



Combining Machine Learning and Metabolomics to Identify Weight Gain Biomarkers

Flávia Luísa Dias-Audibert^{1†}, Luiz Claudio Navarro^{2†}, Diogo Noin de Oliveira¹, Jeany Delafiori¹, Carlos Fernando Odor Rodrigues Melo¹, Tatiane Melina Guerreiro¹, Flávia Troncon Rosa³, Diego Lima Petenuci⁴, Maria Angelica Ehara Watanabe⁴, Lício Augusto Velloso⁵, Anderson Rezende Rocha^{2*} and Rodrigo Ramos Catharino^{1*}

OPEN ACCESS

Edited by:

Pinyi Lu,
Biotechnology HPC Software
Applications Institute (BHSAI),
United States

Reviewed by:

Zheng Wang,
Albert Einstein College of Medicine,
United States
Patrick James Trainor,
New Mexico State University,
United States

*Correspondence:

Anderson Rezende Rocha
anderson.rocha@ic.unicamp.br
Rodrigo Ramos Catharino
rrc@fcm.unicamp.br

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Bioinformatics and Computational
Biology,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

Received: 03 September 2019

Accepted: 06 January 2020

Published: 24 January 2020

Citation:

Dias-Audibert FL, Navarro LC, de
Oliveira DN, Delafiori J, Melo CFOR,
Guerreiro TM, Rosa FT, Petenuci DL,
Watanabe MAE, Velloso LA, Rocha AR
and Catharino RR (2020) Combining
Machine Learning and Metabolomics
to Identify Weight Gain Biomarkers.
Front. Bioeng. Biotechnol. 8:6.
doi: 10.3389/fbioe.2020.00006

¹Innovare Biomarkers Laboratory, School of Pharmaceutical Sciences, University of Campinas, Campinas, Brazil, ²RECOD Laboratory, Institute of Computing (IC), University of Campinas, Campinas, Brazil, ³Centro Universitário Filadélfia, Londrina, Brazil, ⁴Laboratory of Studies and Applications of DNA Polymorphisms, Biological Sciences Center, Londrina State University, Londrina, Brazil, ⁵Department of Internal Medicine, School of Medical Sciences, University of Campinas, Campinas, Brazil

Weight gain is a metabolic disorder that often culminates in the development of obesity and other comorbidities such as diabetes. Obesity is characterized by the development of a chronic, subclinical systemic inflammation, and is regarded as a remarkably important factor that contributes to the development of such comorbidities. Therefore, laboratory methods that allow the identification of subjects at higher risk for severe weight-associated morbidity are of utter importance, considering the health, and safety of populations. This contribution analyzed the plasma of 180 Brazilian individuals, equally divided into a eutrophic control group and case group, to assess the presence of biomarkers related to weight gain, aiming at characterizing the phenotype of this population. Samples were analyzed by mass spectrometry and most discriminant features were determined by a machine learning approach using Random Forest algorithm. Five biomarkers related to the pathogenesis and chronicity of inflammation in weight gain were identified. Two metabolites of arachidonic acid were upregulated in the case group, indicating the presence of inflammation, as well as two other molecules related to dysfunctions in the cycle of nitric oxide (NO) and increase in superoxide production. Finally, a fifth case group marker observed in this study may indicate the trigger for diabetes in overweight and obesity individuals. The use of mass spectrometry combined with machine learning analyses to prospect and characterize biomarkers associated with weight gain will pave the way for elucidating potential therapeutic and prognostic targets.

Keywords: obesity, machine learning, random forest, metabolomics, biomarkers

INTRODUCTION

Weight gain has become a worldwide epidemic, leading to a population of obese individuals of more than 13% of the world population in 2016. According to the World Health Organization, obesity has almost tripled since 1975 and is now linked to a larger number of deaths than underweight alone (World Health Organization, 2018). Obesity is characterized by weight gain, excess body fat and complications in several tissues, and organs. The condition is associated with

type 2 diabetes, hypertension, dyslipidemia and cardiovascular events (Libert et al., 2018). The presence of three or more of these disorders (including elevated fasting glucose and abdominal obesity) define Metabolic Syndrome (MetS), a multifactorial disease that increase substantially the mortality risk (Zhong et al., 2017; Lent-Schochet et al., 2019). According to a MetS prevalence study, 34% of Americans meet the diagnose standards, placing MetS as an important public health issue (Aguilar et al., 2015). Considering the wide variety of possible complications and the different states of metabolic disorders in MetS, it is important to have a better understanding of the metabolic environment of these conditions. In addition, the different profiles found in overweight individuals makes the search for biomarkers - essential for future definitions of diagnosis, prognosis, and treatment (Zhong et al., 2017; Libert et al., 2018).

Weight gain results from a chronic state of positive energy balance due to increased caloric intake and reduced energy expenditure, and may be influenced by a number of genetic and environmental factors (Velloso et al., 2015). Adipose tissue expansion and hypertrophied adipocytes tend to activate intra- and extracellular stress responses, promoting the development of a tissue-specific and systemic proinflammatory state (Sartipy and Loskutoff, 2003; Cooke and Naaz, 2004; Guilherme et al., 2008). This state of metabolic disturbance, activation of immune system and increase in inflammation markers in overweight and obesity is described as a low-grade chronic inflammation, and recent studies have demonstrated some of the pathways associated with this condition. Human and murine research has evinced the relationship between increased nutrient intake, weight gain and inflammatory responses (Gregor and Hotamisligil, 2011; Andersen et al., 2016). Despite that, it is still necessary to unravel the many metabolic and molecular pathways involved in the development of obesity. In this sense, studies in the field of “omics” have provided advance in the physiological and pathological alterations in living organisms. Considering that metabolomics is the study of all metabolites within an organism, including precursors, intermediates and end products, and that mass spectrometry produces a large amount of data for each sample, machine learning techniques provide tools to recognize patterns in this large data environment. In this way, the use of mass spectrometry with machine learning builds an experimental metabolomics platform to develop diagnostics and prevention tools (Kononenko, 2001; Jordan and Mitchell, 2015).

Within this context, this study aims at elucidating molecules that may be involved in the development of obesity in a cohort of overweight and obese individuals and propose the relationship of such species with chronic inflammation. Finally, we understand this study as the basis for potential developments and applications of targeted therapeutics and clinically-relevant prognostic markers.

MATERIALS AND METHODS

Ethics Statement

This study was conducted according to the principles expressed in the Declaration of Helsinki and approved by the Research Ethics Committee of the State University of Londrina, Paraná,

Brazil (No. 2.426.419—No. CAAE: 79277817.8.0000.5231). A written informed consent was obtained from all patients. Samples were obtained from the Health Education Center of Filadelfia University Center—UNIFIL, Londrina, Paraná, Brazil and State University of Londrina, Paraná, Brazil.

Research Participants and Specimen Collection

In total, 180 patients were included in this study, separated into two groups: Case, formed by a combination of obese and overweight individuals, and Control, composed of eutrophic individuals. A questionnaire was used to collect data such as age, gender, and personal and family history of chronic non-communicable diseases. A nutritional assessment was performed including weight and height. The collected specimens from all participants of the present study consisted of blood (plasma) samples.

Body Mass Index (BMI) was calculated as weight in kilograms divided by height in meters squared. Overweight and Obese (case group) classifications were defined based on the respective BMI standard cut points that have been recommended by the World Health Organization (WHO) (World Health Organization, 1995): BMIs of ≥ 25 and < 30 kg/m² for Overweight; BMIs of ≥ 30 and < 34.9 kg/m² for class I obesity ($n = 29$); BMIs of ≥ 35 and < 39.9 kg/m² for class II obesity ($n = 17$); BMIs ≥ 40 kg/m² for class III obesity ($n = 13$).

