0.94, Q2 = 0.63). Detailed model characteristics and validation are given in Supporting information. 2F) Heat map representing the clustering of 24 amino acids across the 3 groups of samples (MPS I, MPS IT and Controls). Columns represent individual samples and rows refer to amino acid. Shades of red or blue represent elevation or decrease, respectively, of an amino acid. 2G and 2H) Spearman rank-order correlation matrix 24 amino acids based on their concentrations profiles across all samples. Shades of red or blue represent low-to-high correlation coefficient between markers. G) MPS I vs Control. H) MPS I vs MPS IT. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Some discriminant features extracted by OPLS-DA models allowing the discrimination of control subjects, MPS I and MPS IT.

HMDB	Putative annotation	Formula	*M*	*m/z*	Adduct	Δ *m/z* (ppm)	t_R (min)	tD (ms)	CCS (A^2)	FDR	%RSD	VIP
MPS I vs. Control												
HMDB00062	Carnitine	$C_7H_{15}NO_3$	161.1053	203.1518	M + ACN + H	0.48	1.41	2.43	140.4	3.22E − 09	10.0	1.90
HMDB00207	Oleic acid	$C_{18}H_{34}O_2$	282.2546	283.2616	M + H	− 4.36	10.75	4.00	183.6	9.20E − 07	24.53	1.85
HMDB29022	Prolyl-Lysine	$C_{11}H_{21}N_3O_3$	243.1595	282.1226	M + K	4.88	1.49	3.08	158.1	7.43E − 08	5.97	1.73
HMDB00268	Tetrahydrocorticosteron	$C_{21}H_{34}O_4$	350.2465	351.2538	M + H	2.33	10.75	4.48	194.2	6.40E − 05	17.26	1.51
HMDB00517	Arginine	$C_6H_{14}N_4O_2$	174.1112	175.1212	M + H	− 2.70	1.23	2.11	130.7	3.21E − 02	22.62	2.38
MPS I vs. MPS IT												
HMDB00517	Arginine	$C_6H_{14}N_4O_2$	174.1112	175.1212	M + H	− 2.70	1.23	2.11	130.7	6.12E − 03	22.62	2.38
HMDB00207	Oleic acid	$C_{18}H_{34}O_2$	282.2546	283.2616	M + H	− 4.36	10.75	4.00	183.6	6.96E − 04	24.53	1.93
HMDB00062	Carnitine	$C_7H_{15}NO_3$	161.1053	203.1518	M + ACN + H	0.48	1.41	2.43	140.4	4.62E − 04	9.97	1.93
HMDB28988	Phenylalanylalanine	$C_{12}H_{16}N_2O_3$	236.1152	237.1225	M + H	− 3.62	7.67	2.70	147.8	2.07E − 03	10.51	1.91
HMDB29022	Prolyl-Lysine	$C_{11}H_{21}N_3O_3$	243.1595	282.1226	M + K	4.88	1.49	3.08	158.1	1.70E − 03	5.97	1.77
HMDB00268	Tetrahydrocorticosteron	$C_{21}H_{34}O_4$	350.2465	351.2538	M + H	2.33	10.75	4.48	194.2	4.00E − 03	17.26	1.61

M: monoisotopic mass, ppm: parts per million, t_R: retention time, tD: drift time, CCS: cross collision section, VIP: variable importance in projection.

