



OPEN ACCESS

EDITED BY

Julia Kuligowski,
La Fe Health Research Institute, Spain

REVIEWED BY

Annalaura Mastrangelo,
Spanish National Centre for
Cardiovascular Research, Spain
Raúl González-Dominguez,
University of Cádiz, Spain
Guillermo Quintas,
Leitat Technological Center, Spain

*CORRESPONDENCE

Gabriel A. Martos-Moreno,
✉ gabrielangelmartos@yahoo.es
Francisco J. Rupérez,
✉ ruperez@ceu.es

RECEIVED 25 September 2023

ACCEPTED 30 November 2023

PUBLISHED 20 December 2023

CITATION

Chamoso-Sanchez D, Rabadán Pérez F,
Argente J, Barbas C, Martos-Moreno GA
and Rupérez FJ (2023), Identifying
subgroups of childhood obesity by using
multiplatform metabotyping.
Front. Mol. Biosci. 10:1301996.
doi: 10.3389/fmolb.2023.1301996

COPYRIGHT

© 2023 Chamoso-Sanchez, Rabadán
Pérez, Argente, Barbas, Martos-Moreno
and Rupérez. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is
permitted, provided the original author(s)
and the copyright owner(s) are credited
and that the original publication in this
journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Identifying subgroups of childhood obesity by using multiplatform metabotyping

David Chamoso-Sanchez¹, Francisco Rabadán Pérez²,
Jesús Argente^{3,4,5}, Coral Barbas¹, Gabriel A. Martos-Moreno^{3,4*}
and Francisco J. Rupérez^{1*}

¹Centro de Metabolómica y Bioanálisis (CEMBIO), Facultad de Farmacia, Universidad San Pablo-CEU, CEU Universities, Boadilla del Monte, Spain, ²Departamento Economía Aplicada I, Universidad Rey Juan Carlos, Madrid, Spain, ³Department of Pediatrics and Pediatric Endocrinology, Hospital Infantil Universitario Niño Jesús, Instituto de Investigación Sanitaria La Princesa, Universidad Autónoma de Madrid, Madrid, Spain, ⁴CIBER Fisiopatología de la Obesidad y Nutrición (CIBEROBN), Instituto de Salud Carlos III, Madrid, Spain, ⁵IMDEA Food Institute, Madrid, Spain

Introduction: Obesity results from an interplay between genetic predisposition and environmental factors such as diet, physical activity, culture, and socioeconomic status. Personalized treatments for obesity would be optimal, thus necessitating the identification of individual characteristics to improve the effectiveness of therapies. For example, genetic impairment of the leptin-melanocortin pathway can result in rare cases of severe early-onset obesity. Metabolomics has the potential to distinguish between a healthy and obese status; however, differentiating subsets of individuals within the obesity spectrum remains challenging. Factor analysis can integrate patient features from diverse sources, allowing an accurate subclassification of individuals.

Methods: This study presents a workflow to identify metabolotypes, particularly when routine clinical studies fail in patient categorization. 110 children with obesity (BMI > +2 SDS) genotyped for nine genes involved in the leptin-melanocortin pathway (CPE, MC3R, MC4R, MRAP2, NCOA1, PCSK1, POMC, SH2B1, and SIM1) and two glutamate receptor genes (GRM7 and GRIK1) were studied; 55 harboring heterozygous rare sequence variants and 55 with no variants. Anthropometric and routine clinical laboratory data were collected, and serum samples processed for untargeted metabolomic analysis using GC-q-MS and CE-TOF-MS and reversed-phase U(H)PLC-QTOF-MS/MS in positive and negative ionization modes. Following signal processing and multialignment, multivariate and univariate statistical analyses were applied to evaluate the genetic trait association with metabolomics data and clinical and routine laboratory features.

Results and Discussion: Neither the presence of a heterozygous rare sequence variant nor clinical/routine laboratory features determined subgroups in the metabolomics data. To identify metabolomic subtypes, we applied Factor Analysis, by constructing a composite matrix from the five analytical platforms. Six factors were discovered and three different metabolotypes. Subtle but neat differences in the circulating lipids, as well as in insulin sensitivity could be established, which opens the possibility to personalize the treatment according to the patients categorization into such obesity subtypes. Metabotyping in clinical contexts poses challenges due to the influence of various uncontrolled variables on metabolic phenotypes. However, this strategy reveals the potential to identify

subsets of patients with similar clinical diagnoses but different metabolic conditions. This approach underscores the broader applicability of Factor Analysis in metabotyping across diverse clinical scenarios.

KEYWORDS

multiplatform metabolomics, factor analysis, data integration, obesity, childhood, leptin-melanocortin pathway

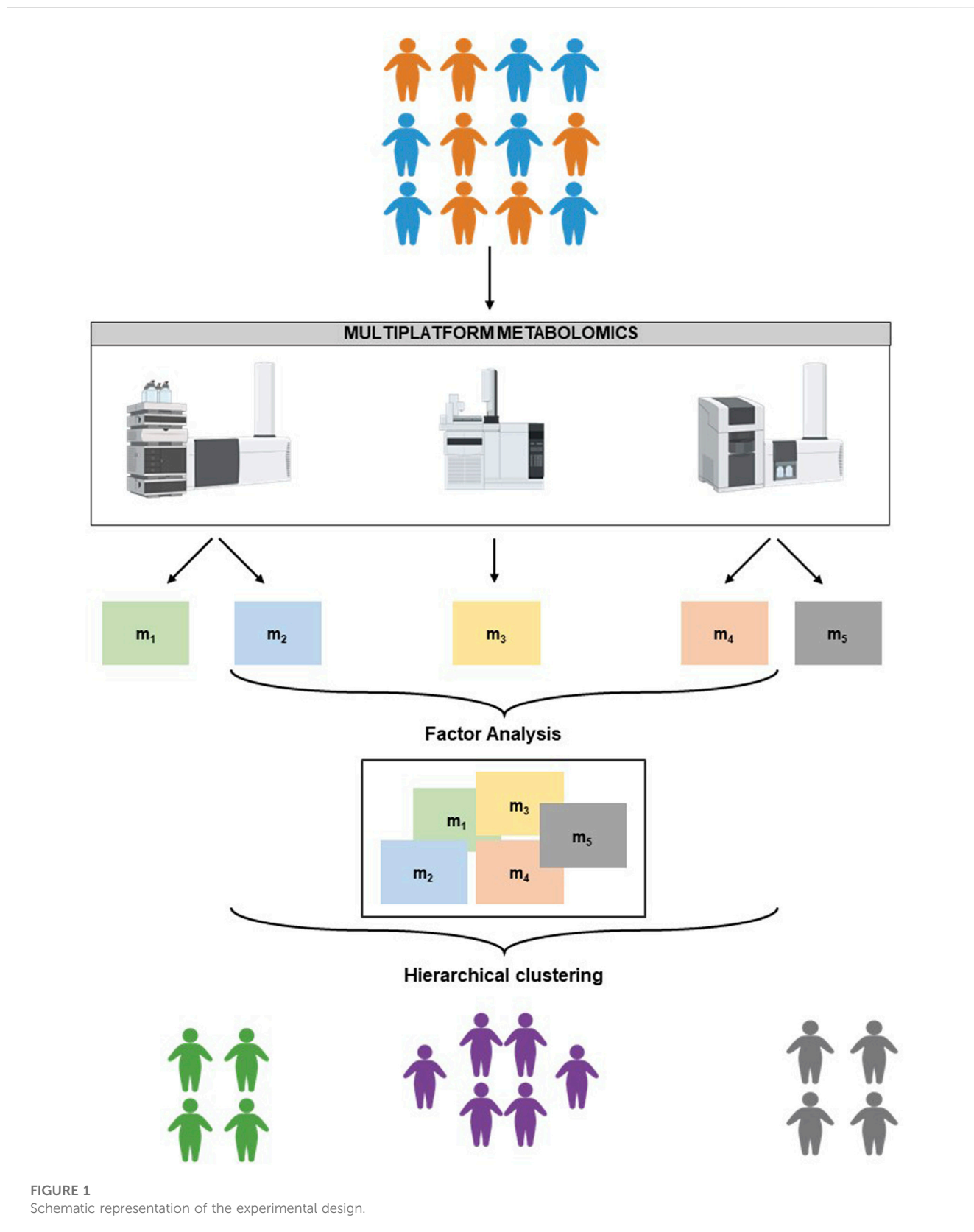
Introduction

Childhood obesity prevalence has increased worldwide in the last decades, including a higher incidence of severe and early onset cases, particularly after the COVID-19 outbreak lockdown (Choi et al., 2023), enhancing the known risk for long-term consequences in these patients (Rupérez et al., 2020; Handakas et al., 2022). Children with obesity are more susceptible to maintain their adiposity in adult life, increasing the risk of multiple comorbidities at an early age, including type 2 diabetes mellitus (T2DM), dyslipidemia, cardiovascular disease (CVD), hypertension, obstructive sleep apnea, cancer and steatohepatitis (da Fonseca et al., 2017; Cote et al., 2013; Butte et al., 2015; Carde et al., 2020; Berger, 2018; Wahl et al., 2012). Obesity has a multifactorial etiology, with lifestyle, including nutritional and physical activity habits, as well as other environmental factors, interacting with an individual's unique genetic background to determine a person's risk to develop obesity (Trang and Grant, 2023). Among the large set of genes influencing obesity, those in the leptin-melanocortin satiety signaling pathway are the most determinant known to date, with homozygous mutations in some causing early onset severe obesity with hyperphagia (Jackson et al., 1997; Chiurazzi et al., 2020; Trang and Grant, 2023). The role of heterozygous variants is under investigation (Trang and Grant, 2023), particularly those with confirmed pathogenicity or high Combined Annotation Dependent Depletion (CADD) scores of "deleteriousness" with low population prevalence [heterozygous rare sequence variants (HetRSVs)]. Additionally, variants in glutamate receptors, pivotal in neuron signaling have also been described in patients with severe obesity (Bell et al., 2005; Fuente-Martín et al., 2016; Serra-Juhé et al., 2017; Fairbrother et al., 2018; Chiurazzi et al., 2020).

Whereas some obesity-associated comorbidities commonly identified in adults can also be observed in children with obesity, others such as T2DM are far less common, with insulin resistance (IR) usually found as the first step in carbohydrate metabolism impairment in childhood obesity (Martos-Moreno et al., 2019). Additionally, not every patient with obesity shows the same risk to develop comorbidities, with the "metabolically healthy obesity" designation proposed for those patients with obesity, even severe obesity, but with no metabolic comorbidities (Wan Mohd Zin et al., 2022). However, this term is under discussion and this condition is known to evolve throughout life in relationship to weight control (Martos-Moreno et al., 2021). The term "metabotype" was defined by Gavaghan et al. (Gavaghan et al., 2000) as "a probabilistic multiparametric description of an organism in a given physiological state based on analysis of its cell types, biofluids, or tissues." Subsequently, this definition has been repeatedly used (Waldram et al., 2009; Sullivan et al., 2011; Palmnäs et al., 2020), establishing itself as the characterization of the metabolic phenotype

of an individual. Recent advances in high-throughput sequencing technologies and computational methods have enabled the generation of large and complex *-omics* datasets, providing an unprecedented opportunity to integrate simultaneous information from multiple molecular levels to investigate the complexity of biological systems (T et al., 2019; Park et al., 2022; Argelaguet et al., 2020; Tanabe et al., 2021; Clark et al., 2021). The integration of various *-omics* data, including genomics, transcriptomics, proteomics, metabolomics, and epigenomics, can help to understand the intricate interplay between different biological molecules and pathways, enabling the identification of key regulators and mechanisms of disease (Hoadley et al., 2014; Meng et al., 2016; Marabita et al., 2022). In metabolomics, a multiplatform strategy combines many analytical tools to study the entire metabolic phenotype. Combining data from multiple sources could result in a better comprehension of the underlying biological mechanisms driving complex diseases including cancer, obesity, and cardiovascular disease (Hoadley et al., 2014; Meng et al., 2016; Marabita et al., 2022; Park et al., 2022). Although much effort has been made in recent years to integrate information from different *-omics* technologies into a single analysis, it is still usual to use a multiplatform strategy individually (T et al., 2019; Argelaguet et al., 2020; Tanabe et al., 2021; Zhang et al., 2022).