For the control group, eutrophic individuals that presented a BMI between ≥ 18 and < 24.9 kg/m² were recruited. In order to exclude from the group any individuals with normal BMI, but with fat percentage above normal, an estimation of body fat composition was performed using a bivalent (foot to foot) bioimpedance analysis (Electronic Scale BC533, Tanita, Tokyo, Japan and Electronic Scale W905, Wiso, São José, Brazil). The percentage of fat within the average for age and gender were determined according to Jackson and Pollock (1978) and Jackson et al. (1980) protocols.

MALDI-MS Analysis

For sample preparation, 5 μ L of plasma and 5 μ L of methanol were mixed in a plastic tube. This solution was diluted in 90 μ L of a MALDI matrix α -cyano-4-hydroxycinnamic acid (Sigma-Aldrich, St. Louis, MO) solution at 10 mg/mL, prepared in 1:1 acetonitrile/methanol. Then, 1 μ L of the obtained solution was spotted on a MALDI 96-well plate and air-dried. Each sample was analyzed in quintuplicates. Spectra were acquired using a MALDI LTQ-XL (Thermo Scientific, San Jose, CA) at the mass range of 250–2,000 m/z, in the positive ion mode.

Machine Learning Method

Machine learning has been intensively used in MS imaging (Hanselmann et al., 2009; Rappez et al., 2019), and is becoming a key methodology in untargeted metabolomics (Li et al., 2019). The strategy applied in this article is an extension of the ML-based platform used by our group for screening ZIKV molecules in blood serum (Melo et al., 2018). The method ranks variables using the feature importance measure provided by Random Forest trained trees (Breiman et al., 1984; Breiman, 2001; Gregor

and Hotamisligil, 2011). It also evaluates the ranking results when discarding least important variables iteratively. Using f1-score (first harmonic average on recall and precision) as the cost function (performance measure), it is possible to determine which features on the mass spectra are most discriminant to predict positive patients based on their observations (Melo et al., 2018). A positive patient is defined as any given individual with the condition of interest herein, i.e., weight gain. At the end of the process, the method provides a Random Forest classifier, called diagnosis classifier, that can predict weight gain markers based on the blood serum mass spectrum, as well as which features are the most important on that prediction, determining the starting point for the metabolomics research.

Starting with data preparation, all vectors were normalized dividing intensity values by the maximum one in each vector. Then, we randomly split the patient samples into two partitions randomly selected, the P_{test} partition, corresponding to 20% of the all patients, i.e., data from 18 controls and 18 patients with weight gain. This partition was kept untouched until the end of the process, when it was used to measure the classification accuracy and precision of the classifier, as well as the potential biomarkers identified by the process. The remaining partition P_{fit} consisted of 80% of patients' data points (82 controls and 82 patients with weight gain) and were used to determine the hyperparameters of the machine-learning algorithm, aiming at ultimately training the classifier. For hyperparameters and determination of the most discriminant features, the P_{fit} partition was randomly shuffled into 10 experiments datasets with 80% for training (65 cases and 65 controls) and 20% validation (17 controls and 17 patients with weight gain) in each experiment. It is important to state that vectors (mass spectra measurements) of the same patient were always in the same partition, so as not to contaminate the training and tests results.

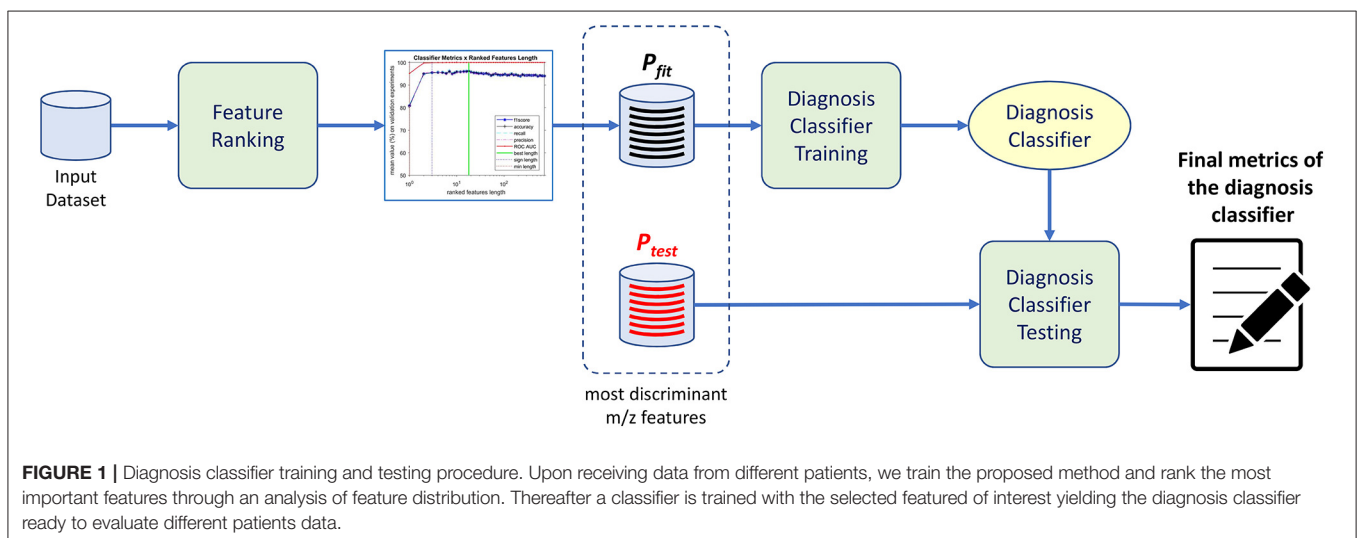
A group of linearly dependent features did not carry more information than a single representative of the group; therefore, in order to enable the following process and improve resource consumption and time in the next steps, linearly dependent features were eliminated using the Gauss-Jordan

elimination algorithm, keeping only features with maximum median across all vectors to represent each group. Thus, the following machine learning steps on most discriminant and putative biomarker features determination rely only on linearly independent features. In the search for molecules based on the determined m/z , dependent features are also used to help on the match of ions for the molecule represented by linearly independent representatives. After the elimination, 713 ions remained represented in the feature vectors from the original 1,752 m/z values.

The purpose of the machine-learning process here devised is 2-fold: first, to generate a diagnosis classifier, which will be used for screening patients to the studied condition (Figure 1); second, to determine potential biomarkers, which will follow to the metabolomics analysis stage, reducing the number of molecules to be analyzed (Figure 2). In the proposed machine learning approach, Random Forest algorithm plays an important role. In addition to be a good choice for a classifier, as shown by Fernández-Delgado et al. (2014), it also provides the variables' importance, which allows them to be ranked based on their influence on the classification results.

Statistical metrics defined in Table 1 are used to evaluate the classifiers' performance on the experiments and final tests. Tp denotes true-positive examples, which are cases that actual positives are classified correctly as positives. Tn stands for true-negative cases, which are examples that actual negatives are classified correctly as negatives. Conversely, Fp denotes false positive cases (those that are actual negatives and were wrongly classified as positives). Finally, Fn represents the false negative examples (those examples that are actual positives, but were incorrectly classified as negatives).