labeled models exhibit lower predictive performance than the true one. The final model had an R2 = 0.96 and Q2 = 0.54. The OPLS-DA scores plot (Fig. 2C) revealed that each class was well separated, suggesting that the OPLS-DA model successfully discriminated samples according to their underlying metabolic profile. This model was internally validated both by CV-ANOVA (p-value = 4×10^{-20}) and by the permutation test (999 permutations gave a negative Q2 intercept). Model validation details are shown in Supplementary information (Fig. S5). To go further in data modeling, binary OPLS-DA classification models have been built. The first OPLS-DA model was built using a dataset including Control and MPS I samples. The model had one predictive and two orthogonal components, and its validation parameters were as follows: R2 = 0.94, Q2 = 0.63 and CV-ANOVA p-value = 1.75×10^{-15} (Fig. S6). The corresponding score plot is shown in Fig. 2E. It exhibited a clear separation between the two classes on the predictive component. A second OPLS-DA model was built using a dataset including MPS I and MPS IT samples following the same procedure. The OPLS-DA model had one predictive and three orthogonal components with R2 = 0.97, Q2 = 0.63 and CV-ANOVA p-value = 5×10^{-4} (Fig. 2D). Selection of discriminant variables was achieved using the VIP scores procedure for each validated OPLS-DA model. Based on 1 as a cutoff value, 216 features out of the 854 were selected for the MPS I vs. Control model and 169 for the MPS I vs. MPS IT model. We then refined the two lists of variables by retaining only the most discriminant variables along with their putative annotation. The list included Carnitine, Arginine, Tetrahydrocorticosteron, Prolyl-Lysine, oleic acid and Phenylalanylalanine. These discriminant variables are presented in Table 1 for both models along with their respective statistical metrics and annotation accuracy. Boxplots of the main discriminant features are presented in Fig. S8. The discriminant performance of these features is also assessed using area under the ROC curves. Carnitine has the highest AUC (0.93). The overall ROC results are shown in Fig. S9. Furthermore, to explore the underlying pathways dysregulated in MPS I we used the Mummichog software to look for significant pathways related to variation in the significantly disturbed features. Different metabolism pathways were affected such as glycerophospholipid metabolism, vitamins and amino acids are shown in Table 2. Interestingly, a series of amino acid metabolic pathways were markedly dysregulated.

Table 2
Significantly dysregulated pathways.

Pathway	Overlap size	p-value (FDR = 5%)
MPS I vs. Control		
Lysine metabolism	13	0.0014
Glycerophospholipid metabolism	11	0.0028
Methionine and cysteine metabolism	8	0.0123
Tyrosine metabolism	26	0.0216
Biopterin metabolism	7	0.0222
Urea cycle/amino group metabolism	14	0.0260
Ascorbate (Vitamin C) and Aldarate metabolism	5	0.0264
Arginine and Proline metabolism	8	0.0341
Glutathione metabolism	3	0.0388
Vitamin H (biotin) metabolism	3	0.0388
MPS I vs. MPS IT		
Aspartate and asparagine metabolism	13	0.0075
Lysine metabolism	9	0.0128
Glutathione metabolism	3	0.0145
Vitamin H (biotin) metabolism	3	0.0145
De novo fatty acid biosynthesis	6	0.0280
Tyrosine metabolism	19	0.0357
Ascorbate (Vitamin C) and Aldarate metabolism	4	0.0359
Omega-3 fatty acid metabolism	3	0.0449
Vitamin B5 - CoA biosynthesis from pantothenate	3	0.0449

FDR: false discovery rate.

3.2. Targeted analysis

The first targeted analysis addressed urinary glycosaminoglycans concentrations. As expected, total GAGs, dermatan sulfate and heparan sulfate are significantly elevated in MPS I patients (Table 3 and Fig. S11). Of note, Keratan sulfate is slightly elevated in MPS I patients compared to control samples. Given the results of the above untargeted approach, we also performed a targeted amino acid profiling on all the samples. Twenty-four amino acids were quantified and their concentrations were subjected to subsequent statistical and pathway analysis. Table S3 presents absolute urine concentrations of amino acids. Boxplots of normalized amino acid concentrations are presented in Fig. S10. The statistical analysis of amino acids is listed in Table 3. Regarding Control vs. MPS I comparison, thirteen amino acids have shown significant differences between the two groups; Arginine, Aspartic acid, Glutamic acid, Proline, Valine, Tryptophan, Lysine, Alanine, Leucine, Histidine, Threonine, Glutamine and Glycine. Besides, six amino acids showed statistically different concentrations between MPS I and MPS IT samples: Glutamic acid, Aspartic acid, Valine, Alanine and Isoleucine. To determine the amino acids profile differences between controls and MPS I and MPS IT patients, the 24 amino acids were first analyzed by an ANOVA test. The analysis yielded eight amino acids above the $p < 0.05$ cut-off (FDR 5%). A hierarchical clustering analysis was first applied to group samples with similar profiles. The heatmap in Fig. 2F represents the 24 amino acids ranked by ANOVA. Even though, there is no an obvious visual pattern, the results show that all samples belonging to the same group were correctly clustered together. The dendrogram structure highlights two main clusters of variables represented by its two longest branches (maximum dissimilarity according to the Euclidean distance). Furthermore, a correlation analysis of the overall concentrations matrix has been performed. Fig. 2G and 2H present the heatmap of the correlation analysis. Both figures show a clear cluster of variables that have high correlation. Fig. 2G (MPS I vs control) showed a main cluster including Alanine, Leucine, Valine, Tyrosine, Threonine, Isoleucine, Histidine, Tryptophane, Serine, Asparagine, Glutamine and Phenylalanine. Regarding MPS I vs MPS IT, Fig. 2H shows two main clusters: The first one includes Isoleucine, Asparagine, Threonine, Serine, Tyrosine, Glutamine and Tryptophane; the second includes Alanine, Phenylalanine, Leucine, Valine, Glycine and Histidine. To assess the diagnostic performance of the different amino acids, univariate ROC curve analyses for MPS I vs. Control groups indicated four amino acids with high AUC above 0.80 and are: Arginine (0.90), Glutamic acid (0.86), Aspartic acid (0.83) and Proline (0.81). The same procedure has been performed for MPS I vs. MPS IT groups and indicated two amino acids with high AUC above 0.80 and were: Aspartic acid (0.85) and Isoleucine (0.82). The overall univariate and ROC analysis results are presented in Table 3. A comparison of different combinations of the main significant amino acids using a PLSDA model with three components each is presented in Fig. S12. Using these quantitative data, we performed pathway analysis that yielded the main impaired metabolisms. For MPS I vs. Control analysis, Arginine and Proline, Malate-Aspartate Schuttle, Cysteine, Urea cycle and alanine metabolism were the most affected pathways. Regarding, MPS I vs. MPS IT analysis, Alanine, Malate-Aspartate Schuttle, branched amino acids metabolisms were the most affected. The overall results are shown in Fig. 3A and 3B for all the studied groups.