Factor Analysis is a multivariate statistical technique that can identify underlying patterns in a large dataset by reducing the number of variables into a smaller number of factors (Lee et al., 2019; Acal et al., 2020). In the context of metabolomics, Factor Analysis can identify metabolite modules, which are groups of metabolites that are highly correlated and potentially involved in a common biological process. This approach provides a more comprehensive understanding of the underlying molecular mechanisms of disease and can identify potential biomarkers and therapeutic targets that may not be identifiable using individual metabolites. Recent studies have demonstrated the potential of Factor Analysis in metabolomics for identifying metabolite modules in various fields of research including cancer biology, metabolic disorders, and neurodegenerative diseases (Shen et al., 2009; Zhao et al., 2013; Argelaguet et al., 2018; Kamleh et al., 2018; Clark et al., 2021). However, there are several challenges associated with the application of Factor Analysis in metabolomics. One of the key challenges is the selection of an appropriate Factor Analysis method (principal component analysis, common Factor Analysis, maximum likelihood method, etc.) which depends on the specific research questions and the characteristics of the metabolomics dataset. Also, multicollinearity is a serious problem that must be solved before performing a Factor Analysis (Chan et al., 2022). Another challenge is the interpretation of the identified metabolite modules, as it may be difficult to determine the biological relevance of the modules. This challenge can be addressed by integrating the



results of Factor Analysis with other omics data types, such as genomics, transcriptomics, and proteomics, to provide a more comprehensive understanding of the underlying biological processes. Combining Factor Analysis with a hierarchical

clustering analysis enables one to classify patients considering all metabolic features detected by a multi-platform approach; to identify patient subgroups based on their metabotype and to provide the optimal treatment for each patient rather than based

TABLE 1 Gene list.

<i>CPE</i> (MIM* 114855. Carboxypeptidase E)
<i>GRIK1</i> (MIM* 138245. GLUTAMATE RECEPTOR, IONOTROPIC, KAINATE 1)
<i>GRM7</i> (MIM* 604101. GLUTAMATE RECEPTOR, METABOTROPIC, 7)
<i>MC3R</i> (MIM* 155540. MELANOCORTIN 3 RECEPTOR)
<i>MC4R</i> (MIM* 155541. MELANOCORTIN 4 RECEPTOR)
<i>MRAP2</i> (MIM* 615410. MELANOCORTIN 2 RECEPTOR ACCESSORY PROTEIN 2)
<i>NCOA1</i> (MIM* 602691. NUCLEAR RECEPTOR COACTIVATOR 1) (Alternative nomenclature: SRC1)
<i>PCSK1</i> (MIM* 162150. PROPROTEIN CONVERTASE, SUBTILISIN/KEXIN-TYPE, 1)
<i>POMC</i> (MIM* 176830. PROOPIOMELANOCORTIN)
<i>SH2B1</i> (MIM* 608937. SH2B ADAPTOR PROTEIN 1)
<i>SIMI</i> (MIM* 603128. SIM bHLH TRANSCRIPTION FACTOR 1)

upon the usual anthropometric and routine laboratory parameters used in the clinical setting and even over the presence or absence of HetRSVs in relevant genes in the studied pathology. Such strategy becomes even more powerful when there is no classification available, or the main goal of the research is to unveil the minimum set of parameters which allow for classification/stratification.

Patients, materials and methods

We tested a multi-platform strategy in combination with Factor Analysis and hierarchical clustering for personalized approaches in the treatment of obesity (Figure 1).

Patients

One hundred and ten children and adolescents (57 females/53 males) affected with severe obesity referred to a specialized clinic in a third level monographic pediatric hospital and genotyped for nine genes in the leptin-melanocortin pathway downstream of the leptin receptor, and two glutamate receptor genes (Table 1) were studied: 55 of them harboring one heterozygous rare sequence variant [HetRSV, defined as populational frequency <0.01 and with a Combined Annotation Dependent Depletion (CADD) score of “deleteriousness” > 20] and/or confirmed pathogenicity according to ACMG criteria in the studied genes [*CPE* ($n = 5$), *MC3R* ($n = 5$), *MC4R* ($n = 5$), *MRAP2* ($n = 5$), *NCOA1* ($n = 7$), *PCSK1* ($n = 5$), *POMC* ($n = 5$), *SH2B1* ($n = 5$), *SIMI* ($n = 5$), *GRM7* ($n = 4$) or *GRIK1* ($n = 4$)] and 55 with no detected variants.

The whole cohort mean age and standardized body mass index (BMI) were 11.01 ± 3.36 years and 4.20 ± 2.20 SDS, respectively with no differences between groups (with vs. without variants) in age, BMI-SDS, routine laboratory metabolic and hormonal features nor in sex, ethnicity, or pubertal status distribution. Their main anthropometric and metabolic features are summarized and compared in Table 2.

All patients and their parents or guardians gave informed written consent as required by the ethics committee at the University Hospital Niño Jesús, which had previously approved the study in accordance with the “Ethical Principles for Medical Research Involving Human Subjects” adopted in the Declaration of Helsinki by the World Medical Association (64th WMA General Assembly, Fortaleza, Brazil, October 2013).

Methods

Weight, height, BMI, waist circumference, and systolic and diastolic blood pressure (BP, mean of three measurements) were recorded and standardized (Cole et al., 2000; Ferná et al., 2004) in all patients. A 12-hour fasting serum sample (drawn, immediately processed, aliquoted and stored at -80°C until assayed) was used to determine glucose, insulin, HbA1c, lipid profile, uric acid, GOT, GPT, GGT, free thyroxin, thyroid stimulating hormone, IGF-I, IGFBP-3, 25-OH-vitamin D and intact parathyroid hormone (iPTH) levels by standardized assays as previously reported (Martos-Moreno et al., 2019). An oral glucose tolerance test (OGTT, 1.75 g/kg, maximum 75 g) for glucose and insulin determination at 30, 60 and 120 min was performed, HOMA (homeostatic model for insulin resistance) and WBISI (whole body insulin sensitivity) indexes were calculated as previously reported (Martos-Moreno et al., 2019).

Multiplatform untargeted metabolomics analysis

Sample treatment

Serum metabolite extraction was carried out according to previously reported standard protocols (Garcia and Barbas, 2011; Pellegrino et al., 2014; Naz et al., 2015). Briefly, for LC-MS analysis, 40 μL of serum was mixed with 800 μL of a cold mixture (-20°C) of methanol:MTBE:Chloroform (1.33:1:1, v/v/v) with Sphinganine (D17:0) and palmitic acid-d31 as internal standards. Samples were vortexed for 30 s and shaken for 20 min at maximum speed at room temperature. Next, samples were centrifuged (13,200 rpm, room temperature, 5 min). After centrifugation, supernatant was directly injected into the system. For GC-MS analysis, protein precipitation was achieved by mixing one volume of serum with three volumes of cold (-20°C) acetonitrile with 25 ppm of palmitic acid-d31 as internal standard, followed by methoximation with O-methoxyamine hydrochloride (15 mg/mL) in pyridine, and silylation with BSTFA: TMCS (99:1). Finally, 20 ppm of tricosane in heptane was added as second internal standard. For CE-MS analysis, 100 μL of serum was mixed with 100 μL of 0.2 M formic acid containing 5% acetonitrile and 0.4 mM methionine sulfone, 2 mM paracetamol and 0.5 mM 4-Morpholineethanesulfonic acid, 2-(N-Morpholino) ethanesulfonic acid (MES) as internal standards. The sample was transferred to an ultracentrifugation device (Millipore Ireland Ltd., Carrigtohill, Ireland) with a 30 kDa protein cutoff for deproteinization through centrifugation ($2000 \times g$, 4°C , 90 min). Detailed version of the sample treatment protocols, the reagents, solvents, standards used for the sample treatment and subsequent analyses, and the analytical setup for the LC-MS, GC-MS, and CE-MS analysis are described in

TABLE 2 Anthropometric and metabolic features.

Clinical parameters	Whole cohort	Variant carriers	NO variant	Variant carriers vs. NO
Age (years)	11.01 ± 3.36	11.12 ± 3.47	10.91 ± 3.28	N.S.
Height (SDS)	0.84 ± 1.12	0.70 ± 1.05	0.99 ± 1.19	N.S.
BMI-SDS	4.20 ± 2.20	4.29 ± 2.42	4.11 ± 1.99	N.S.
Fasting glucose (mg/dL)	91.10 ± 6.81	90.99 ± 5.98	91.91 ± 7.51	N.S.
Glucose at 120'in OGTT (mg/dL)	120.82 ± 18.87	118.11 ± 13.31	123.31 ± 18.29	N.S.
Fasting insulin (μU/mL)	15.13 ± 7.28	14.83 ± 7.56	15.42 ± 7.05	N.S.
HOMA index	3.45 ± 1.75	3.34 ± 1.78	3.56 ± 1.72	N.S.
WBISI index	3.23 ± 1.67	3.35 ± 1.63	3.13 ± 1.71	N.S.
HbA1c (%)	5.43 ± 0.30	5.44 ± 0.24	5.41 ± 0.35	N.S.
Uric acid (mg/dL)	5.18 ± 1.16	5.04 ± 1.21	5.32 ± 1.11	N.S.
GOT (U/L)	27.72 ± 7.35	26.80 ± 6.73	28.68 ± 7.89	N.S.
GPT (U/L)	22.22 ± 9.27	21.38 ± 8.48	23.19 ± 10.03	N.S.
GGT (U/L)	14.16 ± 4.72	13.91 ± 4.31	14.42 ± 5.13	N.S.
HDL-c (mg/dL)	46.72 ± 14.02	46.63 ± 15.93	46.82 ± 11.90	N.S.
LDL-c (mg/dL)	96.21 ± 27.23	96.82 ± 30.57	97.62 ± 23.55	N.S.
Triglycerides (mg/dL)	78.69 ± 48.61	75.58 ± 42.40	81.85 ± 54.43	N.S.
Free thyroxine (T4) (ng/dL)	0.94 ± 0.13	0.95 ± 1.09	0.92 ± 0.154	N.S.
TSH (μU/mL)	2.78 ± 1.57	2.91 ± 1.64	2.65 ± 1.51	N.S.
IGF-I (ng/mL)	310.8 ± 171.7	324.67 ± 191.76	297.77 ± 151.24	N.S.
IGFBP-3 (μg/mL)	4.90 ± 0.99	4.93 ± 0.97	4.86 ± 1.02	N.S.
25-OH-Vitamin D (ng/mL)	23.36 ± 9.22	23.55 ± 8.69	23.21 ± 9.74	N.S.
Intact PTH (pg/mL)	57.13 ± 21.92	57.03 ± 22.04	57.21 ± 22.13	N.S.

Abbreviations: BMI-SDS, Standardized body mass index (Z-score); OGTT, oral glucose tolerance test; HOMA, homeostatic model assessment; WBISI, Whole-body insulin sensitivity index; HDL-c, High density lipoprotein cholesterol; LDL-c, Low density lipoprotein cholesterol; GGT, Gamma-glutamyltransferase; GOT, Glutamic-oxalacetic transaminase; GPT, glutamic-pyruvic transaminase Alanine aminotransferase; IGF-1, Insulin-like growth factor 1; IGFBP3, Insulin-like Growth Factor-binding Protein 3; HbA1c, hemoglobin A1c; TSH, thyroid-stimulating hormone; PTH, parathyroid hormone.

Supplementary Material. Quality control samples (QC) were prepared by pooling and mixing equal volumes of each serum sample and treated as independent samples to check the performance of the systems and the reproducibility of the sample treatment. Then, samples were randomized, and QCs were injected at the beginning, along the sequence, and at the end of the batch. Finally, two blank solutions were prepared along with the rest of the samples and analyzed at the beginning and at the end of the analytical sequence (Dudzik et al., 2018).

LC-MS and CE-MS data pre-processing

The raw data obtained after the LC-MS and CE-MS analysis were processed using Agilent Technologies MassHunter Profinder B.10.0.2.162 (Santa Clara, United States) to clean the background noise and unrelated ions. This algorithm aligns all ions across the samples using mass and retention time (RT) to create a single spectrum for each group of compounds, and finally obtaining a structured data matrix and appropriate format. Missing values were

imputed using the k-nearest neighbors (kNN) algorithm (Armitage et al., 2015) in Matlab R2022a software (Mathworks, Inc., Natick, United States). Then, the data matrix was filtered by coefficient of variation (CV), maintaining those signals that, in the QCs, presented a CV below 30%. The filtered data matrix was imported into SIMCA 17 Sartorius (Goettingen, Germany) to generate a PCA and thus observe the trend of the QCs, detect possible outliers, and look for natural and analytical trends of the samples. To reduce the impact of instrumental and experimental variations that can interfere with the ability to detect biological variations, a correction method called “quality control samples and support vector regression (QC-SVRC)” was used to adjust the data (Kuligowski et al., 2015) implemented in MATLAB R2022a and then normalized by internal standard (IS).

Data pre-processing and compound identification GC-MS analysis

The chromatograms obtained from each of the serum samples, the QCs, and the IS signal were visually examined to ensure the

quality of the obtained profiles and the reproducibility of the IS signal using Agilent MassHunter Qualitative B.10.0.010305.0 software (Santa Clara, United States). Deconvolution and metabolite identification was achieved using the Agilent MassHunter Unknowns Analysis Tool 10.0 (Santa Clara, United States). The software assigned a chemical identity to each of the signals obtained after the search in two commercial libraries: the Fiehn library version 2013, and the NIST library version 2017 and “in-house” libraries. The identities were assigned according to the retention time (RT) and spectra extracted during deconvolution when the software compared them with each compound included in the libraries. Next, the obtained data were aligned using the MassProfiler Professional B.15.1 software (Agilent Technologies) (Santa Clara, United States) and exported to Agilent MassHunter Quantitative Analysis version B10.0.707.0 (Santa Clara, United States) to assign the main ions and the integration of each of the signals. As in the LC-MS and CE-MS analysis, the missing values were estimated using the kNN (k-nearest neighbors) algorithm (Armitage et al., 2015). Experimental and analytical variations were excluded by performing normalization. As in the LC-MS and CE-MS analysis the data matrix was normalized by applying the QC-SRVC correction, normalized by internal standard, and filtered by CV in the QCs (Kuligowski et al., 2015).