Before starting to determine the most discriminant features, we selected the appropriate hyperparameter *number of trees* in the Random Forest method through a grid-search procedure, seeking to maximize the accuracy and f1-score as function of vector length and number of trees. The feature ranking step measures classification results averaging statistical metrics of the trained classifiers' validation (through the 10 repetitions



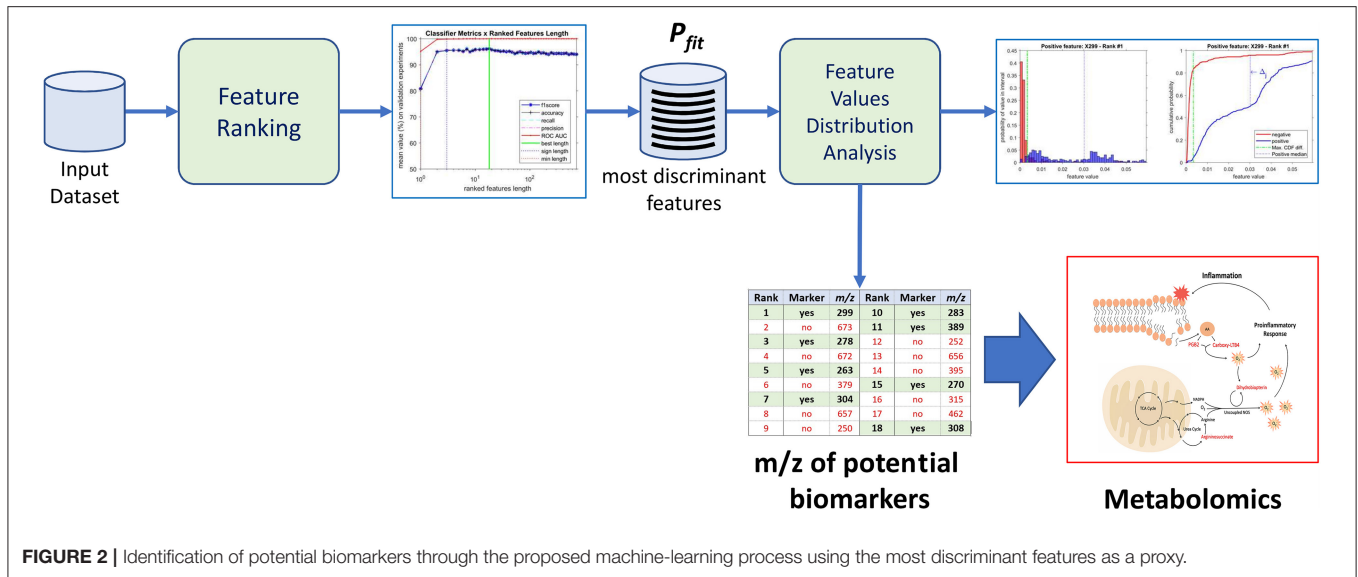


FIGURE 2 | Identification of potential biomarkers through the proposed machine-learning process using the most discriminant features as a proxy.

TABLE 1 | Statistical metrics definition to evaluate classification results.

Metric	Sensitivity (Stv) or Tpr	Specificity (Spc)	Precision (Prc)	F1-score (F1s)	Accuracy (Acc)
Formula	$Tp / (Tp + Fn)$	$Tn / (Tn + Fp)$	$Tp / (Tp + Fp)$	$2 \cdot Prc \cdot Stv / (Prc + Stv)$	$(Rcl + Spc) / 2$

previously described). In each interaction, the metrics on **Table 1** were determined, and features were ranked according to their feature importance. Vector length was reduced by a factor of 0.9 in each interaction, discarding the less significant features until it achieved length 1.

Marker features represents the ions that are potential biomarkers to be further analyzed in metabolomics research. From the 18 most discriminant features, eight were identified as markers by the definition (see equation below) that a marker feature has the difference between the cumulative distribution function (CDF) of the feature values for the negative class and the CDF of positive class vectors, at the median value of positive class, exceeding the threshold β , that was defined as 0.4 (40%) for the analysis herein.

$$F_j \text{ is a marker feature, if: } \Delta_j = (1 - P(m_j)) - (1 - Q(m_j)) = Q(m_j) - P(m_j) > \beta$$

and

$$Q(F_j) \geq P(F_j) \mid \forall F_j > m_j$$

where

- y_j is a F_j value for a positive patient;
- m_j is the median of y_j , i.e., median of F_j values of positive patients;
- \bar{y}_j is a F_j value for a negative patient;
- $p(y_j)$ is the probability distribution function of positive patients, and $q(\bar{y}_j)$ the probability distribution function of negative patients;

$P(y_j)$ is the cumulative distribution function (CDF) of y values, and $Q(\bar{y}_j)$ is the CDF of \bar{y}_j ;

$0 < \beta < 0.5$ CDF is the threshold of the difference over median of the feature j for the positive patients (e.g., $\beta = 40\%$).

To evaluate how the most discriminant features (18 variables) perform in different classifiers, we submitted the same partitions used to train and validate the Random Forest classifier to a SVM classifier with two different optimization algorithms: (SMO) Sequential Minimal Optimization (Fan et al., 2005), and (ISDA) Iterative Single Data Algorithm (Kecman et al., 2005), and to a single decision tree classifier using (GDI) Gini Diversity Index (Breiman et al., 1984) as the split criterion (same criterium used by the Random Forest).

The software platform used for machine learning processing was the MATLAB 2018b version 9.5.0.1067069 (R2018b) February 28, 2019 Update 4, especially the bagged ensemble decision tree algorithm and related functions available in the package Statistics and Machine Learning Toolbox. All experiments ran on a Samsung Notebook NP500R5H with Intel® Core™ i7-5500U CPU @ 2.40 GHz, RAM 8 GB on Windows 10 v1903.

The datasets and codes used to perform these experiments are available in public repository through the link <http://dx.doi.org/10.21227/k446-fp12> (Dias-Audibert et al., 2019).

HRMS Analyses

In order to determine the exact masses of the markers discriminated by machine learning, case group samples were analyzed in an ESI-LTQ-XL Orbitrap Discovery (Thermo Scientific, Bremen, Germany). For sample preparation, following

an adapted protocol by Melo et al. (2017), 20 μL of plasma were diluted in 200 μL of tetrahydrofuran and homogenized under vortex for 30 s; the volume was completed to 1 mL with methanol, with further homogenization, and then centrifuged for 5 min under 3,600 rpm. Hundred microliter of the supernatant was diluted in 400 μL of methanol. Formic acid was added to 0.1% in the final solution. Samples were directly infused into an ESI-LTQ-XL Orbitrap Discovery (Thermo Scientific, Bremen, Germany), and data were acquired according to the following parameters: flow rate at 10 $\mu\text{L}\cdot\text{min}^{-1}$, capillary temperature at 280°C, spray current at 5 kV, and sheath gas at five arbitrary units, in the mass range of 200–2,000 m/z .

Structure Elucidation

Human Metabolome database version 3.6 (www.hmdb.ca), LipidMaps (www.lipidmaps.org), and METLIN (Scripps Center for Metabolomics, La Jolla, CA) were consulted to elect the most suitable marker obtained from HRMS and mass accuracy;

only molecules with a mass shift lower than 2 ppm were considered. Aiming to confirm the proposed molecules, *Tandem* mass spectrometry (MS/MS) data were acquired and structure proposals were carried out with the assistance of METLIN MS/MS databases and fragmentation calculations/ modeling predicted by Mass Frontier software (v. 6.0, Thermo Scientific, San Jose, CA).