4. Discussion

In this study, the potential of metabolomics to identify biomarkers related to MPS I in urine was investigated. The data demonstrates that lysosomal accumulation of GAGs triggers deep metabolic turnover in MPS I patients. Urinary global metabolomics profiling may provide better understanding MPS I disease mechanisms and may pave the way for potential biomarkers. The unveiled metabolic alterations were mainly relevant to amino acid pathways contributing significantly to

Table 3
t-Test statistics, fold change and area under the curve (AUC) of the receiver operating curves (ROC) for 24 amino acids and the Glycoaminoglycans (GAGs) ($p < 0.05$). Significant features are highlighted in bold (false discovery rate FDR = 5%).

	MPS I vs Control			MPS I vs MPS IT		
	AUC	q-Value (FDR)	Fold change	AUC	q-Value (FDR)	Fold change
Amino acids						
L-Arginine	0.904	**3.14E − 06**	− 3.75	0.51	4.78E − 01	− 0.41
L-Aspartic acid	0.83	**2.25E − 04**	− 2.19	0.86	**1.52E − 02**	1.3
L-Glutamic acid	0.858	**4.78E − 04**	− 1.85	0.75	**3.58E − 02**	0.28
L-Proline	0.816	**4.79E − 04**	− 1.89	0.73	7.34E − 02	0.43
L-Valine	0.786	**1.50E − 02**	− 1.87	0.78	**3.33E − 02**	0.5
L-Tryptophane	0.752	**2.41E − 02**	− 2.72	0.66	2.08E − 01	0.16
L-Lysine	0.751	**2.41E − 02**	− 2.56	0.75	7.34E − 02	0.47
L-Alanine	0.784	**2.41E − 02**	− 2.13	0.76	**3.58E − 02**	0.83
L-Leucine	0.734	**2.51E − 02**	− 1.45	0.76	**4.72E − 02**	0.35
L-Histidine	0.733	**4.30E − 02**	− 2.53	0.61	2.82E − 01	− 0.05
L-Threonine	0.687	**4.77E − 02**	− 1.13	0.67	2.01E − 01	0.12
L-Glutamine	0.746	**4.77E − 02**	− 2.08	0.68	2.01E − 01	0.15
Glycine	0.754	**4.77E − 02**	− 2.05	0.58	2.01E − 01	0.04
Cystathionine	0.709	7.17E − 02	− 1.85	0.65	3.65E − 01	0.15
L-Serine	0.718	7.19E − 02	− 2.06	0.68	2.82E − 01	0.4
L-Isoleucine	0.685	7.19E − 02	− 1.1	0.83	**1.85E − 02**	1.81
L-Phenylalanine	0.704	7.80E − 02	− 1.3	0.75	1.51E − 01	0.7
L-Ornithine	0.652	1.23E − 01	− 1.54	0.57	7.94E − 01	− 0.19
L-Citrulline	0.657	1.72E − 01	− 1	0.66	3.96E − 01	0.19
L-Tyrosine	0.636	2.15E − 01	− 0.96	0.69	2.08E − 01	0.55
L-Cystine	0.53	2.50E − 01	− 0.98	0.58	5.65E − 01	− 0.31
L-Asparagine	0.638	2.61E − 01	− 0.73	0.7	2.31E − 01	0.74
Taurine	0.549	3.06E − 01	− 1.22	0.59	4.00E − 01	− 1.95
L-Methionine	0.524	3.65E − 01	− 0.48	0.65	1.93E − 01	0.09
GAGs						
Total GAGs	0.92	**4.76E − 04**	− 3.74	0.79	6.39E − 01	1.56
Dermatan sulfate	0.92	**1.37E − 06**	− 5.14	0.78	3.65E − 01	2.48
Heparan sulfate	0.89	**1.14E − 03**	− 3.13	0.75	6.02E − 01	1.46
Keratan sulfate	0.91	**2.75E − 02**	− 1.84	0.77	5.65E − 01	1.24