Compound identification LC-MS and CE-MS analysis

For the metabolite tentative annotation initially the m/z was searched against multiple databases available online, including METLIN (<http://metlin.scripps.edu>), LipidsMAPS (<http://lipidMAPS.org>) and KEGG (<http://www.genome.jp/kegg/>), all of which have been joined into an “in-house” developed search engine, CEU MassMediator (<http://ceumass.eps.uspceu.es/>) (Gil-de-la-Fuente et al., 2019). Aiming to obtain additional information for some identities, HMDB (<http://hmdb.ca>) was also consulted. In parallel, three complementary software, MS-DIAL (<http://prime.psc.riken.jp/>), LipidAnnotator (Agilent Technologies) and LipidHunter (Ni et al., 2017; Koelme et al., 2020; Tsugawa et al., 2020) by fragmentation mass/mass spectra were used for LC-MS identification. Features that were tentatively assigned to metabolites from the databases were based on (1): mass accuracy (maximum error mass 20 ppm) (2), isotopic pattern distribution (3), possibility of cation and anion formation (4), adduct formation (5), elution order of the compounds based on the chromatographic conditions, and (6) MS/MS spectra. Additionally, an “in-house” CE-MS library built with authentic standards was used to compare the relative migration time (RMT) to increase the confidence of the annotations. The confidence levels established by the Compound Identification group of the Metabolomics Society at the 2017 annual meeting of the Metabolomics Society (Brisbane, Australia) have been used. The new identification levels (Blaženović et al., 2018) range from level 0 with full identification based on knowledge of its 3D structure, level 1 2D confidence using comparison with reference standards, level 2 probable structure when compared with database, level 3 possible structure or class and level 4 as unidentified compound.

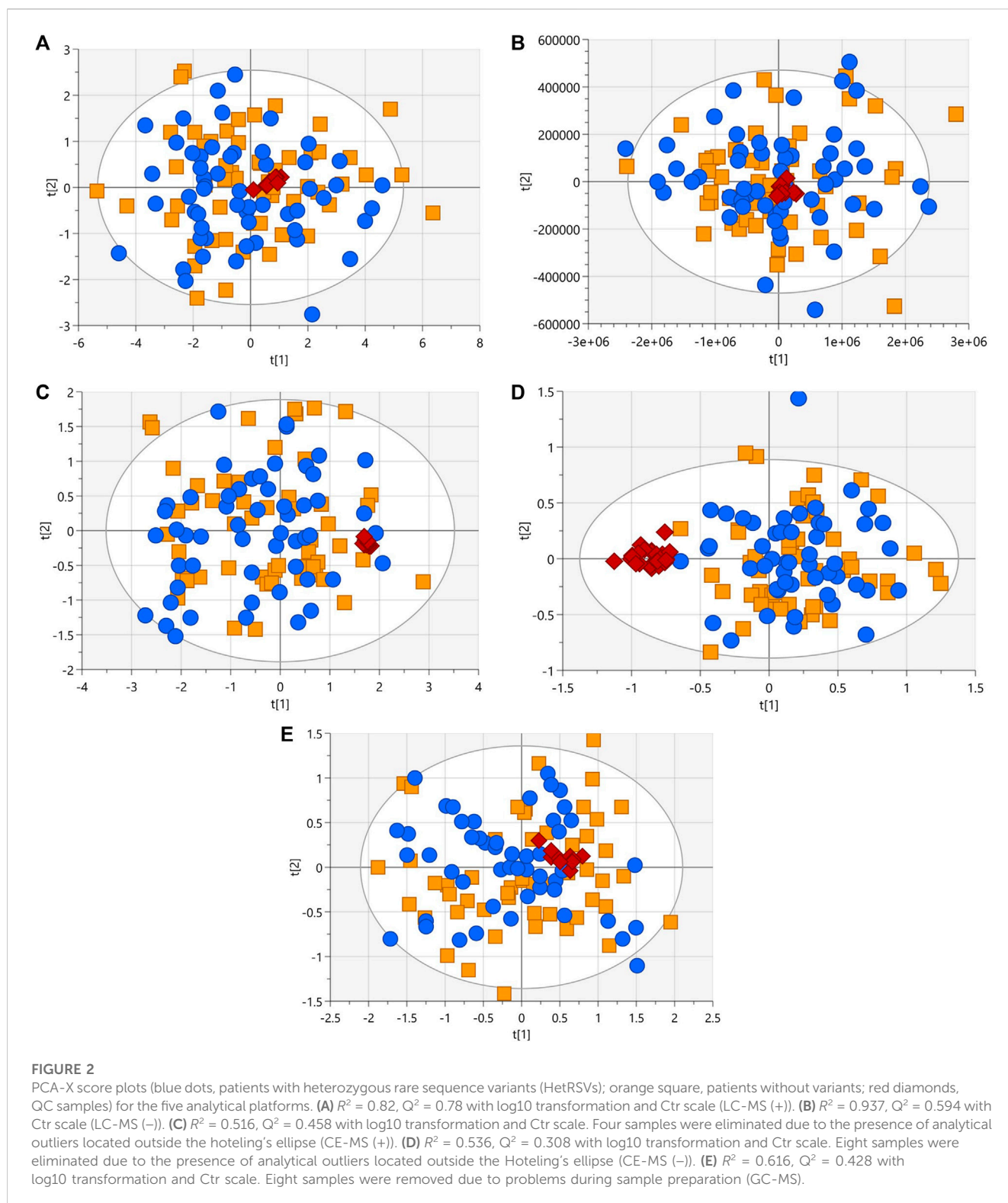
Statistical analysis

Statistical analysis was carried out by univariate (UVA, Matlab R2022a) and multivariate analysis [MVA, SIMCA 17, R v4.1.2 and IBM SPSS v27 (Armonk, NY, United States)]. For the UVA, parametric (unpaired t -test) with a Benjamini–Hochberg False Discovery Rate *post hoc* correction ($q < 0.05$) was applied. For MVA, the PCA plot, PLS-DA plot and OPLS-DA plot was built. The data matrix was analyzed using unsupervised machine learning using R environment (<https://www.r-project.org/>), applying clustering technique to obtain pattern in our data independently of the initial groups.

The raw data from the various analytical platforms were merged using Factor Analysis and hierarchical clustering to generate a broad perspective of the results and to assign metabolotypes based on the metabolic phenotypes of each patient with obesity. The whole process of Factor Analysis and hierarchical clustering was carried out by using IBM SPSS software and Microsoft Excel. First, the Pearson correlations between the variables in each of the matrices were analyzed to eliminate multicollinearity. Correlations between the various matrices (inter-matrix correlations) were examined after filtering by the specific correlations of each matrix (intra-matrix correlations). The individual matrices with the resulting variables were subjected to principal component analysis with varimax rotation (Acal et al., 2020) to reduce dimensionality. Three rules were applied to select the number of principal components in each of the individual matrices, the “Scree plot elbow,” the Kaiser-Guttman test (Eigenvalue greater than unity) and a total explained variance greater than 60% (Cattell, 1966). The principal component scores have been analyzed. The variables that present a principal component score higher than 0.5 in any of the selected components and that do not present double saturation are kept for the subsequent Factor Analysis. We consider double saturation to be when the smallest difference in the principal component score of a variable between two components is less than 0.1. The resulting variables have been subjected to a Factor Analysis by maximum likelihood (Babakus et al., 1987) with varimax rotation of each of the matrices separately to further reduce dimensionality. The same three rules were applied as in the PCA. For the final Factor Analysis, variables that displayed double saturation or had factor scores lower than 0.5 in any of the chosen factors were excluded. Finally, all the resulting variables were pooled into a single matrix after applying all these filters and a Factor Analysis was performed using a maximum likelihood extraction method and a varimax rotation method. All variables that entered the combined Factor Analysis of the different platforms were identified using internal databases and mass/mass fragmentation spectra software (LipidAnnotator, MS-Dial, LipidHunter). Following the criteria applied above, the appropriate number of factors was selected for our data and by regression new variables were created for each of the factors. A hierarchical clustering analysis with squared Euclidean distance and Ward method was applied on the created factors. To select the appropriate number of metabolotypes, a discriminant analysis (DA) was performed (Lee et al., 2019).

Results

1) The presence of heterozygous rare sequence variant in the studied genes, associated to human obesity and energy



homeostasis, does not determine different metabolomic phenotypes.

After following the procedure described in the patients and methods section, we obtained 345 and 170 metabolic features in LC-MS performed in positive and negative ionization modes, 63 signals in GC-MS, and finally in CE-MS we obtained 242 signals in positive

ionization and 91 in negative ionization mode. The visual inspection of the PCA plots built for all techniques revealed a tight cluster of the QCs assessing the analytical stability and reproducibility (Figure 2). A homogeneous distribution of patients with and without heterozygous variant was seen in PCA plots.

The identified HetRSVs did not allow for the construction of multivariate supervised model from the results of only one of the

analytical techniques. Only LC-MS (+), enabled the creation of an OPLS-DA model ($R_2X = 0.72$, $R_2Y = 0.82$; $Q_2 = 0.64$; p CV-ANOVA_{OPLS-DA} = $3.1 \cdot 10^{-19}$; as illustrated in [Supplementary Figure S1](#)). Correspondingly, the results of three metabolites derived from LC-MS (+) displayed statistically significant differences in means between the variant carrier/no variant groups, whereas no discernible differences between both groups were observed in the means of all the variables from LC-MS (-), GC-MS, CE-MS (+), or CE-MS (-). Furthermore, despite the limited number of samples for each individual gene, the presence of singular metabolic patterns was not observed for any of the studied genes in any of the employed platforms (data not shown).

2) Factor Analysis groups the variability into six factors

Imprecise information is obtained from the examination of the metabolic phenotype using a single analytical platform, which might result in the description of erroneous metabolotypes in patients, generating different classifications depending on the analytical platform used ([Supplementary Figure S2](#)). Furthermore, the use of a classification based on anthropometric and routine laboratory metabolic and hormonal parameters available in daily clinical practice does not appear to be sufficient to establish distinct metabolotype among patients. We also performed a hierarchical clustering analysis, with anthropological and clinical parameters (data not shown). We observed dissimilar outcomes when compared to the classifications produced by individual analytical platforms. Additionally, no statistically significant differences were observed in any of the analytical platforms with the clusters (possible groups) generated after analyzing these parameters.

The multiplatform strategy provided five matrices with 345, 170, 53, 242 and 91 variables analyzed by LC-MS (+), LC-MS (-), GC-MS, CE-MS (+) and CE-MS (-), respectively, from 100 of the studied samples. Due to the presence of analytical outliers caused by errors during sample preparation in GC-MS and analytical error in CE-MS (+), 10 samples had to be eliminated from the analysis of the total of 110 patients enrolled (three patients without genetic variants, and 7 with genetic variants, with a maximum of two individuals per gene studied). As described in detail above (see materials and methods) a Factor Analysis of each of the matrices was performed to subsequently combine the variables present in each Factor Analysis into a single combined Factor Analysis. To eliminate multicollinearity, Pearson correlations were used to analyze the relationships between variables in each matrix (intra-matrix correlations) and between matrices (inter-matrix correlations). In the final Factor Analysis by maximum likelihood and varimax rotation performed on the LC-MS matrix (+), three factors were chosen that explained 75% of the variability accumulated in the matrix, saturating 57 variables that were kept for the final combined Factor Analysis. The Kaiser-Meyer-Olkin (KMO) test was used to determine whether the Factor Analysis was effective, and a result of 0.85 was obtained. In LC-MS (-), two factors were chosen to explain 74.68% of the variability accumulated in the matrix, obtaining a KMO of 0.85 and saturating 17 variables. In GC-MS, three factors were chosen that explain 73.54% of the accumulated variability, obtaining a KMO of 0.89 with 18 variables independently saturated in these factors. In CE-MS (+) 1 factor was chosen that explained 64.09% of the accumulated variation, obtaining a KMO of 0.77 with

five saturated variables. In CE-MS (-) no satisfactory factor extraction was achieved, so no variable was retained for the final Factor Analysis. Using in-house databases and mass/mass fragmentation spectra software (LipidAnnotator, MSDial, LipidHunter) all variables that remained after all of these pre-filtering stages for the combined Factor Analysis of the various platforms were identified. The removal of non-annotated variables from the combined Factor Analysis (23 out of 97 variables were removed due to unsuccessful identification) is performed to determine the biological interpretation of the obtained factors. Therefore, variables 43, 11, 14 and 5 analyzed by LC-MS (+), LC-MS (-), GC-MS and CE-MS (+), respectively, were pooled together and the Factor Analysis was performed. The adequacy of the Factor Analysis was tested using the KMO test, obtaining a value of 0.76. Finally, six factors that explained 75% of the accumulated variability were selected. The results show a clustering of the variables into factors depending on the analytical technique. [Table 3](#) shows the variables corresponding to each of the factors (see identification details in [Supplementary Table S2](#)). As the resulting factors can be employed to predict discrete clusters of samples, we used all the inferred factors to cluster the patients in the latent factor space, collectively implementing collectively all information from the different analytical platforms.