RESULTS

Anthropometric data of the study groups are shown as means in **Table 2**. Regarding non-communicable diseases, the control group had no comorbidities, whereas in the 90 patients from case group, 37.5% had hypertension and 17.5% had diabetes. Mass spectral data obtained from 180 patient samples in MALDI-MS measurements fed the machine learning process with 893 (about 5 per patient) feature vectors of length 1,752 (m/z values as features).

TABLE 2 | Anthropometric data of subjects.

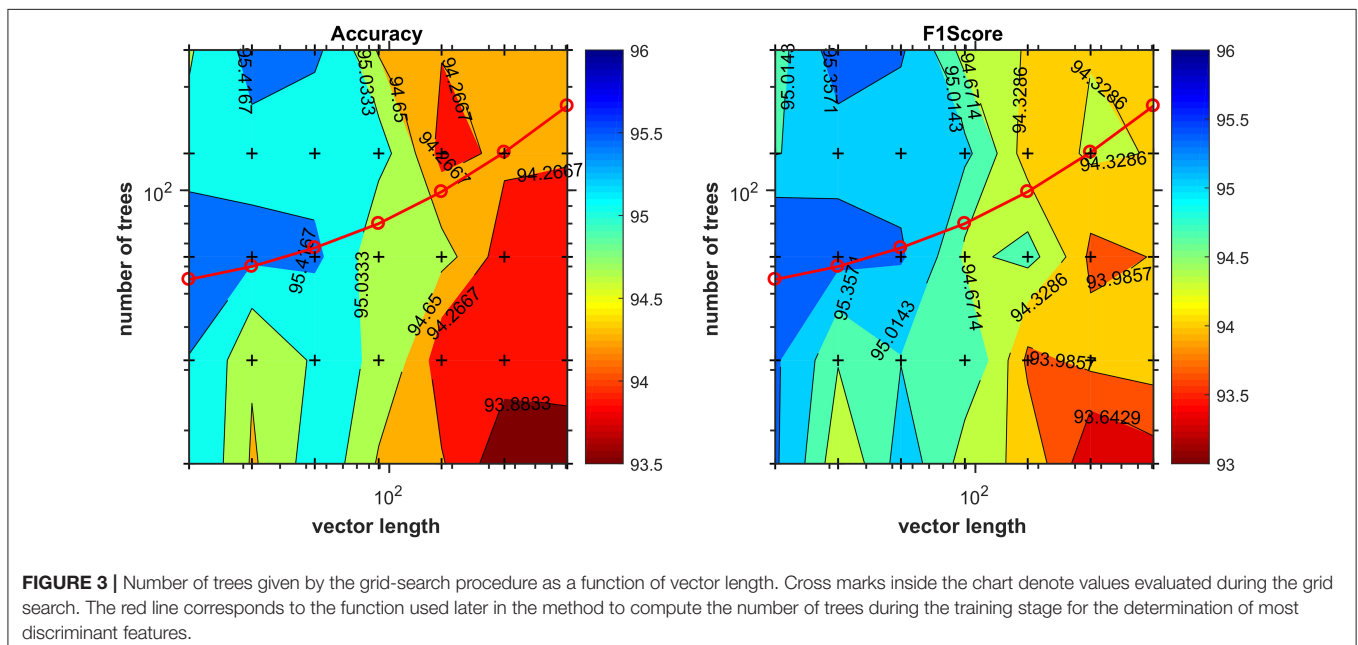
Metric	Group	All		Men		Women	
		Mean	RSD	Mean	RSD	Mean	RSD
No. patients	Control	90		31		59	
	Case	90		18		72	
Age (y)	Control	35.1	12.5	34.5	13.2	35.4	12.2
	Case	39.0	15.1	44.3	17.2	37.7	14.4
Weigh (kg)	Control	64.0	11.1	75.3	8.2	57.8	6.7
	Case	88.6	18.9	103.5	22.5	84.8	16.0
BMI (kg/m ²)	Control	22.3	2.8	24.7	2.3	21.0	2.0
	Case	33.3	6.5	34.4	9.3	33.0	5.7
Body fat percentage (%)	Control	40.2	7.6	33.7	5.9%	41.7	7.2

Ranked Results of MI Analyses

The appropriate hyperparameter *number of trees* in the Random Forest are shown in **Figure 3**, that depicts the plot of results of the grid search step, showing the curve (red line) that was selected to compute the number of trees in the further steps of the process. By looking at the chart on **Figure 4**, we found 18 features responsible for the maximum f1-score that represent good candidates for the metabolomics analysis.

Table 3 shows the ranked results for the 10 validation experiments using the optimal feature vectors and the Diagnosis Classifier final test results trained with P_{fit} and tested with P_{test} partitions respectively.

From the 18 most discriminant features, eight were identified as markers, as shown in **Table 4**. The first-ranked feature



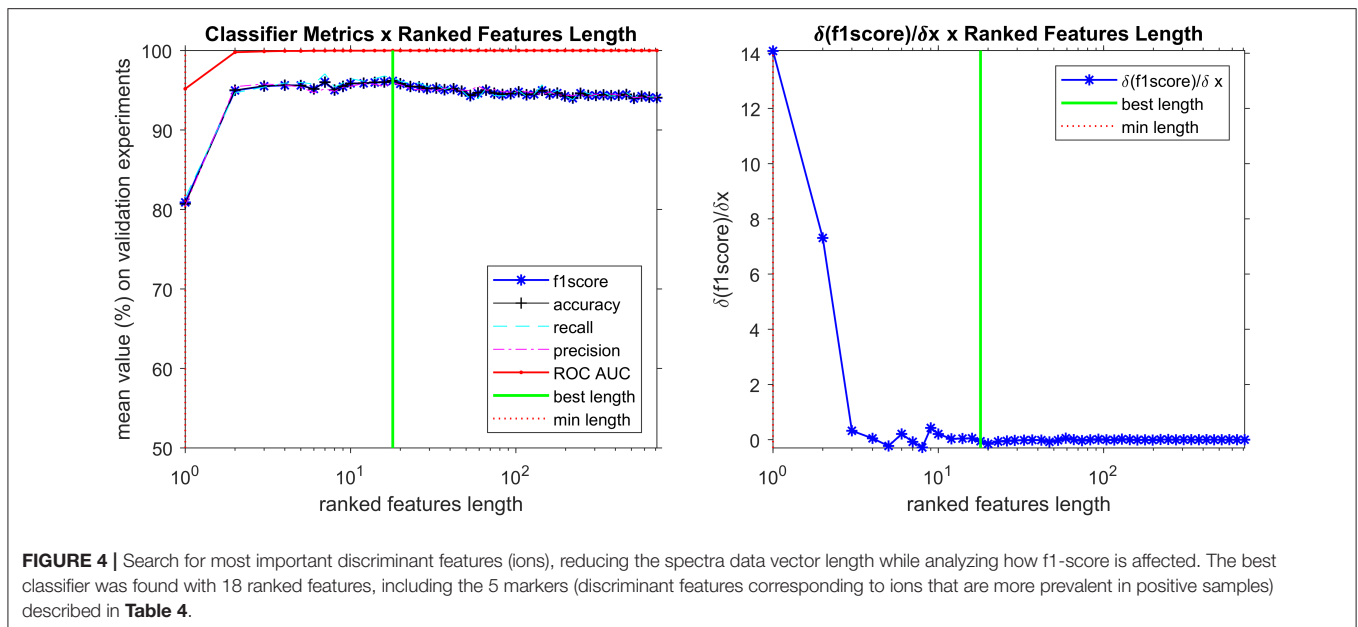


TABLE 3 | Classification results of the validation tests and the final test of the Diagnosis Classifier using the 18 most discriminant features.