the clear discrimination of the different studied groups, MPS I, MPS IT and control samples using their metabolic differences. Indeed, based on the untargeted urinary metabolic profiles retrieved from the different studied groups, we were able to build a predictive model that clearly separates the different studied groups, MPS I, MPS IT and control samples using their metabolic differences. This study showed metabolic impairments mainly in amino acid metabolism and related metabolisms such vitamin and glutathione metabolisms. In the light of these results, we performed a targeted analysis focusing on amino acids profiles which confirmed the amino acids profiles alterations. This pathway analysis yielded different dysregulated metabolic pathways. Furthermore, we performed a comparative analysis between the pathway analysis results from both untargeted and targeted data along with the recently in silico systems analysis data reported by Salazar et al. [27]. These authors performed a system biology approach using a genome-scale human metabolic reconstruction to understand the effect of metabolism alterations in MPS. The in silico MPS I model was generated by silencing *IDUA* gene then this model was analyzed through a flux balance and variability analysis. Thus, to depict the interrelationships between our untargeted and targeted results along with these in silico metabolic impairment data, we used a Venn diagram approach (Fig. 3C). Thus, two main metabolisms were identified: Arginine-Proline metabolism and Cysteine-Glutathione metabolism. Detailed data are presented in Table S4. Arginine-Proline metabolism is depicted in Fig. S13 and cysteine-glutathione metabolism is presented in Fig. S14. The later metabolism is tightly linked to oxidative stress. Recent studies have shown oxidative damage involvement in the pathophysiology of several genetic diseases, including LSD [28]. The GAG biosynthesis requires recycled substrates, however, the lack of recycled substrates in MPS may lead to an increase in cellular energy needs [29]. This energy requirement may trigger the active mitochondrion turnover and results in an excess production of reactive oxidative species. Moreover, mitochondrion turnover is also triggered by the alteration of mitophagy process [30]. Oxidative stress has been recognized as a mechanism of cell damage in MPSs and has been reported in MPS patients undergoing ERT [31–35]. Fillipon et al. demonstrated that a reduction of oxidative defenses induces lipid and protein oxidative damages in MPS II patients; ERT plays a protective role et restores the antioxidant response [36]. Besides, Donida et al. showed that pro-inflammatory and pro-oxidant features occur in MPS IVA patients and that ERT has no effect. The authors suggested an antioxidant supplementation in combination with ERT to enhance the therapeutic effect [35]. Concerning MPS I, oxidative stress has been observed in a mouse model [37,38]. Pereira et al. assessed oxidative stress in MPS I patients, compared with control subjects [31]. The authors detected a decrease in superoxide dismutase activity in erythrocytes from MPS I patients after ERT, while catalase activity increased after ERT compared to baseline levels. These findings could suggest that potential antioxidants might be included as adjuvants for current MPS therapies. Regarding arginine, its classification performance is interestingly comparable with that of the quantified GAGs (Table 3). Arginine is an amino acid which is involved in several key metabolisms, urea cycle, nitric oxid, polyamins, glutamate, proline and homoarginine. Furthermore, changes in arginine levels act as nutritional sensors and regulate cellular metabolism through its interaction with mammalian target of rapamycin complex 1 (mTORC1) [39]. Recently, Chantranupong et al. demonstrated that arginine sensing by mTORC1 is specific and depends on CASTOR1. The later interacts with GATOR2 to inhibit mTORC1 in low arginine condition. In the presence of arginine, CASTOR1 is bound to arginine, and thus free up GATOR2 and activate mTORC1 [40]. mTORC1 is a key regulator of protein synthesis, cell growth and autophagy [41]. Autophagy has been described as impaired in several LSDs [42] which may be attributed to lysosome dysfunction but may also be linked to arginine metabolism impairment. Woloszynek et al. observed profound metabolic alterations