3) Hierarchical clustering of factors permits to classify patients into metabolotypes

A hierarchical clustering analysis with Ward method and squared Euclidean distance was applied in SPSS statistical software ([Figure 3A](#)). To determine the optimal number of metabolotypes, a discriminant was applied to 2, 3 and 4 metabolotypes. It appears that grouping the samples into three distinct metabolotypes provides the most robust explanation for the observed relationships, where 94% accuracy was observed after cross-validation, demonstrating the existence of 3 clearly differentiated metabolotypes [metabolotype 1 (G1) ($n = 74$), metabolotype 2 (G2) ($n = 10$), metabolotype 3 (G3) ($n = 16$)] ([Figure 3B](#)).

The identified factors enabled metabolotypes to be characterized. The components of the first two factors, F1 (10 sphingolipids, 4 ether-linked phosphatidylcholines, 2 phosphatidylcholines and 3 cholesterol esters, named "*Lipids1*") and F2 (1 di- and 12 triacylglycerols, named "*Lipids2*"), accounted for 44% of the variability and were increased in metabolotype 3 and decreased in metabolotype 1. F3 components (named "*nutritional amino acids*," including eight amino acids) showed increased levels in metabolotype 2. F4 elements (including nine circulating free fatty acids, named "*Lipids3*") showed increased levels in metabolotype 1. F5 elements (including 10 phosphatidylcholines, named "*Lipids4*") and the components in F6 (glutamic acid, choline, aspartic acid, glutamine, and arginine, named "*signaling amino acids*") showed an accumulated variation of 75% and were increased in metabolotype 3.

Univariate statistics were performed on each of the matrices. Each variable was analyzed by ANOVA or its corresponding non-parametric method (Kruskal-Wallis). The overall results show the greatest differences between metabolotype 1 and metabolotype 3

TABLE 3 Metabolites included in each of the six final factors obtained with their factor scores associated with the factors. Saturations above 0.5 are indicated in dark red. Confidence level in annotation based on Metabolomics Society (Blaženović et al., 2018).

Factor	Identification	Confidence level	Factor score					
			Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
1	SM (d41:2)	2	0.83	-0.08	-0.06	0.03	0.19	-0.09
1	SM (d42:3)	2	0.85	0	-0.09	0.05	0.03	-0.08
1	SM (d40:1)	2	0.83	0.09	-0.07	0.11	0.1	0.07
1	SM (d39:1)	2	0.72	0	-0.04	0.13	0.24	-0.04
1	SM (d32:1)	2	0.77	0.07	-0.07	0.02	0.3	0.04
1	SM (d36:2)	2	0.82	0.22	-0.1	0.15	0	-0.04
1	SM (d34:2)	2	0.86	0.1	-0.06	-0.03	0.17	0.03
1	SM (d34:0)	2	0.9	0.09	-0.02	0.11	0.1	-0.03
1	SM (d40:2)	2	0.87	0.03	-0.08	-0.02	0.24	-0.01
1	SM (d38:1)	3	0.88	0.16	-0.12	0.1	0.16	-0.01
1	PC (16:0/16:0)	3	0.78	0.26	-0.03	0.04	0.44	-0.06
1	PC (O-34:1)	3	0.82	-0.08	0.01	0.03	0.2	-0.06
1	PC (O-40:4)	3	0.82	-0.03	-0.13	0.2	0.06	-0.06
1	PC (O-32:0)	3	0.8	0.04	-0.01	0.01	0.17	-0.03
1	PC (O-38:4)	3	0.81	-0.02	-0.04	0.04	0.17	0.01
1	PC (O-36:5)	3	0.75	-0.06	0.01	0.01	0.26	0.01
1	PC (16:0/18:2)	3	0.7	0.28	0.11	-0.14	0.28	0
1	CE (18:2)	2	0.86	0.11	0.05	-0.1	0.18	0.06
1	CE (20:4)	2	0.83	0.05	-0.12	0.06	0.15	0.03
1	CE (18:1)	2	0.83	0.23	-0.03	0.01	0.3	0.02
2	DG (36:4)	3	0.13	0.81	0.11	0.06	0.23	0.25
2	TG (16:0_18:0_18:1)	3	0.11	0.77	0.05	-0.1	0.44	0.09
2	TG (56:6)	3	0.36	0.85	-0.07	0.02	0.14	-0.08
2	TG (18:1_18:2_20:4)	3	-0.16	0.74	0.07	0.03	0	-0.07
2	TG (56:3)	3	-0.08	0.83	0.05	-0.04	0.28	0.06
2	TG (58:5)	3	0.05	0.83	0.02	-0.11	0.29	-0.06
2	TG (56:2)	3	0.11	0.76	0.02	-0.12	0.3	0.06
2	TG (54:3)	3	0.21	0.87	0.01	-0.07	0.08	0.05
2	TG (18:1_18:2_18:2)	3	0.05	0.8	-0.02	-0.15	0.01	-0.04
2	TG (57:2)	3	0.08	0.85	0.03	-0.09	0.31	-0.02
2	TG (53:3)	3	-0.06	0.84	0.08	-0.02	0.38	0
2	TG (54:4)	3	0.09	0.89	0.01	-0.11	-0.03	0
2	TG (16:0_18:1_18:2)	3	0.19	0.94	-0.02	-0.07	0.13	0
3	Phenylalanine	2	-0.08	0.02	0.84	0.03	0.05	0.33
3	Oxalic acid	2	-0.17	-0.08	0.64	-0.1	-0.05	-0.24
3	Myo-Inositol	2	0.03	0.02	0.65	0.25	0	0.19
3	Cholesterol	2	0.14	0.15	0.72	0.04	0.06	-0.19

(Continued on following page)

TABLE 3 (Continued) Metabolites included in each of the six final factors obtained with their factor scores associated with the factors. Saturations above 0.5 are indicated in dark red. Confidence level in annotation based on Metabolomics Society (Blaženović et al., 2018).

Factor	Identification	Confidence level	Factor score					
			Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
3	Proline	2	-0.05	0.18	0.78	-0.12	0.14	0.17
3	Serine	2	-0.02	-0.09	0.93	-0.03	-0.13	0.16
3	Glycine	2	0	-0.14	0.77	0.05	-0.02	0.15
3	Alanine	2	-0.06	0.14	0.84	-0.13	0.13	0.11
3	Methionine	2	-0.09	0.01	0.84	-0.14	0.03	0.11
3	5-Oxoproline/Pyroglutamic acid	2	-0.12	-0.03	0.81	-0.06	-0.02	-0.08
3	Valine	2	0	0.15	0.81	-0.01	-0.02	0.05
3	Threonine	2	-0.03	-0.03	0.85	-0.21	-0.05	-0.01
4	FA (20:4)	2	0.09	-0.1	-0.03	0.73	-0.08	0.45
4	FA (20:3)	2	-0.04	0.01	-0.06	0.89	0.04	0.2
4	FA (17:0)	2	0.04	-0.14	0.05	0.86	-0.05	0.1
4	FA (22:4)	2	0.07	0	-0.15	0.86	-0.08	0.06
4	FA (18:3)	2	0.04	0.02	0	0.88	0.11	0.03
4	FA (18:0;O6)	4	0.04	-0.03	-0.04	0.92	-0.21	0.03
4	FA (14:0)	2	0.07	-0.03	0	0.87	0.07	0.02
4	FA (22:5)	2	0.16	-0.08	-0.08	0.91	0.01	-0.01
4	FA (22:6)	2	0.2	-0.09	0.04	0.64	-0.05	0
4	FA (14:1)	2	0.04	-0.11	-0.03	0.88	0.01	0
4	FAHFA (2:0_20:4)	4	-0.06	-0.04	0.02	0.88	-0.22	0.05
5	LPC (20:3/0:0)	2	0.21	0.18	0.04	-0.02	0.67	0.16
5	PC (18:0_20:3)	3	0.37	0.33	0.03	-0.06	0.73	0.1
5	PC (30:0)	3	0.32	0.21	0.05	-0.06	0.76	-0.09
5	PC (16:0_16:1)	3	0.36	0.3	0.01	0.02	0.78	-0.07
5	PC (34:3)	3	0.39	0.27	0.03	-0.15	0.79	-0.06
5	PC (40:5)	3	0.51	0.28	0.1	-0.02	0.69	-0.06
5	PC (33:1)	3	0.42	0.22	0.03	-0.04	0.68	-0.06
5	PC (18:0_18:1)	3	0.44	0.21	0.04	-0.07	0.71	0.06
5	PC (18:0_22:4)	3	0.3	0.28	-0.01	-0.02	0.72	0.03
5	PC (38:1)	3	0.52	0.18	-0.04	-0.11	0.65	0.01
6	Glutamic acid	2	-0.04	0.12	0.14	0.18	-0.09	0.85
6	Choline	2	0.01	0.13	0.17	0.03	0.07	0.74
6	Aspartic acid	2	0.02	-0.03	-0.01	0.19	0.04	0.72
6	Glutamine	2	-0.16	-0.1	0.18	0.07	-0.12	0.67
6	Arginine	2	0.01	0.02	0.12	0.02	0.14	0.66

Bold values means the factor score of each variable in its factor.

(208 metabolites with p -Bonferroni < 0.05 out of a total of 964 variables) mainly in triglyceride, diglyceride and phosphatidylcholine levels. In addition, there are also differences (125 metabolites) between metabotype 2 and metabotype 3. Only one of these 125 significant metabolites is different from the comparison between metabotype 1 and

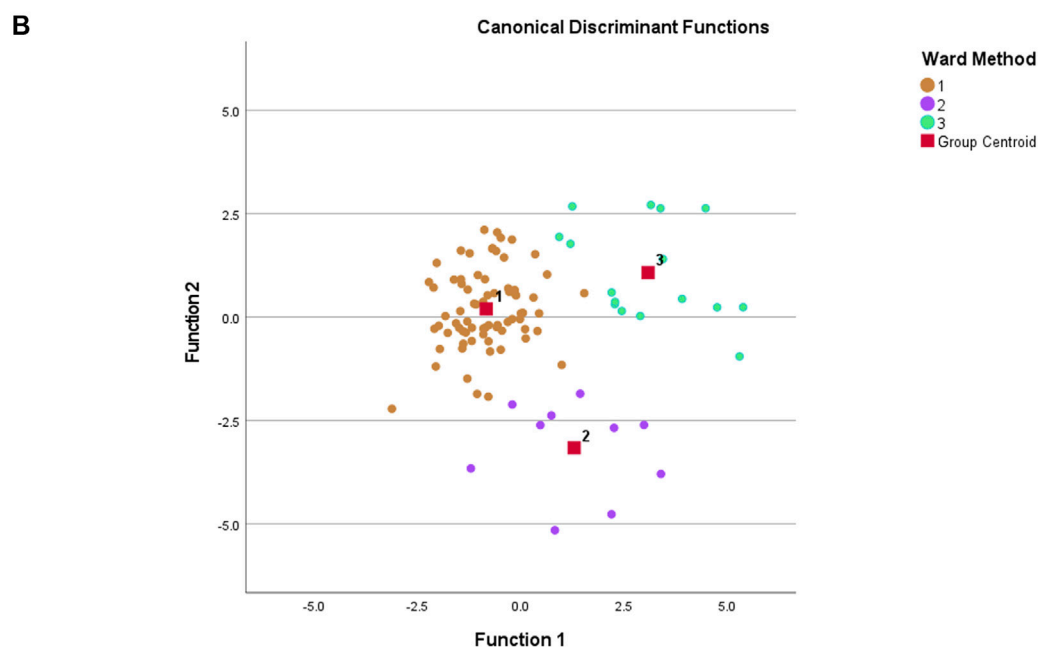
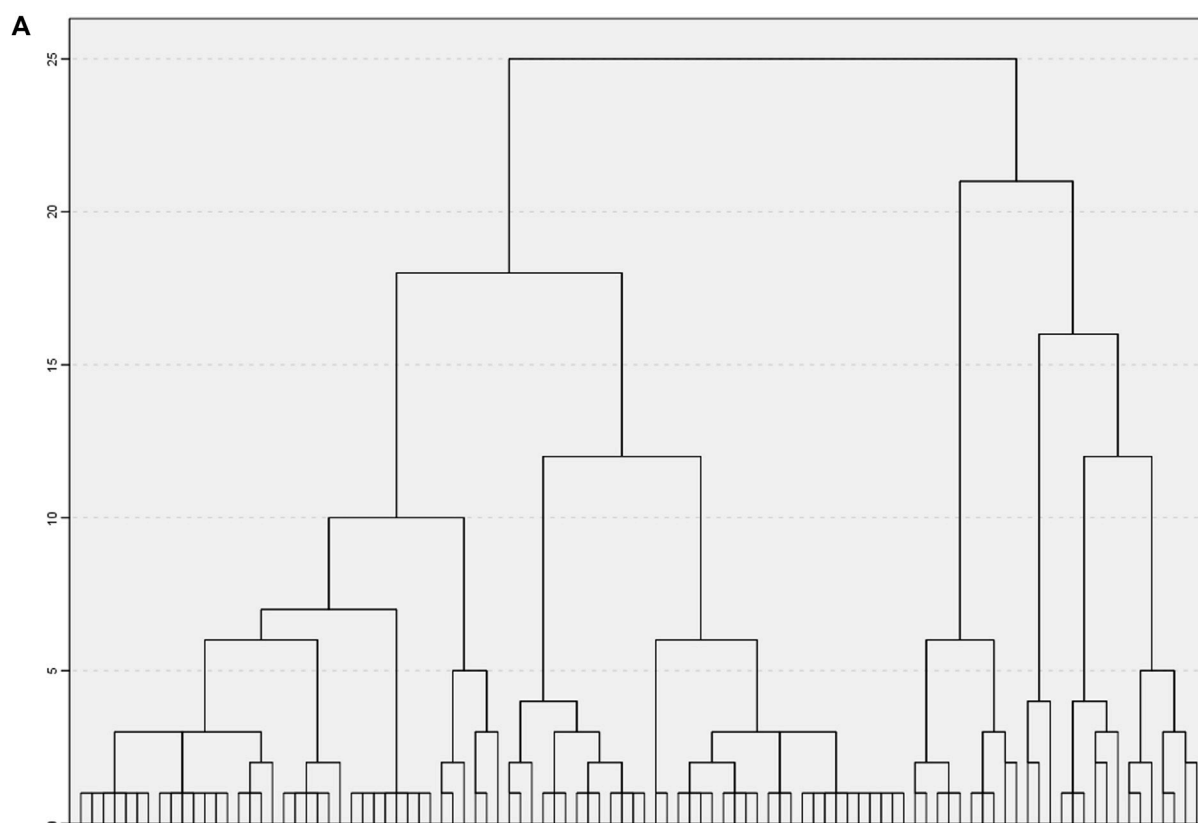