Metric	Validation		Final test
	Mean	RSD	
Vector length	18		18
Number of trees	58		58
Accuracy (%)	96.2	3.0	86.1
Sensitivity (%)	96.6	5.2	94.4
Specificity (%)	95.8	3.5	77.8
Precision (%)	95.9	3.3	81.0
F1-score (%)	96.2	3.2	87.2

TABLE 4 | The 18 most discriminant features found by the Random Forest analysis, highlighting the eight markers found herein.

Rank	Marker	m/z	Rank	Marker	m/z
1	Yes	299	10	Yes	283
2	No	673	11	Yes	389
3	Yes	278	12	No	252
4	No	672	13	No	656
5	Yes	263	14	No	395
6	No	379	15	Yes	270
7	Yes	304	16	No	315
8	No	657	17	No	462
9	No	250	18	Yes	308

analysis is displayed on **Figure 5**. Distribution Analysis Chart for 18 Most Discriminant Features are displayed in **Figure S1**. **Table 5** presents the ranked results when using only the markers features on the validation experiments, showing their capability to identify positive samples. Results of 18 most discriminant features are show in **Table 6**. All results are statistically comparable.

Biomarkers Characterization

Thereafter, the eight markers identified underwent metabolomics techniques to identify which ones are weight gain biomarkers, as described next herein. After checking the exact masses in the metabolomics databases and literature, five biomarkers were characterized: Dihydrobiopterin, Argininosuccinate, 3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid (CMPF), a prostaglandin (PGB2) and a metabolite of leukotriene (Carboxy-LTB4). The description of identified markers is listed in **Table 7**. The Heatmap analysis distribution of 5 biomarkers identified is displayed in **Figure S2**.

With the support of previous studies, it was possible to demonstrate a link between weight gain and systemic inflammation, and describe the roles of biomarkers in the pathways associated to an imbalance in the Nitric Oxide (NO) cycle and the overweight status-related metabolic stress.

DISCUSSION

Upon elucidation of weight gain-associated biomarkers, it was possible to demonstrate the presence of Argininosuccinate [*m/z* 308.1565], a basic amino acid and immediate precursor of arginine in the urea cycle. Numerous studies have shown a relationship between arginine, obesity and other metabolic disturbance (Pallares-Méndez et al., 2016; Zhang et al., 2017), but its exact role and related metabolic pathways still require a more robust amount of studies (Lent-Schochet et al., 2019). Argininosuccinate is synthesized from the condensation of citrulline and aspartate by the enzyme argininosuccinate synthase; then, it is cleaved into fumarate and arginine by the enzyme argininosuccinate lyase (Haines et al., 2011).

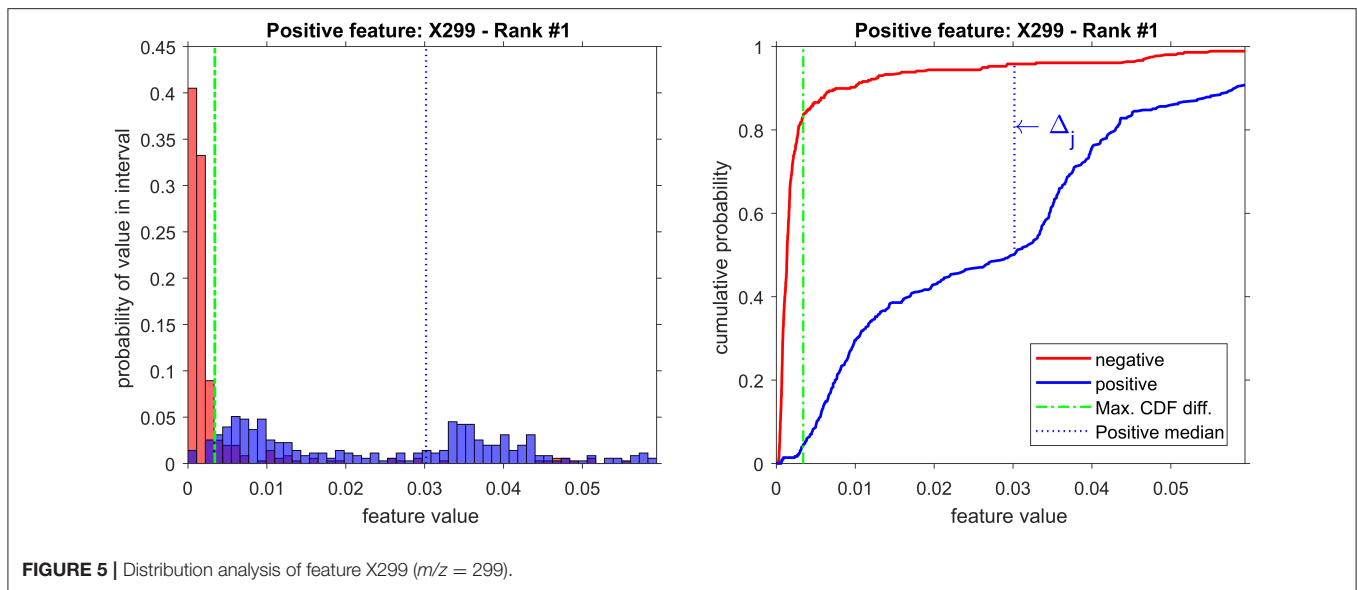


FIGURE 5 | Distribution analysis of feature X299 ($m/z = 299$).

TABLE 5 | Results of the 10 experiments validation with Random Forest classifiers trained with the eight markers.

Metric	Mean	RSD
Vector length	8	
Number of trees	58	
Accuracy (%)	90.9	4.0
Sensitivity (%)	93.5	3.9
Specificity (%)	88.4	7.3
Precision (%)	89.5	6.2
F1-score (%)	91.3	3.6

TABLE 6 | Comparison of validation results of 18 most discriminant features using different classifiers.

Metric	SVM (SMO)		SVM (ISDA)		Random Forest (GDI)		Tree (GDI)	
	Mean	RSD	Mean	RSD	Mean	RSD	Mean	RSD
Accuracy (%)	96.6	2.3	97.2	2.3	96.5	3.3	96.2	3.5
Sensitivity (%)	97.4	4.5	98.6	3.0	97.1	5.0	95.9	5.7
Specificity (%)	95.8	5.0	95.8	5.0	95.8	4.8	96.5	3.7
Precision (%)	96.2	4.3	96.2	4.3	96.0	4.7	96.6	3.7
F1-score (%)	96.6	2.2	97.3	2.1	96.5	3.5	96.1	3.7

SVM with two different optimization algorithms: SMO, Sequential Minimal Optimization (Fan et al., 2005); ISDA, Iterative Single Data Algorithm (Kecman et al., 2005), and single decision tree classifier using (GDI) Gini Diversity Index (Breiman et al., 1984) as the split criterion. Green font means best classifier and red font worst classifier for each metric (table line).

Interestingly, according to studies carried out with rodents, argininosuccinate may actively participate in the shift of M2 macrophages to M1 in adipose tissue, indicating that this molecule is a potential marker of activation of proinflammatory macrophages in human obesity as well (Fraternal et al., 2015; Jha et al., 2015; Kuda et al., 2016).