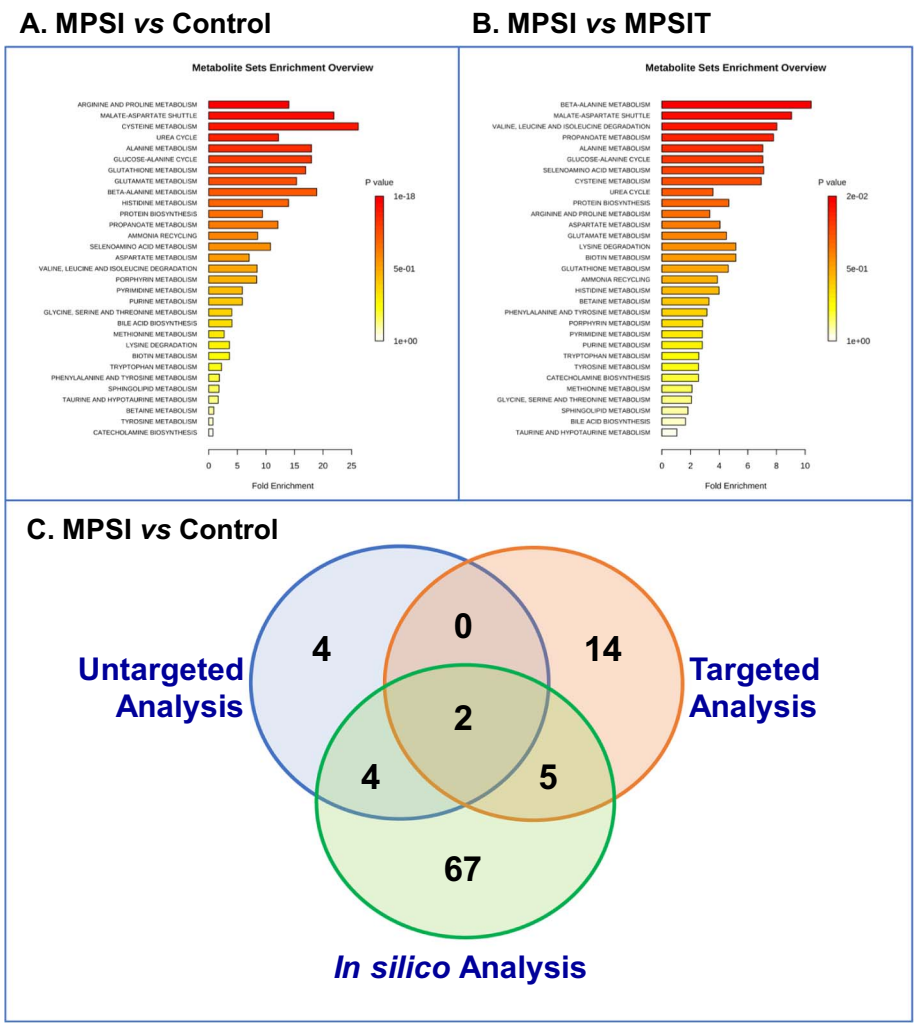

Fig. 3. Metabolite Set Enrichment Analysis using amino acid concentrations. 3A) MPS I vs. Control. 3B) MPS I vs. MPS IT. 3C) Venn diagram of the significant pathways retrieved from untargeted, targeted approaches and in silico systems biology approach from Salazar DA et al. [25]. The diagram shows two common metabolisms: Arginine-Proline metabolism and Cysteine-Glutathione metabolism. Detailed pathway information is given in Supporting information (Table S4).

in energy expenditure in MPS I mice, similar to those observed in the current study with an increase in most amino acids concentrations including dipeptides, amino acid derivatives, and urea [43]. The authors attributed these changes to an increase of protein catabolism and an autophagy disruption as a consequence of lysosome dysfunction. Interestingly, autophagic vacuoles number are increased in several LSDs, and reduced in MPS I mice fed a high-fat diet [43].