FIGURE 3
(A) Hierarchical clustering performed on the factors obtained after Factor Analysis. Ward’s method and Euclidean distance squared. **(B)** Graph of individuals on the discriminant dimensions. Shows the relative location of the different groups.

metabotype 3, this was lactic acid. However, differences between metabotype 1 and metabotype 2 are minimal (7 metabolites). It is important to note that the levels of different triglycerides in

metabotype 3 are found to be increased over two-fold over those in metabolotypes 1 and 2. In addition, we observed a significant reduction of proline in metabotype 1. Intergroup

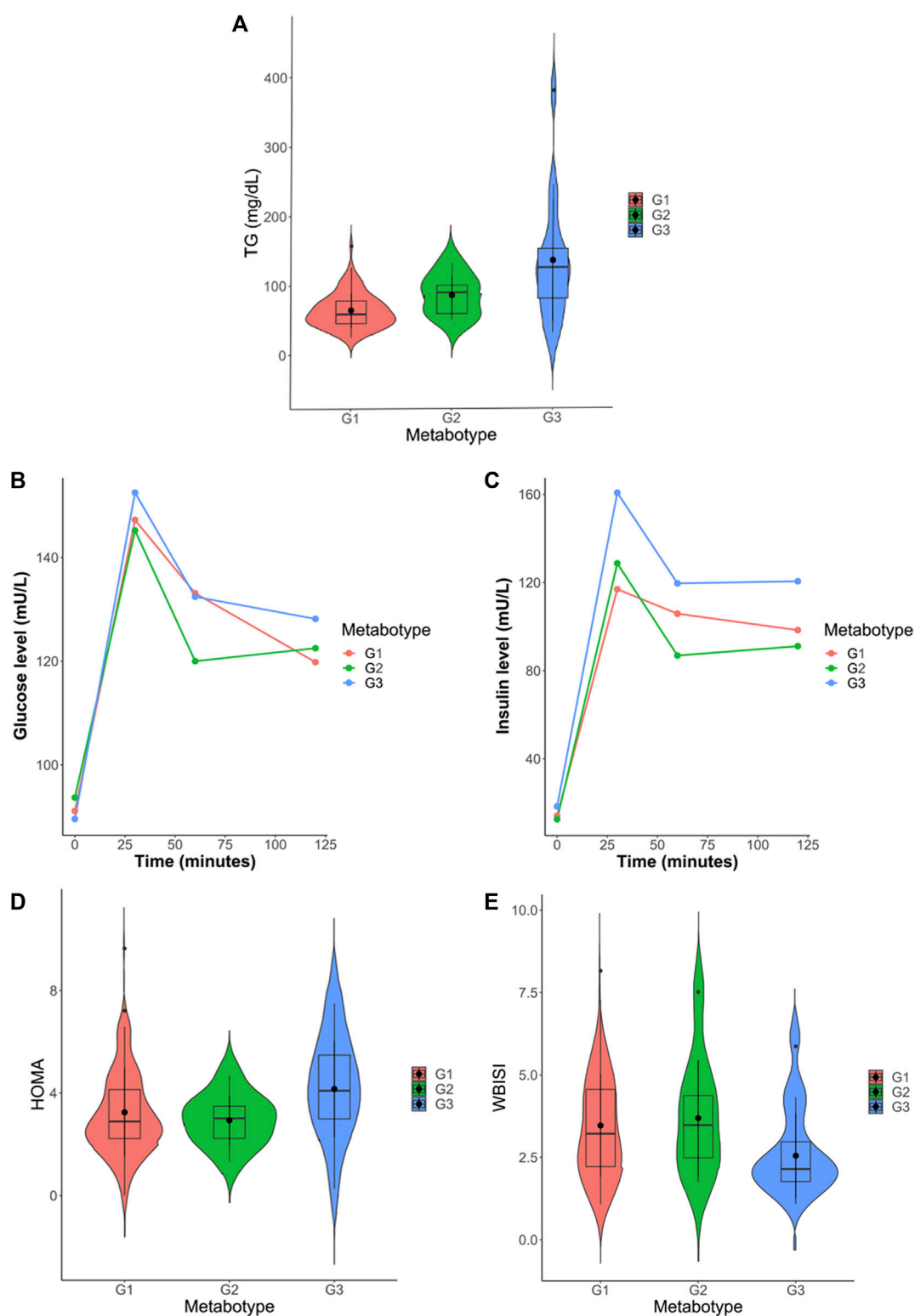


FIGURE 4

(A) Violin plot of total triglyceride levels in the three identified metatypes. (B) Progression plot of insulin levels throughout the oral glucose tolerance test (OGTT). (C) Progression plot of glucose levels throughout the oral glucose tolerance test (OGTT). (D) Violin plot of HOMA-IR (Homeostatic Model Assessment for Insulin Resistance) levels in the three identified metatypes. (E) Violin plot of WBISI (whole-body insulin sensitivity index) levels in the three identified metatypes.

TABLE 4 Clinical/biochemical parameters in the studied metabotypes. Values are average \pm SEM. *p*-value was computed according to the parametric or non-parametric tests applied (ANOVA/Kruskal–Wallis), selected accordingly. Groups homogeneity (Bonferroni) is indicated with superscript letters. Shared letter involves homogeneous groups.

Clinical features	G1	G2	G3	<i>p</i> -value
Age (years)	10.8 \pm 0.4	12 \pm 0.9	11.8 \pm 1	0.34
BMI (kg/m ²)	28.4 \pm 0.5	28 \pm 0.9	29.8 \pm 1.5	0.65
BMI-SDS	4.2 \pm 2.4	3.6 \pm 1.0	3.9 \pm 1.6	0.87
HOMA	3.2 \pm 0.2	2.9 \pm 0.3	4.2 \pm 0.5	0.08
Insulin levels 0 min (μ U/mL)	14.2 \pm 0.8 ^a	12.6 \pm 1.2 ^{ab}	18.4 \pm 1.9 ^b	0.03
Insulin levels 30 min (μ U/mL)	116.9 \pm 8 ^a	128.7 \pm 24.5 ^{ab}	160.7 \pm 17.8 ^b	0.03
Insulin levels 60 min (μ U/mL)	105.8 \pm 9.3	86.9 \pm 24.7	119.6 \pm 14.7	0.17
Insulin levels 120 min (μ U/mL)	98.4 \pm 8.6	91.1 \pm 17.9	120.6 \pm 18.1	0.36
WBISI	3.5 \pm 0.2	3.7 \pm 0.6	2.6 \pm 0.3	0.05
Glucose levels 0 min (mg/dL)	91.1 \pm 0.8	93.7 \pm 1.9	89.6 \pm 1.9	0.32
Glucose levels 30 min (mg/dL)	147.2 \pm 2.9	145.2 \pm 7.7	152.5 \pm 5.9	0.68
Glucose levels 60 min (mg/dL)	133.1 \pm 3.3	120 \pm 5.9	132.4 \pm 5.7	0.35
Glucose levels 120 min (mg/dL)	119.8 \pm 2	122.5 \pm 5.1	128.1 \pm 5	0.24
Total Cholesterol (mg/dL)	154.2 \pm 3.7 ^a	160.9 \pm 8.9 ^{ab}	178.9 \pm 9.9 ^b	0.03
TG (mg/dL)	64.2 \pm 2.9 ^a	85.5 \pm 8.9 ^{ab}	135.3 \pm 22 ^b	0.00
HDL-c (mg/dL)	48 \pm 1.8	45.3 \pm 3.4	41.4 \pm 2.2	0.19
LDL-c (mg/dL)	92.7 \pm 3.2	99.5 \pm 7	110.5 \pm 7.6	0.19
GGT (U/L)	14 \pm 0.5	15 \pm 1.6	15.9 \pm 1.3	0.35
GOT (U/L)	27.4 \pm 0.8	26.9 \pm 1.4	30.1 \pm 2.7	0.87
GPT (U/L)	21.2 \pm 0.9	23 \pm 3.7	26.9 \pm 3.4	0.13
IGF-1 (ng/mL)	320.3 \pm 19.9	289.4 \pm 41.3	293 \pm 49.5	0.64
IGFBP3 (μ g/mL)	4.9 \pm 0.1	4.6 \pm 0.3	5.1 \pm 0.3	0.52
HbA1c (%)	5.4 \pm 0	5.5 \pm 0.1	5.3 \pm 0.1	0.58
Free T4 (ng/dL)	0.9 \pm 0.1	1.0 \pm 0.1	0.9 \pm 0.1	0.61
TSH (μ U/mL)	2.7 \pm 0.2 ^{ab}	1.8 \pm 0.3 ^a	3.3 \pm 0.5 ^b	0.03
DBP mmHg	61.7 \pm 0.9	62.2 \pm 2.1	64.7 \pm 1.9	0.22
SBP mmHg	116.3 \pm 1.5	115.6 \pm 3.7	123.1 \pm 3.5	0.17
Uric Acid mg/dL	5.2 \pm 0.1	5.0 \pm 0.4	5.3 \pm 0.3	0.76
Vitamin D (ng/mL)	24.4 \pm 1.3	23.1 \pm 3.8	20.6 \pm 2.8	0.28

Abbreviations: BMI-SDS, Standardized body mass index (Z-score); HOMA, homeostatic model assessment; WBISI, Whole-body insulin sensitivity index; TG, triglycerides; HDL-c, High density lipoprotein cholesterol; LDL-c, Low density lipoprotein cholesterol; GGT, Gamma-glutamyltransferase; GOT, aspartate transaminase; GPT, glutamic-pyruvic transaminase Alanine aminotransferase; IGF-1, Insulin-like growth factor 1; IGFBP3, Insulin-like Growth Factor-binding Protein 3; HbA1c: hemoglobin A1c; TSH, thyroid-stimulating hormone; DBP, diastolic blood pressure; SBP, diastolic blood pressure.

comparison of routine clinical laboratory data revealed significant differences in total triglyceride levels, along with fasting and glucose-stimulated serum insulin but not glucose level among these metabotypes (Figure 4; Table 4), with individuals in metabotype 3 showing lower insulin sensitivity and hypertriglyceridemia, in a higher risk metabolic profile than patients in metabotypes 1 and 2.

Discussion

In this study, we have highlighted the significance of conducting an in-depth analysis of individuals' metabolic phenotypes, yielding a classification that cannot be attained through anthropometric features or routine clinical laboratory analyses. Furthermore, we have observed an absence of pathognomonic metabolic or

metabolomics signatures due to the presence of specific HetRSVs. In this context, Factor Analysis assumes particular importance in the integration of data from various analytical platforms, bringing us closer to personalized medicine.

The term “personalized medicine” stands for the most suitable specific therapeutic interventions for an individual patient, underscoring the relevance of developing management strategies on specific individuals and not average group response to treatments. This concept has also expanded to nutrition (i.e., personalized nutrition) and current research focuses on the intricate interaction between diet, (epi)genome, and the microbiome, which can determine the effects of bioactive compounds (González-Sarrías et al., 2017).