Dihydrobiopterin (BH_2) [m/z 278.0650] is a product of the oxidation of Tetrahydrobiopterin (BH_4), an important cofactor

for NO synthase (NOS). Whenever there is a balance between cofactors and substrates in relation to Nicotinamide adenine dinucleotide phosphate (NADPH), NOS catalyzes the reduction of O_2 and incorporates it into the guanidine group of L-arginine, generating NO. Nonetheless, in specific situations, such as the oxidation of BH_4 into BH_2 , a molecule that cannot work as a cofactor for NOS, a phenomenon called “uncoupling” occurs. Uncoupled NOS generates superoxide instead of nitric oxide, leading to an oxidative stress state. Further, produced reaction oxygen species (ROS) may oxidize BH_4 into BH_2 , exacerbating this state (Incalza et al., 2018). It is known that reactive oxygen species may induce and increase the inflammatory process by upregulating different genes involved in the inflammatory response, such as those that induce the production of cytokines and adhesion molecules (Galvão et al., 2018). One of the effects of ROS on cells is the modification of fatty acids from the phospholipid membrane, a process that alters membrane fluidity and cell signaling. In oxidative stress, the enzyme phospholipase A presents increased activity, producing active mediators such as arachidonic acid (AA) metabolites (Balboa and Balsinde, 2006).

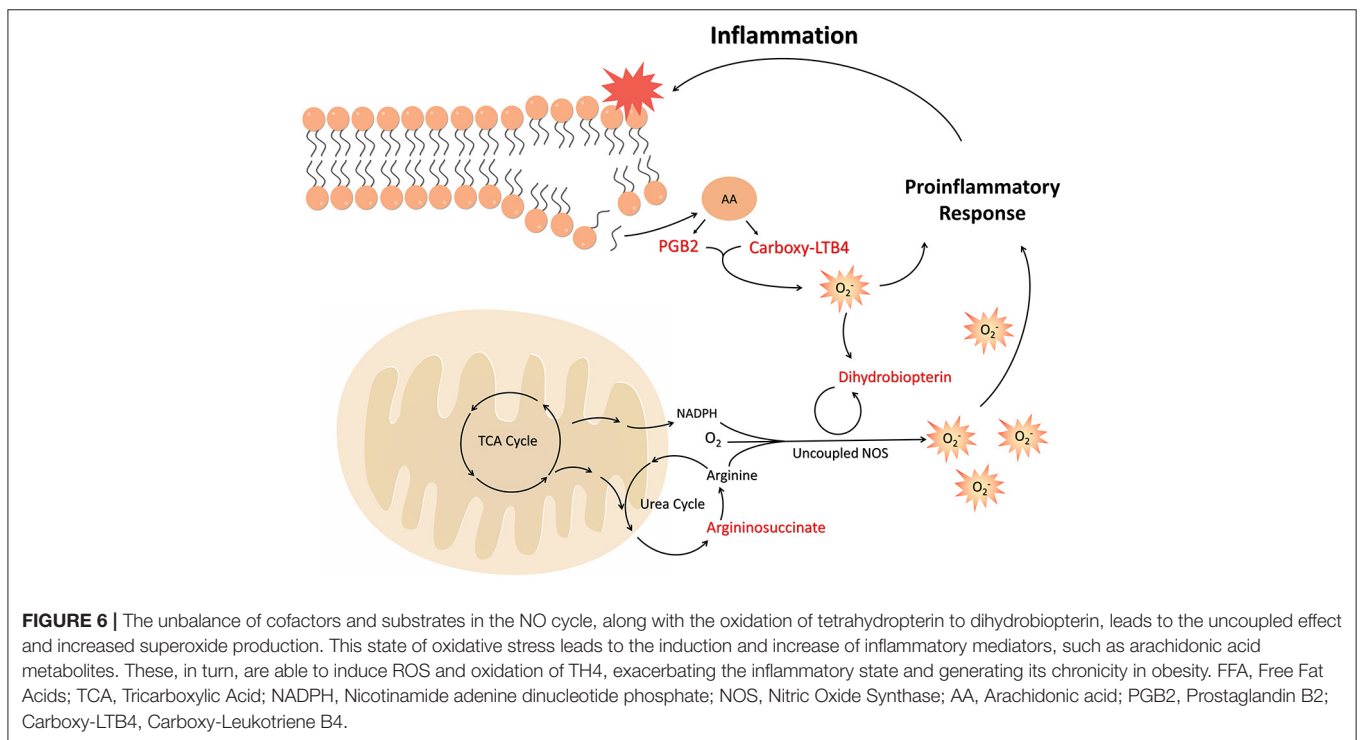
AA metabolites, namely Eicosanoids, are a family of lipid mediators generated from phospholipid precursors, involved in distinct cellular processes. In an inflammatory environment, the enzyme Phospholipase A2 catalyzes the esterified AA present in the phospholipid membrane into free AA, which may be then oxidized in four different pathways: lipoxygenase (LOX), cyclooxygenase (COX), epoxygenase (CYTP450) and isoprostane. Prostaglandins are produced through the COX pathway, and leukotrienes through the LOX pathway. Two markers found in the plasma of the case group are AA metabolites: Prostaglandin B2 [m/z 299.2017] and Carboxy-Leukotriene B4 [m/z 389.1935], indicating that at least two AA oxidation pathways are putatively more active in the case group.

The most studied Prostaglandin, with the highest pro-inflammatory effects described in prior art, is PGE2. The biomarker found herein is obtained from the dehydration of

TABLE 7 | Markers elected by Random Forest from plasma analysis of case group.

Nominal mass	Exact mass	Mass fragments	Adduct	Theoretical mass	Error (ppm)	Metlin ID	Molecule
278	278.0655	232–260–236–246	[M+K] ⁺	278.0650	1.79	65,872	Dihydrobiopterin
299	299.2022	263–271–213–281	[M+H-2H ₂ O] ⁺	299.2017	1.67	3,466	PGB2
308	308.1571	290–248–209	[M+NH ₄] ⁺	308.1565	1.94	389	Argininosuccinate
389	389.1942	353–319–371–285	[M+Na] ⁺	389.1935	1.79	36,247	Carboxy-LTB4
263	263.0885	245 [227*]–165	[M+Na] ⁺	263.0890	–1.90	45,041	CMPF

*MS3.



PGE2 and has a number of studies describing its effects. In a study with cell cultures, Cattani et al. evaluated the regulation of different prostaglandins in an inflammation site. The main finding was that PGB2 induced Interleukin 2 (IL-2), activation of tyrosine kinase activities, and Nuclear factor kappa B (NF-κB). According to the authors, PGB2 behaves as a potent lipid mediator and mimics almost all effects of PGE2 (Cattani et al., 2000).

LTB4, the precursor of the other biomarker involved with AA pathways in this study, is a potent lipid mediator capable of promoting chemotaxis, neutrophil degranulation, and release of enzymes and ROS. LTB4 has a receptor with high specificity, B leukotriene receptor 1 (BLT1), expressed exclusively in inflammatory cells such as neutrophils, macrophages and eosinophils. *In vivo*, LTB4 is rapidly metabolized by different pathways in an attempt to decrease the inflammatory effects caused by this mediator. From neutrophil action, and with the participation of ω-hydroxylase, LTB4 is oxidized to 20-carboxy LTB4. The oxidized products of LTB4 were believed to be inactive species, but binding studies demonstrated that 20-carboxy LTB4 has the ability to bind to BLT1 and had the same

chemotactic effects, suggesting a similar action of this molecule on inflammation (Wang et al., 2000; Toda et al., 2002).