5. Conclusion

Metabolic phenotyping enabled us to unveil profound metabolic impairments beyond the primary deficiency in MPS I. The understanding of disease pathophysiological bases may open new therapeutic strategies such as antioxidants adjuvants and diet intervention as complementary treatments for MPS and maybe for other LSDs.

Acknowledgements

The authors are grateful to Carine Pilon, Thomas Plichet and Tony Pereira (Rouen University Hospital, France) for their technical assistance. The authors gratefully acknowledge the Région Normandie, the Labex SynOrg (ANR-11-LABX-0029), the European Regional Development Fund (ERDF 31708), Sanofi-Genzyme, Shire and Vaincre les Maladies Lysosomales for financial support.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cca.2017.09.024.

References

[1] A. Tebani, C. Afonso, S. Marret, S. Bekri, Omics-based strategies in precision medicine: toward a paradigm shift in inborn errors of metabolism investigations, Int. J. Mol. Sci. 17 (9) (2016).
[2] J.K. Nicholson, J.C. Lindon, E. Holmes, 'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data, Xenobiotica 29 (11) (1999) 1181–1189 the fate of foreign compounds in biological systems.
[3] S. Bekri, The role of metabolomics in precision medicine, Expert Rev. Precis. Med. Drug Dev. (2016) just-accepted.
[4] H.P. Benton, E. Want, H.C. Keun, A. Amberg, R.S. Plumb, F. Goldfain-Blanc, B. Walther, M.D. Reily, J.C. Lindon, E. Holmes, J.K. Nicholson, T.M. Ebbels, Intra- and interlaboratory reproducibility of ultra performance liquid chromatography-time-of-flight mass spectrometry for urinary metabolic profiling, Anal. Chem. 84 (5) (2012) 2424–2432.
[5] A. Tebani, L. Abily-Donval, C. Afonso, S. Marret, S. Bekri, Clinical metabolomics: the new metabolic window for inborn errors of metabolism investigations in the post-genomic era, Int. J. Mol. Sci. 17 (7) (2016).
[6] J.J. Pitt, Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry, Clin. Biochem. Rev. 30 (1) (2009) 19–34. Australian Association of Clinical Biochemists.
[7] J.J. Pitt, Newborn screening, Clin. Biochem. Rev. 31 (2) (2010) 57–68. Australian Association of Clinical Biochemists.
[8] Z. Spacil, H. Tatipaka, M. Barcenas, C.R. Scott, F. Turecek, M.H. Gelb, High-throughput assay of 9 lysosomal enzymes for newborn screening, Clin. Chem. 59 (3) (2013) 502–511.
[9] A. Tebani, C. Afonso, S. Bekri, Advances in metabolome information retrieval:

turning chemistry into biology. Part I: analytical chemistry of the metabolome, J. Inherit. Metab. Dis. (2017).
[10] A. Ballabio, V. Gieselmann, Lysosomal disorders: from storage to cellular damage, Biochim. Biophys. Acta 1793 (4) (2009) 684–696.
[11] J.E. Wraith, The mucopolysaccharidoses: a clinical review and guide to management, Arch. Dis. Child. 72 (3) (1995) 263–267.
[12] M.J. Neufeld EF, The mucopolysaccharidoses, in: B.A. Scrive C, W. Sly, D. Vaele (Eds.), The metabolic and molecular basis of inherited disease, Mc Graw-Hill, New York, NY, 2001, pp. 3421–3452.
[13] F. Baehner, C. Schmiedeskamp, F. Krummenauer, E. Miebach, M. Bajbouj, C. Whybra, A. Kohlschutter, C. Kampmann, M. Beck, Cumulative incidence rates of the mucopolysaccharidoses in Germany, J. Inherit. Metab. Dis. 28 (6) (2005) 1011–1017.
[14] G. Parenti, G. Andria, A. Ballabio, Lysosomal storage diseases: from pathophysiology to therapy, Annu. Rev. Med. 66 (2015) 471–486.
[15] A. Tebani, L. Zanoutene-Cheriet, Z. Adjtoutah, L. Abily-Donval, C. Brasse-Lagnel, A. Laquerriere, S. Marret, A. Chalabi Benabdellah, S. Bekri, Clinical and molecular characterization of patients with mucopolysaccharidosis type I in an Algerian series, Int. J. Mol. Sci. 17 (5) (2016).
[16] B. OA, Clinical Characteristics of MPS I Patients in the MPS I Registry, 2007 The American Society of Human Genetics, San Diego, 2007.
[17] A. Tebani, I. Schmitz-Afonso, D.N. Rutledge, B.J. Gonzalez, S. Bekri, C. Afonso, Optimization of a liquid chromatography ion mobility-mass spectrometry method for untargeted metabolomics using experimental design and multivariate data analysis, Anal. Chim. Acta 913 (2016) 55–62.
[18] J.G. de Jong, R.A. Wevers, R. Liebrand-van Sambeek, Measuring urinary glycosaminoglycans in the presence of protein: an improved screening procedure for mucopolysaccharidoses based on dimethylmethylene blue, Clin. Chem. 38 (6) (1992) 803–807.
[19] E.J. Langereis, T. Wagemans, W. Kulik, D.J. Lefeber, H. van Lenthe, E. Oussoren, A.T. van der Ploeg, G.J. Ruijter, R.A. Wevers, F.A. Wijburg, N. van Vlies, A multiplex assay for the diagnosis of mucopolysaccharidoses and mucolipidoses, PLoS One 10 (9) (2015) e0138622.
[20] X. Shen, X. Gong, Y. Cai, Y. Guo, J. Tu, H. Li, T. Zhang, J. Wang, F. Xue, Z.-J. Zhu, Normalization and integration of large-scale metabolomics data using support vector regression, Metabolomics 12 (5) (2016) 89.
[21] L. Eriksson, J. Trygg, S. Wold, A chemometrics toolbox based on projections and latent variables, J. Chemom. 28 (5) (2014) 332–346.
[22] B. Galindo-Prieto, L. Eriksson, J. Trygg, Variable influence on projection (VIP) for orthogonal projections to latent structures (OPLS), J. Chemom. (2014).
[23] Z. Zhou, X. Xiong, Z.J. Zhu, MetCCS Predictor: a web server for predicting collision cross-section values of metabolites in ion mobility-mass spectrometry based metabolomics, Bioinformatics (Oxford, England) (2017).
[24] A. Tebani, C. Afonso, S. Bekri, Advances in metabolome information retrieval: turning chemistry into biology. Part II: biological information recovery, J. Inherit. Metab. Dis. (2017).
[25] S. Li, Y. Park, S. Duraisingham, F.H. Strobel, N. Khan, Q.A. Soltow, D.P. Jones, B. Pulendran, Predicting network activity from high throughput metabolomics, PLoS Comput. Biol. 9 (2013) e1003123.
[26] J. Xia, I.V. Sinelnikov, B. Han, D.S. Wishart, MetaboAnalyst 3.0-making metabolomics more meaningful, Nucleic Acids Res. (2015).
[27] D.A. Salazar, A. Rodriguez-Lopez, A. Herreno, H. Barbosa, J. Herrera, A. Ardila, G.E. Barreto, J. Gonzalez, C.J. Almeciga-Diaz, Systems biology study of mucopolysaccharidosis using a human metabolic reconstruction network, Mol. Genet. Metab. 117 (2) (2016) 129–139.