Using a multiplatform untargeted metabolomics-based approach, we determined the metabolic fingerprint of children with obesity, and by integrating all the data generated by using Factor Analysis to stratify individuals with obesity according to their metabolic phenotype, we defined three different “metabotypes.” Bioinformatics tools are currently available to combine information from different omics technologies or from different analytical platforms. Some of these tools allow the performance of supervised multivariate analysis (Westerhuis et al., 1998; Löfstedt and Trygg, 2011; Boccard and Rutledge, 2013) to determine the existing differences between different groups combining the obtained data. Other integrative multi-omics clustering tools are specific unsupervised integrative methods to find coherent groups between samples or features using the information obtained in a multi-omics analysis (Multiblock PCA, iClusterPlus, iClusterBayes, moCluster, LRAcluster, PINsplus, SNF, etc.) (Mo et al., 2013; Wu et al., 2015; Meng et al., 2016; Nguyen et al., 2017; Wang et al., 2017; Mo et al., 2018; Rappoport and Shamir, 2018; Nguyen et al., 2019; Tanabe et al., 2021; Zhang et al., 2022). However, most of these algorithms require knowledge about the parameters to be applied, and some exhibit complex interpretability. The advantage of Factor Analysis is that it allows us to reduce dimensionality (without losing statistically relevant information), which facilitates the discovery of potential biomarkers, as well as simplifies the biological interpretation of differences between individuals’ metabolic phenotypes. Data integration based on dimensionality reduction approaches seems to be a powerful tool to combine all metabolomic information obtained from different platforms (Zhang et al., 2022). This study proposes the use of Factor Analysis to combine and summarize the information from the different data matrices. The use of Factor Analysis combined with a hierarchical clustering analysis has made it possible to identify three clearly differentiated metabotypes between children with obesity. It is known that cluster analysis has the potential to yield clusters that are either arbitrary or devoid of biological significance. One strength of the results obtained relies on the fact that the acquisition of a notably elevated score in a supervised analysis (discriminant analysis) employing the metabotypes derived from the cluster analysis, serves to not only validate the efficacy of the Factor Analysis but also to enhance the concrete manifestation of the three identified metabotypes.

In routine clinical laboratories, serum levels of triglycerides, lipoproteins, and transaminases are frequently increased in patients with obesity, revealing underlying dyslipidemia and liver dysfunction (Rauschert et al., 2016). Several studies indicate that some amino acids,

such as the branched chain amino acids (BCAA), tyrosine, valine, leucine, or isoleucine, can be used as indicators in early stages of carbohydrate metabolism impairment (Wang et al., 2011; Michaliszyn et al., 2012; McCormack et al., 2013; Butte et al., 2015; Mastrangelo et al., 2016; Suzuki et al., 2019). Moreover, Suzuki et al. reported a correlation between insulin resistance and free amino acid levels in a cohort of patients with moderate to severe obesity (Suzuki et al., 2019). Our results suggest the existence of large metabolic differences between the identified metabotypes, with a singularly differentiated fingerprint in metabotype 3. Factor Analysis indicates that metabotype 3 is characterized by increased levels of “Lipids1,” “Lipids2,” “Lipids4,” and amino acids related to cell signaling. In addition, univariate analysis showed mainly significant differences in triglycerides, diglycerides, and phosphatidylcholines between metabotype 3 and the rest of the metabotypes, with increased levels of these lipid species in metabotype 3. These results suggest the presence of combined hyperlipidemia (cholesterol + triglyceride) in individuals integrated within metabotype 3. Routine clinical laboratory analyses are partially in agreement with these results as individuals in metabotype 3 had increased total triglyceride levels, as well as impairment of insulin, including increased fasting and glucose stimulated insulin secretion and lower WBISI, along with significantly increased levels of isoleucine and proline (UVA, data not shown), in concordance with Suzuki et al. (Suzuki et al., 2019). However, routine clinical analysis of cholesterol species did not detect the higher cholesterol ester levels in metabotype 3 observed by using metabolomics, even when other studies have associated increased cholesterol and triglyceride levels with a decrease in HDL-c levels (Brown et al., 2000) and higher BMI (Huynh et al., 2019). It is pertinent to emphasize that the more pronounced metabolic perturbation of individuals in metabotype 3 is not correlated with a higher BMI-SDS among these individuals compared to those in other metabotypes. Nevertheless, a higher representation of Hispanic ethnicity was observed in metabotype 3 (25%) compared to metabotypes 1 (12%) or 2 (0%). This is consistent with the lower insulin sensitivity and higher triglyceride levels reported in Hispanic children with obesity compared to Caucasians (Martos-Moreno et al., 2020), thus suggesting an eventual ethnic driven influence in obesity associated metabolomic profiles (Butte et al., 2015), although this does not extend to all difference observed (i.e., higher cholesterol ester levels in metabotype 3, not endorsed in inter-ethnic comparisons) (Martos-Moreno et al., 2019; Martos-Moreno et al., 2020). In contrast, proven the role of pubertal status on the development of obesity associated metabolic comorbidities (particularly insulin resistance), we compared the relative frequency within each defined metabotype of patient Tanner stage and, additionally, of prepubertal vs. pubertal patients (pooling TII to T-V in the latter). No significant differences between metabotypes were observed regarding the distribution of Tanner stages within each metabotype (χ^2 0.491; p = 0.782) nor in the relative proportion of prepubertal vs. pubertal patients (χ^2 8.596; p = 0.378). Despite these results not being supportive, the possibility of pubertal influence on the patient metabotype cannot be completely ruled out and this should be further explored in larger patient cohorts.

Interestingly, Factor Analysis splits the relevant amino acids into two subsets. Phenylalanine, proline, serine, glycine, alanine, methionine, valine, and threonine were part of Factor 3, and were higher in Metabotype 2. Glutamic acid, glutamine, and

arginine, together with choline were higher in Metabotype 3. Even though these amino acids have all been shown to correlate with insulin resistance in childhood obesity (Suzuki et al., 2019), such grouping points towards a non-homogeneous involvement of the different amino acids in the complications of obesity. Besides their role in protein synthesis, each amino acid can be involved in different functions and processes, and it is beyond the possibilities of this observational study to determine the exact relationships between the differences found and the therapeutic approach to treat obesity. Those amino acids grouped in Factor 3 include 4 essential (phenylalanine, threonine, valine, methionine) and 3 of the most abundant amino acids (glycine, serine, alanine), and therefore this factor could be strongly related to the nutritional status of the patients, as it would represent protein intake and turnover in the body. In Factor 6, increased in Metabotype 3, glutamic acid, aspartic acid and glutamine, were grouped with choline and arginine. In addition to also reflecting the nutritional status, this group of factors is of particular relevance as they can be related to neurotransmission (Dalangin et al., 2020), and their circulating levels have been proposed as biomarkers of visceral obesity and metabolic alterations (Maltais-Payette et al., 2018) and have been associated with metabolic stress (Yan et al., 2012).

As stated above, childhood obesity is the result of the action of multiple environmental factors on eating and activity habits and lifestyle, in combination with an individual's unique genetic fingerprint. GWAS studies yielded a large list of genes with SNPs, or variants associated to human obesity, but in the vast majority of cases, a single determinant of childhood obesity cannot be identified, thus classifying these cases as "polygenic" or "idiopathic" obesity. In contrast, the rare cases of monogenic forms of obesity, are mainly caused by biallelic mutations in a single gene, usually in the leptin-melanocortin satiety pathway, and are characterized by vary severe, early-onset obesity, usually with evident hyperphagia, and in some cases associated to other metabolic comorbidities and influencing growth pattern even in the first years of life (Handakas et al., 2022). Several metabolomic studies have been performed in childhood obesity, comprehensively characterizing the metabolic alterations in these conditions, as well as in animal models of leptin resistance thus exploring the effect of the impairment of the leptin-POMC satiety pathway (Pietiläinen et al., 2007; Mastrangelo et al., 2016; Martos-Moreno et al., 2017; Rauschert et al., 2017; Kim et al., 2019; Lawler et al., 2020; Rupérez et al., 2020; Sanz-Fernandez et al., 2020). However, the pathogenic role of heterozygous rare sequence variants in the genes of the leptin-melanocortin pathway (Le Collen et al., 2023), as in other genes relevant for central energy and glucose homeostasis is under discussion (Trang and Grant, 2023). Following previous observations by us and other groups (Serra-Juhé et al., 2017; Gerl et al., 2019; Gonzalez-Riano et al., 2021), the hypothesis that they are eventual pathogenicity was proposed. However, the results of this study, showing no anthropometric, metabolic nor metabolomic differences between patients with or without HetRSVs in the studied genes, and the lack of differences in the prevalence of the different metabolotypes between these groups does not verify their pathogenic role (Supplementary Tables), at least from a metabolic and metabolomic point of view. To our knowledge, this study is the first metabolomic study attempting

TABLE 5 Mean values of the factor scores of the new factors obtained after Factor Analysis.

Factor name		G1	G2	G3
		Mean	Mean	Mean
Factor 1	"Lipids1" (SM/CE)	-0.06	-0.03	0.32
Factor 2	"Lipids2" (DAG/TAG)	-0.30	0.01	1.38
Factor 3	"Nutritional amino acids"	-0.33	2.10	0.20
Factor 4	"Lipids3" (FFA)	0.05	-0.13	-0.16
Factor 5	"Lipids4" (PC)	-0.15	-0.29	0.88
Factor 6	"Signaling amino acids"	-0.08	-0.59	0.75

to identify a specific metabolic phenotype associated with the presence of HetRSVs in the leptin-POMC pathway, as well as in glutamate receptors, and demonstrates the absence of a clear differential metabolic phenotype due to the presence of these variants.

The association of obesity with lipidomes and the use of technologies to stratify obesity based on lipidomic data has been previously investigated by means of machine learning algorithms (Gerl et al., 2019). However, this is the first study, to combine metabolomic data from different analytical platforms and genetic data to stratify obesity. Moreover, Factor Analysis has not been previously employed for the subclassification of patients with obesity by using the adequate combination of multiplatform MS metabolomics data. We distinguished two antagonistic metabolotypes (1 and 2 vs. 3) that can be deduced from the examination of the contributing factors. Such subclassification was not possible from the information derived from the routine clinical examination and laboratory analyses. With our approach, subtle but clear differences arose between the three metabolotypes: Six groups of metabolites can be combined to evaluate the metabolic phenotype, and promising associations between this metabolic phenotype and insulin sensitivity, circulating triglycerides and TSH levels and ethnicity have been uncovered. Metabolotypes 1 and 2 have lower levels of the factors corresponding to "Lipids1" (F1), "Lipids2" (F2), "Lipids4" (F5) and "Signaling amino acids" (F6), as compared to Metabotype 3, suggesting a higher metabolic risk phenotype in patients with childhood obesity in Metabotype 3 (Table 5). Among the most promising results, the separation of amino acids into two different factors, and the differential association of these factors with different phenotypes opens the possibility of treating the obese subjects in these two metabolotypes with different approaches. Metabotype 2 was associated with higher levels of F2, or *Nutritional amino acids*, than in the other two metabolotypes, suggesting a healthier metabolic phenotype of these patients, that could speculatively be associated to higher protein intake in their diet, whereas F6, higher in metabolotype 3, could speculatively be associated to a behavioral component of the children in this group. However, the lack of precise control of feeding behavior in these patients, feeding due to the ambulatory modality of management is a limitation to test this hypothesis. Apart from this, the limited number of patients studied, along with the potential confounding factors (such as sex, race or

pubertal status) potentially influencing the described metabolotypes raise the need of validating the presented results in larger and independent cohorts, to enhance the reliability and generalizability of the results, i.e., to support the metabolotypes here identified and to explore an eventual role of these factors.

The challenge for the near future will be to use new technological advances such as that used here to accurately stratify the state/stage of different diseases, in order to precisely predict disease progression and to provide appropriate treatment for each patient, as well as to monitor their evolution.

Data availability statement

This study is available at the NIH Common Fund's National Metabolomics Data Repository (NMDR) website, the Metabolomics Workbench, <https://www.metabolomicsworkbench.org> where it has been assigned Study ID ST002993. The data can be accessed directly via its Project DOI: <http://dx.doi.org/10.21228/M8WX4S>. This work is supported by NIH grant U2C-DK119886. The study is available for review at <http://dev.metabolomicsworkbench.org:2222/data/DRCCMetadata.php?Mode=Study&StudyID=ST002993&Access=ZeeZ9799>.

Ethics statement

The studies involving humans were approved by University Hospital Niño Jesús. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

DC-S: Conceptualization, Data curation, Formal Analysis, Investigation, Methodology, Software, Visualization, Writing—original draft, Writing—review and editing. FRP: Data curation, Formal Analysis, Writing—review and editing. JA: Conceptualization, Funding acquisition, Resources, Supervision, Writing—review and editing. CB: Conceptualization, Resources, Supervision, Writing—review and editing. GM-M: Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Writing—review and editing. FR: Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing—original draft, Writing—review and editing.