Given the fact that these AA-derived metabolites are closely associated with inflammation, and since this study was not tissue-specific, their presence reinforce the concept that the inflammatory activity associated with obesity occurs systemically. This result agrees with several previous studies that have demonstrated the relation of obesity to other complications and diseases in several organs, such as insulin resistance, type 2 diabetes, dyslipidemia, and disorders in immune function (Doupa et al., 2014; Finn, 2015; Kohlgruber and Lynch, 2015; Velloso et al., 2015).

A number of studies published previously have shown alterations in mitochondrial activity, NO cycle, and increase of reactive oxygen species at the onset of obesity, including childhood and juvenile obesity (Williams et al., 2002; Gruber et al., 2008). According to Muñoz and Costa, the increase in glucose and free fatty acids from energy imbalance causes an increase in the production of Acetyl-Coenzyme A (Acetyl-CoA) in the Tricarboxylic Acid (TCA) cycle, which stimulates the excessive production of superoxide and, consequently, the

activation of inflammatory stimuli (Muñoz and Costa, 2013). In a study that evaluated gene expression and cytokine dosage in plasma and tissue before and after induction of obesity in animals fed a high-fat diet, Matsuzawa-Nagata et al. demonstrated that up-regulation of ROS gene expression preceded the onset of obesity, and also the production of Tumor Necrosis Factor alpha (TNF- α), suggesting that the onset of obesity is related to an imbalance in ROS production (Matsuzawa-Nagata et al., 2008).

Considering that obesity is related to the increase of Acetyl-CoA combined with disturbances in the NO cycle and consequent activation of inflammatory mediators from the release of ROS, the biomarkers found are actively involved in the process of onset of obesity and development of chronicity of low-grade inflammation in a pathway described in **Figure 6**.

Finally, another marker found was 3-carboxy-4-methyl-5-propyl-2-furanpropanoic acid (CMPF) [m/z 263.0890], a dibasic urofuran acid catabolized from furan fatty acids incorporated into phospholipids and cholesterol esters, already described in the literature as related to uremia. Recent studies have found a relationship of this metabolite with a beta cell dysfunction and the development of diabetes (Prentice et al., 2014; Liu et al., 2016). Considering that there were diabetic patients within our case group, further studies are needed to attest whether this molecule may also be related to obesity.

Several studies have sought to elucidate biomarkers through metabolomic techniques in the contexts of obesity, diabetes, cardiovascular diseases, and MetS. The main purposes supporting these contributions are to broaden the understanding on the metabolic environment of individuals in these different conditions, as well as potential molecular targets that link them all (Newgard, 2017). Studies have already shown the relationship between cardiometabolic diseases, obesity and diabetes with disturbances in the amino acid pathways such as aromatics and branched-chain amino acids (BCAA) (Newgard, 2017; Rauschert et al., 2017; Zhong et al., 2017; Libert et al., 2018; Del Coco et al., 2019).

In an extensive review of the literature, Zhang et al. have collected several studies that associate obesity with markers present in different metabolic pathways, such as lipids, lysophosphatidylcholines, monosaccharides, acylcarnitines, and metabolites related with TCA cycle, tryptophan, phenylalanine, and tyrosine pathways (Zhang et al., 2017). In a study conducted in Australia, Huynh et al. used robust lipidomics techniques to determine markers related to cardiometabolic risk factors and anthropometric measures in human plasma. According to the study, there are 338 plasma lipid species related with BMI, being positively associated species such as sphingomyelin, phosphatidylcholine, ceramides, and phosphatidylethanolamine (Huynh et al., 2019).

Our results also demonstrate that the field of metabolomics can contribute comprehensively to the study of the onset of obesity and its development. From the plasma of case patients, with the metabolomics strategies combined with advances in data processing using machine learning, it was viable to elect biomarkers of obesity and describe a possible pathway of inflammation in this and other associated comorbidities. This study contributes with knowledge on aspects of the relationship between obesity, systemic inflammation and chronicity. Our

approach—contribute for elucidation of potential therapeutic targets in obesity and other disorders and pathologies. Despite that metabolomics is a sophisticated method, that requires expensive equipment and trained personnel, the simplicity in sample preparation, short time to analyze metabolites, and the use of modern data processing techniques for the determination of obesity biomarkers are the new insights proposed in this study. Finally, it also collaborates with upcoming studies in the fields of prevention and diagnosis of obesity.

DATA AVAILABILITY STATEMENT

All data will be available in MetaboLights accession MTBLS1400 as soon as the curation process at Metabolights is finished. Thus, we cannot provide anything immediately, as at this stage our actions are all complete and we depend on the system workflow; we expect, however, that the process will be over within the next couple of weeks.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Human Ethics Committee of the State University of Londrina, Paraná, Brazil. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

FD-A and DO performed mass spectrometry experiments. LN and AR conceived and executed the ML method. FD-A, DO, and LN wrote the manuscript. FD-A, DO, TG, CM, and JD performed data analysis. FD-A, DP, FR, and MW performed patient recruitment, biofluids management, and clinical support. FD-A and JD processed serum samples. FD-A, LN, DO, LV, AR, and RC performed manuscript proofreading and prepared tables and figures. RC idealized all experiments and managed the research group.

FUNDING

This work was supported by funds from CAPES and CNPq (Grant No. 141200/2018-9 for FD-A fellowship); and by the São Paulo Research Foundation (FAPESP) Grant No. 2019/05718-3 for JD fellowship.

ACKNOWLEDGMENTS

This work was supported by the OCRC (Obesity Comorbidities Research Center) CEPID: CMPO—Centro Multidisciplinar de Pesquisa em Obesidade e Doenças Associadas (FAPESP number: 13/07607-8).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fbioe.2020.00006/full#supplementary-material>