[28] B. Donida, C.E.D. Jacques, C.P. Mescka, D.G.B. Rodrigues, D.P. Marchetti, G. Ribas, R. Giugliani, C.R. Vargas, Oxidative damage and redox in lysosomal storage disorders: biochemical markers, Clin. Chim. Acta 466 (2017) 46–53.
[29] J.C. Woloszynek, A. Kovacs, K.K. Ohlemiller, M. Roberts, M.S. Sands, Metabolic adaptations to interrupted glycosaminoglycan recycling, J. Biol. Chem. 284 (43) (2009) 29684–29691.
[30] A.V. Pshezhetsky, Crosstalk between 2 organelles: lysosomal storage of heparan sulfate causes mitochondrial defects and neuronal death in mucopolysaccharidosis III type C, Rare Dis. (Austin, Tex.) 3 (1) (2015) e1049793.
[31] V.G. Pereira, A.M. Martins, C. Micheletti, V. D'Almeida, Mutational and oxidative stress analysis in patients with mucopolysaccharidosis type I undergoing enzyme replacement therapy, Clin. Chim. Acta 387 (1–2) (2008) 75–79 International Journal of Clinical Chemistry.
[32] C.E. Jacques, B. Donida, C.P. Mescka, D.G. Rodrigues, D.P. Marchetti, F.H. Bitencourt, M.G. Burin, C.F. de Souza, R. Giugliani, C.R. Vargas, Oxidative and nitrative stress and pro-inflammatory cytokines in Mucopolysaccharidosis type II patients: effect of long-term enzyme replacement therapy and relation with glycosaminoglycan accumulation, Biochim. Biophys. Acta 1862 (9) (2016) 1608–1616.
[33] G.W. Negretto, M. Deon, G.B. Biancini, M.G. Burin, R. Giugliani, C.R. Vargas, Glycosaminoglycans can be associated with oxidative damage in mucopolysaccharidosis II patients submitted to enzyme replacement therapy, Cell Biol. Toxicol. 30 (4) (2014) 189–193.
[34] L. Filippon, C.A. Wayhs, D.M. Atik, V. Manfredini, S. Herber, C.G. Carvalho, I.V. Schwartz, R. Giugliani, C.R. Vargas, DNA damage in leukocytes from pre-treatment mucopolysaccharidosis type II patients; protective effect of enzyme replacement therapy, Mutat. Res. 721 (2) (2011) 206–210.
[35] B. Donida, D.P. Marchetti, G.B. Biancini, M. Deon, P.R. Manini, H.T. da Rosa, D.J. Moura, J. Saffi, F. Bender, M.G. Burin, A.S. Coitinho, R. Giugliani, C.R. Vargas, Oxidative stress and inflammation in mucopolysaccharidosis type IVA patients treated with enzyme replacement therapy, Biochim. Biophys. Acta 1852 (5) (2015) 1012–1019.
[36] L. Filippon, C.S. Vanzin, G.B. Biancini, I.N. Pereira, V. Manfredini, A. Sitta, C. Peralba Mdo, I.V. Schwartz, R. Giugliani, C.R. Vargas, Oxidative stress in patients with mucopolysaccharidosis type II before and during enzyme replacement therapy, Mol. Genet. Metab. 103 (2) (2011) 121–127.
[37] C.M. Simonaro, M. D'Angelo, X. He, E. Eliyahu, N. Shtraizent, M.E. Haskins, E.H. Schuchman, Mechanism of glycosaminoglycan-mediated bone and joint disease: implications for the mucopolysaccharidoses and other connective tissue diseases, Am. J. Pathol. 172 (1) (2008) 112–122.
[38] G.K. Reolon, A. Reinke, M.R. de Oliveira, L.M. Braga, M. Camassola, M.É. Andrades, J.C.F. Moreira, N.B. Nardi, R. Roesler, F. Dal-Pizzol, Alterations in oxidative markers in the cerebellum and peripheral organs in MPS I mice, Cell. Mol. Neurobiol. 29 (4) (2009) 443–448.
[39] S.M. Morris Jr., Arginine metabolism revisited, J. Nutr. 146 (12) (2016) 2579s–2586s.
[40] L. Chantranupong, S.M. Scaria, R.A. Saxton, M.P. Gygi, K. Shen, G.A. Wyant, T. Wang, J.W. Harper, S.P. Gygi, D.M. Sabatini, The CASTOR proteins are arginine sensors for the mTORC1 pathway, Cell 165 (1) (2016) 153–164.
[41] D.C. Goberdhan, C. Wilson, A.L. Harris, Amino acid sensing by mTORC1: intracellular transporters mark the spot, Cell Metab. 23 (4) (2016) 580–589.
[42] M. Chévrier, N. Brakch, L. Céline, D. Genty, Y. Ramdani, S. Moll, M. Djavaheri-Mergny, C. Brasse-Lagnel, A.L. Annie Laquerrière, F. Barbey, Autophagosome maturation is impaired in Fabry disease, Autophagy 6 (5) (2010) 589–599.
[43] J.C. Woloszynek, A. Kovacs, K.K. Ohlemiller, M. Roberts, M.S. Sands, Metabolic adaptations to interrupted glycosaminoglycan recycling, J. Biol. Chem. 284 (43) (2009) 29684–29691.