References

- Acal, C., Aguilera, A. M., and Escabias, M. (2020). New modeling approaches based on varimax rotation of functional principal components. *Mathematics* 8 (11), 2085. doi:10.3390/math8112085
- Argelaguet, R., Arnol, D., Bredikhin, D., Deloro, Y., Velten, B., Marioni, J. C., et al. (2020). MOFA+: a statistical framework for comprehensive integration of multi-modal single-cell data. *Genome Biol.* 21 (1), 111. doi:10.1186/s13059-020-02015-1
- Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., et al. (2018). Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets. *Mol. Syst. Biol.* 14 (6), e8124. doi:10.15252/msb.20178124
- Armitage, E. G., Godzien, J., Alonso-Herranz, V., López-González, Á., and Barbas, C. (2015). Missing value imputation strategies for metabolomics data. *Electrophoresis* 36 (24), 3050–3060. doi:10.1002/elps.201500352
- Babakus, E., Ferguson, C. E., and Jöreskog, K. G. (1987). The sensitivity of confirmatory maximum likelihood factor analysis to violations of measurement scale and distributional assumptions. *J. Mark. Res.* 24 (2), 222–228. doi:10.2307/3151512
- Bell, C. G., Walley, A. J., and Froguel, P. (2005). The genetics of human obesity. *Nat. Rev. Genet.* 6 (3), 221–234. doi:10.1038/nrg1556

Funding

The authors declare financial support was received for the research, authorship, and/or publication of this article. This study was supported by the Ministry of Science and Innovation of Spain (MICIN), PID2021-122490NB-I00/AEI/10.13039/501100011033 by ERDF-“A way of making Europe”; the Autonomous Community of Madrid P2022/BMD-7232 (TomoXliver2-CM); Instituto de Salud Carlos III (Spain) FIS PI09/91060, FIS PI10/00747, PI13/02195, FIS PI16/00485, FIS PI 19/00166 and FIS PI 22/01820. GM-M and JA are part of the CIBER Fisiopatología de la Obesidad y Nutrición (CB06/03), supported by Instituto de Salud Carlos III. DC-S is supported by a fellowship from CEU International Doctoral School (CEINDO) and Bank Santander.

Acknowledgments

The authors wish to thank Prof. Luis A. Pérez-Jurado for his collaboration in the classification and CADD scoring of the genetic variants harbored by the patients and Julie Chowen for the critical review and English edition of the manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2023.1301996/full#supplementary-material>

- Berger, N. A. (2018). Young adult cancer: influence of the obesity pandemic. *Obesity* 26 (4), 641–650. doi:10.1002/oby.22137
- Blaženović, I., Kind, T., Ji, J., and Fiehn, O. (2018). Software tools and approaches for compound identification of LC-MS/MS data in metabolomics. *Metabolites*, 8 (2), 31. doi:10.3390/metabo8020031
- Boccard, J., and Rutledge, D. N. (2013). A consensus orthogonal partial least squares discriminant analysis (OPLS-DA) strategy for multiblock Omics data fusion. *Anal. Chim. Acta* 769, 30–39. doi:10.1016/j.aca.2013.01.022
- Brown, C. D., Higgins, M., Donato, K. A., Rohde, F. C., Garrison, R., Obarzanek, E., et al. (2000). Body mass index and the prevalence of hypertension and dyslipidemia. *Obes. Res.* 8 (9), 605–619. doi:10.1038/oby.2000.79
- Butte, N. F., Liu, Y., Zakeri, I. F., Mohnhey, R. P., Mehta, N., Voruganti, V. S., et al. (2015). Global metabolomic profiling targeting childhood obesity in the Hispanic population. *Am. J. Clin. Nutr.* 102 (2), 256–267. doi:10.3945/ajcn.115.111872
- Cardel, M. I., Atkinson, M. A., Taveras, E. M., Holm, J. C., and Kelly, A. S. (2020). Obesity treatment among adolescents: a review of current evidence and future directions. *JAMA Pediatr.* 174 (6), 609–617. doi:10.1001/jamapediatrics.2020.0085
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivar. Behav. Res.* 1 (2), 245–276. doi:10.1207/s15327906mbr0102_10
- Chan, J. Y. L., Leow, S. M. H., Bea, K. T., Cheng, W. K., Phoong, S. W., Hong, Z. W., et al. (2022). Mitigating the multicollinearity problem and its machine learning approach: a review. *Mathematics* 10 (8), 1283. doi:10.3390/math10081283
- Chiurazzi, M., Cozzolino, M., Orsini, R. C., Di Maro, M., Di Minno, M. N. D., and Colantuoni, A. (2020). Impact of genetic variations and epigenetic mechanisms on the risk of obesity. *Int. J. Mol. Sci.* 21 (23), 9035. doi:10.3390/ijms21239035
- Choi, J. E., Lee, H. A., Park, S. W., Lee, J. W., Lee, J. H., Park, H., et al. (2023). Increase of prevalence of obesity and metabolic syndrome in children and adolescents in Korea during the COVID-19 pandemic: a cross-sectional study using the knhanes. *Children* 10 (7), 1105. doi:10.3390/children10071105
- Clark, C., Dayon, L., Masoodi, M., Bowman, G. L., and Popp, J. (2021). An integrative multi-omics approach reveals new central nervous system pathway alterations in Alzheimer's disease. *Alzheimers Res. Ther.* 13 (1), 71. doi:10.1186/s13195-021-00814-7
- Cole, T. J., Bellizzi, M. C., Flegal, K. M., and Dietz, W. H. (2000). Establishing a standard definition for child overweight and obesity world-wide: International survey. *BMJ Evid. Based Med.* 320, 1240–1243. doi:10.1136/bmj.320.7244.1240
- Cote, A. T., Harris, K. C., Panagiotopoulos, C., Sandor, G. G. S., and Devlin, A. M. (2013). Childhood obesity and cardiovascular dysfunction. *J. Am. Coll. Cardiol.* 62 (15), 1309–1319. doi:10.1016/j.jacc.2013.07.042
- da Fonseca, A. C. P., Mastrorandi, C., Johar, A., Arcos-Burgos, M., and Paz-Filho, G. (2017). Genetics of non-syndromic childhood obesity and the use of high-throughput DNA sequencing technologies. *J. Diabetes Complicat.* 31 (10), 1549–1561. doi:10.1016/j.jdiacomp.2017.04.026
- Dalagin, R., Kim, A., and Campbell, R. E. (2020). The role of amino acids in neurotransmission and fluorescent tools for their detection. *Int. J. Mol. Sci.* 21 (17), 6197. doi:10.3390/ijms21176197
- Dudzik, D., Barbas-Bernardos, C., García, A., and Barbas, C. (2018). Quality assurance procedures for mass spectrometry untargeted metabolomics: a review. *J. Pharm. Biomed. Anal.* 147, 149–173. doi:10.1016/j.jpba.2017.07.044
- Fairbrother, U., Kidd, E., Malagamuwa, T., and Walley, A. (2018). Genetics of severe obesity. *Curr. Diab Rep.* 18 (10), 85. doi:10.1007/s11892-018-1053-x
- Fernández, J. R., Redden, D. T., Pietrobelli, A., and Allison, D. B. (2004). Waist circumference percentiles in nationally representative samples of African-American, European-American, and Mexican-American children and adolescents. *J. Pediatr.* 145, 439–444. doi:10.1016/j.jpeds.2004.06.044
- Fuente-Martín, E., García-Cáceres, C., Argente-Arizona, P., Díaz, F., Granado, M., Freire-Regatillo, A., et al. (2016). Ghrelin regulates glucose and glutamate transporters in hypothalamic astrocytes. *Sci. Rep.* 6, 23673. doi:10.1038/srep23673
- García, A., and Barbas, C. (2011). Gas chromatography-mass spectrometry (GC-MS)-based metabolomics. *Methods Mol. Biol.* 708, 191–204. doi:10.1007/978-1-61737-985-7_11
- Gavaghan, C. L., Holmes, E., Lenz, E., Wilson, I. D., and Nicholson, J. K. (2000). An NMR-based metabolomic approach to investigate the biochemical consequences of genetic strain differences: application to the C57BL/10J and Alpk:ApfCD mouse. *FEBS Lett.* 484 (3), 169–174. doi:10.1016/s0014-5793(00)02147-5
- Gerl, M. J., Klose, C., Surma, M. A., Fernandez, C., Melander, O., Männistö, S., et al. (2019). Machine learning of human plasma lipidomes for obesity estimation in a large population cohort. *PLoS Biol.* 17 (10), e3000443. doi:10.1371/journal.pbio.3000443
- Gil-de-la-Fuente, A., Godzien, J., Saugar, S., Garcia-Carmona, R., Badran, H., Wishart, D. S., et al. (2019). CEU mass mediator 3.0: a metabolite annotation tool. *J. Proteome Res.* 18 (2), 797–802. doi:10.1021/acs.jproteome.8b00720
- Gonzalez-Riano, C., Gradillas, A., and Barbas, C. (2021). Exploiting the formation of adducts in mobile phases with ammonium fluoride for the enhancement of annotation in liquid chromatography-high resolution mass spectrometry based lipidomics. *J. Chromatogr. Open* 1, 100018. doi:10.1016/j.jcoa.2021.100018
- González-Sarriás, A., García-Villalba, R., Romo-Vaquero, M., Alasalvar, C., Örem, A., Zafilla, P., et al. (2017). Clustering according to urolithin metabotype explains the interindividual variability in the improvement of cardiovascular risk biomarkers in overweight-obese individuals consuming pomegranate: a randomized clinical trial. *Mol. Nutr. Food Res.* 61 (5), 1600830. doi:10.1002/mnfr.201600830
- Handakas, E., Lau, C. H., Alfano, R., Chatzi, V. L., Plusquin, M., Vineis, P., et al. (2022). A systematic review of metabolomic studies of childhood obesity: state of the evidence for metabolic determinants and consequences. *Obes. Rev.* 23 (S1), 23. doi:10.1111/obr.13384
- Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., et al. (2014). Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 158 (4), 929–944. doi:10.1016/j.cell.2014.06.049
- Huynh, K., Barlow, C. K., Jayawardana, K. S., Weir, J. M., Mellett, N. A., Cinel, M., et al. (2019). High-throughput plasma lipidomics: detailed mapping of the associations with cardiometabolic risk factors. *Cell Chem. Biol.* 26 (1), 71–84. doi:10.1016/j.chembiol.2018.10.008
- Jackson, R. S., Creemers, J. W., Ohagi, S., Raffin-Sanson, M. L., Sanders, L., Montague, C. T., et al. (1997). Obesity and impaired prohormone processing associated with mutations in the human prohormone convertase 1 gene. *Nat. Genet.* 16 (3), 303–306. doi:10.1038/ng0797-303
- Kamleh, M. A., McLeod, O., Checa, A., Baldassarre, D., Veglia, F., Gertow, K., et al. (2018). Increased levels of circulating fatty acids are associated with protective effects against future cardiovascular events in nondiabetics. *J. Proteome Res.* 17 (2), 870–878. doi:10.1021/acs.jproteome.7b00671
- Kim, M. J., Kim, J. H., Kim, M. S., Yang, H. J., Lee, M., and Kwon, D. Y. (2019). Metabolomics associated with genome-wide association study related to the basal metabolic rate in overweight/obese Korean women. *J. Med. Food* 22 (5), 499–507. doi:10.1089/jmf.2018.4310
- Koelmel, J. P., Li, X., Stow, S. M., Sartain, M. J., Murali, A., Kemperman, R., et al. (2020). Lipid annotator: towards accurate annotation in non-targeted liquid chromatography high-resolution tandem mass spectrometry (LC-HRMS/MS) lipidomics using a rapid and user-friendly software. *Metabolites* 10 (3), 101. doi:10.3390/metabo10030101
- Kuligowski, J., Sánchez-Illana, Á., Sanjuán-Herráez, D., Vento, M., and Quintás, G. (2015). Intra-batch effect correction in liquid chromatography-mass spectrometry using quality control samples and support vector regression (QC-SVRC). *Analyst* 140 (22), 7810–7817. doi:10.1039/c5an01638j
- Lawler, K., Huang-Doran, I., Sonoyama, T., Collet, T. H., Keogh, J. M., Henning, E., et al. (2020). Leptin-mediated changes in the human metabolome. *J. Clin. Endocrinol. Metab.* 105 (8), 2541–2552. doi:10.1210/clinem/daa251
- Le Collen, L., Delemer, B., Poitou, C., Vaxillaire, M., Toussaint, B., Dechaume, A., et al. (2023). Heterozygous pathogenic variants in POMC are not responsible for monogenic obesity: implication for MC4R agonist use. *Genet. Med.* 25 (7), 100857. doi:10.1016/j.gim.2023.100857
- Lee, C. F., Chen, H. Y., and Lee, J. (2019). “Multivariate analysis: discriminant analysis and factor analysis,” in *Financial econometrics, mathematics and statistics* (New York, NY: Springer), 439–457.
- Löfstedt, T., and Trygg, J. (2011). OnPLS—a novel multiblock method for the modelling of predictive and orthogonal variation. *J. Chemom.* 25 (8), 441–455. doi:10.1002/cem.1388
- Maltais-Payette, I., Boulet, M. M., Prehn, C., Adamski, J., and Tcherno, A. (2018). Circulating glutamate concentration as a biomarker of visceral obesity and associated metabolic alterations. *Nutr. Metab.* 15 (1), 78. doi:10.1186/s12986-018-0316-5
- Marabita, F., James, T., Karhu, A., Virtanen, H., Kettunen, K., Stenlund, H., et al. (2022). Multiomics and digital monitoring during lifestyle changes reveal independent dimensions of human biology and health. *Cell Syst.* 13 (3), 241–255.e7. doi:10.1016/j.cels.2021.11.001
- Martos-Moreno, G., Martínez-Villanueva, J., González-Leal, R., Chowen, J. A., and Argente, J. (2019). Sex, puberty, and ethnicity have a strong influence on growth and metabolic comorbidities in children and adolescents with obesity: report on 1300 patients (the Madrid Cohort). *Pediatr. Obes.* 14 (12), e12565. doi:10.1111/jipo.12565
- Martos-Moreno, G. Á., Martínez-Villanueva, J., González-Leal, R., Barrios, V., Sirvent, S., Hawkins, F., et al. (2020). Ethnicity strongly influences body fat distribution determining serum adipokine profile and metabolic derangement in childhood obesity. *Front. Pediatr.* 8, 8. doi:10.3389/fped.2020.551103
- Martos-Moreno, G. Á., Martínez-Villanueva Fernández, J., Frías-Herrero, A., Martín-Rivada, A., and Argente, J. (2021). Conservative treatment for childhood and adolescent obesity: real world follow-up profiling and clinical evolution in 1300 patients. *Nutrients* 13 (11), 3847. doi:10.3390/nu13113847
- Martos-Moreno, G. Á., Mastrangelo, A., Barrios, V., García, A., Chowen, J. A., Rupérez, F. J., et al. (2017). Metabolomics allows the discrimination of the pathophysiological relevance of hyperinsulinism in obese prepubertal children. *Int. J. Obes.* 41 (10), 1473–1480. doi:10.1038/ijo.2017.137
- Mastrangelo, A., Martos-Moreno, G. Á., García, A., Barrios, V., Rupérez, F. J., Chowen, J. A., et al. (2016). Insulin resistance in prepubertal obese children correlates with sex-dependent early onset metabolomic alterations. *Int. J. Obes.* 40 (10), 1494–1502. doi:10.1038/ijo.2016.92

- Mccormack, S. E., Shaham, O., Mccarthy, M. A., Deik, A. A., Wang, T. J., Gerszten, R. E., et al. (2013). Circulating branched-chain amino acid concentrations are associated with obesity and future insulin resistance in children and adolescents. *Pediatr. Obes.* 8 (1), 52–61. doi:10.1111/j.2047-6310.2012.00087.x
- Meng, C., Helm, D., Frejno, M., and Kuster, B. (2016). moCluster: identifying joint patterns across multiple omics data sets. *J. Proteome Res.* 15 (3), 755–765. doi:10.1021/acs.jproteome.5b00824
- Michalyszyn, S. F., Sjaarda, L. A., Mihalik, S. J., Lee, S. J., Bacha, F., Chace, D. H., et al. (2012). Metabolomic profiling of amino acids and β -cell function relative to insulin sensitivity in youth. *J. Clin. Endocrinol. Metabolism* 97 (11), 2119–2124. doi:10.1210/jc.2012-2170
- Mo, Q., Shen, R., Guo, C., Vannucci, M., Chan, K. S., and Hilsenbeck, S. G. (2018). A fully Bayesian latent variable model for integrative clustering analysis of multi-type omics data. *Biostatistics* 19 (1), 71–86. doi:10.1093/biostatistics/kxx017
- Mo, Q., Wang, S., Seshan, V. E., Olshen, A. B., Schultz, N., Sander, C., et al. (2013). Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc. Natl. Acad. Sci.* 110 (11), 4245–4250. doi:10.1073/pnas.1208949110
- Naz, S., Calderón, Á. A., García, A., Gallafrio, J., Mestre, R. T., González, E. G., et al. (2015). Unveiling differences between patients with acute coronary syndrome with and without ST elevation through fingerprinting with CE-MS and HILIC-MS targeted analysis. *Electrophoresis* 36 (18), 2303–2313. doi:10.1002/elps.201500169
- Nguyen, H., Shrestha, S., Draghici, S., and Nguyen, T. (2019). PINSPlus: a tool for tumor subtype discovery in integrated genomic data. *Bioinformatics* 35 (16), 2843–2846. doi:10.1093/bioinformatics/bty1049
- Nguyen, T., Tagett, R., Diaz, D., and Draghici, S. (2017). A novel approach for data integration and disease subtyping. *Genome Res.* 27 (12), 2025–2039. doi:10.1101/gr.215129.116
- Ni, Z., Angelidou, G., Lange, M., Hoffmann, R., and Fedorova, M. (2017). LipidHunter identifies phospholipids by high-throughput processing of LC-MS and shotgun lipidomics datasets. *Anal. Chem.* 89 (17), 8800–8807. doi:10.1021/acs.analchem.7b01126
- Palmnäs, M., Brunius, C., Shi, L., Rostgaard-Hansen, A., Torres, N. E., González-Domínguez, R., et al. (2020). Perspective: metabolotyping - A potential personalized nutrition strategy for precision prevention of cardiometabolic disease. *Adv. Nutr.* 11, 524–532. Oxford University Press. doi:10.1093/advances/nmz121
- Park, J., Barahona-Torres, N., Jang, S., Mok, K. Y., Kim, H. J., Han, S., et al. (2022). Multi-omics-based autophagy-related untypical subtypes in patients with cerebral amyloid pathology. *Adv. Sci.* 9 (23), e2201212. doi:10.1002/advs.202201212
- Pellegrino, R. M., Di Veroli, A., Valeri, A., Goracci, L., and Cruciani, G. (2014). LC/MS lipid profiling from human serum: a new method for global lipid extraction. *Anal. Bioanal. Chem.* 406 (30), 7937–7948. doi:10.1007/s00216-014-8255-0
- Pietiläinen, K. H., Sysi-Aho, M., Rissanen, A., Seppänen-Laakso, T., Yki-Järvinen, H., Kaprio, J., et al. (2007). Acquired obesity is associated with changes in the serum lipidomic profile independent of genetic effects—a monozygotic twin study. *PLoS One* 2 (2), e218. doi:10.1371/journal.pone.0000218
- Rappoport, N., and Shamir, R. (2018). Multi-omic and multi-view clustering algorithms: review and cancer benchmark. *Nucleic Acids Res.* 46 (20), 10546–10562. doi:10.1093/nar/gky889
- Rauschert, S., Kirchberg, F. F., Marchioro, L., Koletzko, B., Hellmuth, C., and Uhl, O. (2017). Early programming of obesity throughout the life course: a metabolomics perspective. *Ann. Nutr. Metab.* 70 (3), 201–209. doi:10.1159/000459635
- Rauschert, S., Uhl, O., Koletzko, B., Kirchberg, F., Mori, T. A., Huang, R. C., et al. (2016). Lipidomics reveals associations of phospholipids with obesity and insulin resistance in young adults. *J. Clin. Endocrinol. Metab.* 101 (3), 871–879. doi:10.1210/jc.2015-3525
- Rupérez, F. J., Martos-Moreno, G. Á., Chamoso-Sánchez, D., Barbas, C., and Argente, J. (2020). Insulin resistance in obese children: what can metabolomics and adipokine modelling contribute? *Nutrients* 12 (11), 3310. doi:10.3390/nu12113310
- Sanz-Fernandez, M. V., Torres-Rovira, L., Pesantez-Pacheco, J. L., Vazquez-Gomez, M., Garcia-Contreras, C., Astiz, S., et al. (2020). A cross-sectional study of obesity effects on the metabolomic profile of a leptin-resistant swine model. *Metabolites* 10 (3), 89. doi:10.3390/metabo10030089
- Serra-Juhé, C., Martos-Moreno, G. Á., Bou de Pieri, F., Flores, R., González, J. R., Rodríguez-Santiago, B., et al. (2017). Novel genes involved in severe early-onset obesity revealed by rare copy number and sequence variants. *PLoS Genet.* 13 (5), e1006657. doi:10.1371/journal.pgen.1006657
- Shen, R., Olshen, A. B., and Ladanyi, M. (2009). Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* 25 (22), 2906–2912. doi:10.1093/bioinformatics/btp543
- Sullivan, A., Gibney, M. J., Connor, A. O., Mion, B., Kaluskar, S., Cashman, K. D., et al. (2011). Biochemical and metabolomic phenotyping in the identification of a vitamin D responsive metabolotype for markers of the metabolic syndrome. *Mol. Nutr. Food Res.* 55 (5), 679–690. doi:10.1002/mnfr.201000458
- Suzuki, Y., Kido, J., Matsumoto, S., Shimizu, K., and Nakamura, K. (2019). Associations among amino acid, lipid, and glucose metabolic profiles in childhood obesity. *BMC Pediatr.* 19 (1), 273. doi:10.1186/s12887-019-1647-8
- Tanabe, K., Hayashi, C., Katahira, T., Sasaki, K., and Igami, K. (2021). Multiblock metabolomics: an approach to elucidate whole-body metabolism with multiblock principal component analysis. *Comput. Struct. Biotechnol. J.* 19, 1956–1965. doi:10.1016/j.csbj.2021.04.015
- Tini, G., Marchetti, L., Priami, C., and Scott-Boyer, M. P. (2019). Multi-omics integration—a comparison of unsupervised clustering methodologies. *Brief. Bioinform.* 20 (4), 1269–1279. doi:10.1093/bib/bbx167
- Trang, K., and Grant, S. F. A. (2023). “Genetics and epigenetics in the obesity phenotyping scenario,” in *Reviews in endocrine and metabolic disorders* (Springer).
- Tsugawa, H., Ikeda, K., Takahashi, M., Satoh, A., Mori, Y., Uchino, H., et al. (2020). A lipidome atlas in MS-DIAL 4. *Nat. Biotechnol.* 38 (10), 1159–1163. doi:10.1038/s41587-020-0531-2
- Wahl, S., Yu, Z., Kleber, M., Singmann, P., Holzapfel, C., He, Y., et al. (2012). Childhood obesity is associated with changes in the serum metabolite profile. *Obes. Facts* 5 (5), 660–670. doi:10.1159/000343204
- Waldram, A., Holmes, E., Wang, Y., Rantalainen, M., Wilson, I. D., Tuohy, K. M., et al. (2009). Top-down systems biology modeling of host metabolite-microbiome associations in obese rodents. *J. Proteome Res.* 8 (5), 2361–2375. doi:10.1021/pr8009885
- Wang, B., Zhu, J., Pierson, E., Ramazzotti, D., and Batzoglou, S. (2017). Visualization and analysis of single-cell RNA-seq data by kernel-based similarity learning. *Nat. Methods* 14 (4), 414–416. doi:10.1038/nmeth.4207
- Wang, T. J., Larson, M. G., Vasani, R. S., Cheng, S., Rhee, E. P., McCabe, E., et al. (2011). Metabolite profiles and the risk of developing diabetes. *Nat. Med.* 17 (4), 448–453. doi:10.1038/nm.2307
- Wan Mohd Zin, R. M., Jalaludin, M. Y., Yahya, A., Nur Zati Iwani, A. K., Md Zain, F., Hong, J. Y. H., et al. (2022). Prevalence and clinical characteristics of metabolically healthy obese versus metabolically unhealthy obese school children. *Front. Endocrinol.* 13, 13. doi:10.3389/fendo.2022.971202
- Westerhuis, J. A., Kourti, T., and Macgregor, J. F. (1998). *Analysis of multiblock and hierarchical PCA and pls models*.
- Wu, D., Wang, D., Zhang, M. Q., and Gu, J. (2015). Fast dimension reduction and integrative clustering of multi-omics data using low-rank approximation: application to cancer molecular classification. *BMC Genomics* 16 (1), 1022. doi:10.1186/s12864-015-2223-8
- Yan, J., Winter, L. B., Burns-Whitmore, B., Vermeylen, F., and Caudill, M. A. (2012). Plasma choline metabolites associate with metabolic stress among young overweight men in a genotype-specific manner. *Nutr. Diabetes* 2 (10), e49. doi:10.1038/nutd.2012.23
- Zhang, X., Zhou, Z., Xu, H., and Liu, C. (2022). Integrative clustering methods for multi-omics data. *WIREs Comput. Stat.* 14 (3), e1553. doi:10.1002/wics.1553
- Zhao, C., Mao, J., Ai, J., Shenwu, M., Shi, T., Zhang, D., et al. (2013). Integrated lipidomics and transcriptomic analysis of peripheral blood reveals significantly enriched pathways in type 2 diabetes mellitus. *BMC Med. Genomics* 6 (S1), S12. doi:10.1186/1755-8794-6-S1-S12