REFERENCES

- Aguilar, M., Bhuket, T., Torres, S., Liu, B., and Wong, R. J. (2015). Prevalence of the metabolic syndrome in the United States, 2003–2012. *JAMA* 313, 1973–1974. doi: 10.1001/jama.2015.4260
- Andersen, C. J., Murphy, K. E., and Fernandez, M. L. (2016). Impact of obesity and metabolic syndrome on immunity. *Adv. Nutr.* 7, 66–75. doi: 10.3945/an.115.010207
- Balboa, M. A., and Balsinde, J. (2006). Oxidative stress and arachidonic acid mobilization. *Biochim. Biophys. Acta* 1761, 385–391. doi: 10.1016/j.bbali.2006.03.014
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324
- Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Pacific, CA: Wadsworth & Brooks/Cole Advanced Books & Software.
- Cattan, N., Mary, D., Peleraux, A., Mari, B., Aussel, C., and Rossi, B. (2000). Prostaglandin B2 delivers a co-stimulatory signal leading to T cell activation. *Eur. Cytokine Netw.* 11, 293–300.
- Cooke, P. S., and Naaz, A. (2004). Role of estrogens in adipocyte development and function. *Exp. Biol. Med.* 229, 1127–1135. doi: 10.1177/153537020422901107
- Del Coco, L., Vergara, D., De Matteis, S., Mensà, E., Sabbatinelli, J., Prattichizzo, F., et al. (2019). NMR-based metabolomic approach tracks potential serum biomarkers of disease progression in patients with Type 2 diabetes mellitus. *J. Clin. Med.* 8:720. doi: 10.3390/jcm8050720
- Dias-Audibert, F. L., Navarro, L. C., Oliveira, D. N., Melo, C. F. O. R., Guerreiro, T. M., Dabaja, M. Z., et al. (2019). *Combining Machine Learning and Metabolomics to Identify Weight Gain Biomarkers*. Campinas: IEEE Dataport.
- Doupa, D., Seck, S. M., Dia, C. A., Diallo, F. A., Kane, M. O., Kane, A., et al. (2014). Dyslipidemia, obesity and other cardiovascular risk factors in the adult population in Senegal. *Pan Afr. Med. J.* 19:181. doi: 10.11604/pamj.2014.19.181.4872
- Fan, R. E., Chen, P. H., and Lin, C. J. (2005). Working set selection using second order information for training support vector machines. *J. Mach. Learn. Res.* 6, 1889–1918. doi: 10.5555/1046920.1194907
- Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *J. Mach. Learn. Res.* 15, 3133–3181. doi: 10.5555/2627435.2697065
- Finn, P. (2015). Dyslipidemia in overweight and obese school-aged children. *NASN Sch. Nurse* 30, 255–257. doi: 10.1177/1942602X15588369
- Fraternali, A., Brundu, S., and Magnani, M. (2015). Polarization and repolarization of macrophages. *J. Clin. Cell. Immunol.* 6:2. doi: 10.4172/2155-9899.1000319
- Galvão, I., Sugimoto, M. A., Vago, J. P., Machado, M. G., and Sousa, L. P. (2018). “Mediators of Inflammation,” in *Immunopharmacology and Inflammation*, eds C. Riccardi, F. Levi-Schaffer, and E. Tiligada (Belo Horizonte: Springer), 3–32. doi: 10.1007/978-3-319-77658-3_1
- Gregor, M. F., and Hotamisligil, G. S. (2011). Inflammatory mechanisms in obesity. *Annu. Rev. Immunol.* 29, 415–445. doi: 10.1146/annurev-immunol-031210-101322
- Gruber, H., Mayer, C., Mangge, H., Fauler, G., Grandits, N., and Wilders-Truschning, M. (2008). Obesity reduces the bioavailability of nitric oxide in juveniles. *Int. J. Obes.* 32, 826–831. doi: 10.1038/sj.ijo.0803795
- Guilherme, A., Virbasius, J. V., Puri, V., and Czech, M. P. (2008). Adipocyte dysfunctions linking obesity to insulin resistance and type 2 diabetes. *Nat. Rev. Mol. Cell. Biol.* 9, 367–377. doi: 10.1038/nrm2391
- Haines, R. J., Pendleton, L. C., and Eichler, D. C. (2011). Argininosuccinate synthase: at the center of arginine metabolism. *Int. J. Biochem. Mol. Biol.* 2, 8–23.
- Hanselmann, M., Köthe, U., Kirchner, M., Renard, B. Y., Amstalden, E. R., Glunde, K., et al. (2009). Toward digital staining using imaging mass spectrometry and random forests. *J. Proteome Res.* 8, 3558–3567. doi: 10.1021/pr900253y
- 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, 71–84.e74. doi: 10.1016/j.chembiol.2018.10.008
- Incalza, M. A., D’Oria, R., Natalicchio, A., Perrini, S., Laviola, L., and Giorgino, F. (2018). Oxidative stress and reactive oxygen species in endothelial dysfunction associated with cardiovascular and metabolic diseases. *Vascul. Pharmacol.* 100, 1–19. doi: 10.1016/j.vph.2017.05.005
- Jackson, A. S., and Pollock, M. L. (1978). Generalized equations for predicting body density of men. *Br. J. Nutr.* 40, 497–504. doi: 10.1079/BJN19780152
- Jackson, A. S., Pollock, M. L., and Ward, A. (1980). Generalized equations for predicting body density of women. *Med. Sci. Sports Exerc.* 12, 175–181. doi: 10.1249/00005768-198023000-00009
- Jha, A. K., Huang, S. C.-C., Sergushichev, A., Lampropoulou, V., Ivanova, Y., Loginicheva, E., et al. (2015). Network integration of parallel metabolic and transcriptional data reveals metabolic modules that regulate macrophage polarization. *Immunity* 42, 419–430. doi: 10.1016/j.immuni.2015.02.005
- Jordan, M. I., and Mitchell, T. M. (2015). Machine learning: trends, perspectives, and prospects. *Science* 349, 255–260. doi: 10.1126/science.aaa8415
- Kecman, V., Huang, T. M., and Vogt, M. (2005). “Iterative single data algorithm for training kernel machines from huge data sets: theory and performance,” in *Support Vector Machines: Theory and Applications. Studies in Fuzziness and Soft Computing, Vol. 177*, ed L. Wang (Berlin; Heidelberg: Springer). doi: 10.1007/10984697_12
- Kohlgruber, A., and Lynch, L. (2015). Adipose tissue inflammation in the pathogenesis of type 2 diabetes. *Curr. Diab. Rep.* 15:92. doi: 10.1007/s11892-015-0670-x
- Kononenko, I. (2001). Machine learning for medical diagnosis: history, state of the art and perspective. *Art. Intell. Med.* 23, 89–109. doi: 10.1016/S0933-3657(01)00077-X
- Kuda, O., Rombaldova, M., Janovska, P., Flachs, P., and Kopecky, J. (2016). Cell type-specific modulation of lipid mediator’s formation in murine adipose tissue by omega-3 fatty acids. *Biochem. Biophys. Res. Commun.* 469, 731–736. doi: 10.1016/j.bbrc.2015.12.055
- Lent-Schochet, D., McLaughlin, M., Ramakrishnan, N., and Jialal, I. (2019). Exploratory metabolomics of metabolic syndrome: a status report. *World J. Diabetes* 10:23. doi: 10.4239/wjcd.v10.i1.23
- Li, Y., Kuhn, M., Gavin, A. C., and Bork, P. (2019). Identification of metabolites from tandem mass spectra with a machine learning approach utilizing structural features. *BioRxiv* 573790. doi: 10.1101/573790
- Libert, D. M., Nowacki, A. S., and Natowicz, M. R. (2018). Metabolomic analysis of obesity, metabolic syndrome, and type 2 diabetes: amino acid and acylcarnitine levels change along a spectrum of metabolic wellness. *PeerJ* 6:e5410. doi: 10.7717/peerj.5410
- Liu, Y., Prentice, K. J., Eversley, J. A., Hu, C., Batchuluun, B., Leavey, K., et al. (2016). Rapid elevation in CMPF may act as a tipping point in diabetes development. *Cell Rep.* 14, 2889–2900. doi: 10.1016/j.celrep.2016.02.079
- Matsuzawa-Nagata, N., Takamura, T., Ando, H., Nakamura, S., Kurita, S., Misu, H., et al. (2008). Increased oxidative stress precedes the onset of high-fat diet-induced insulin resistance and obesity. *Metab. Clin. Exp.* 57, 1071–1077. doi: 10.1016/j.metabol.2008.03.010
- Melo, C. F. O., Delafiori, J., de Oliveira, D. N., Guerreiro, T. M., Esteves, C. Z., Lima, E., et al. (2017). Serum metabolic alterations upon Zika infection. *Front. Microbiol.* 8:1954. doi: 10.3389/fmicb.2017.01954
- Melo, C. F. O. R., Navarro, L. C., De Oliveira, D. N., Guerreiro, T. M., Lima, E. O., Delafiori, J., et al. (2018). A machine learning application based in random forest for integrating mass spectrometry-based metabolomic data: a simple screening method for patients with zika virus. *Front. Bioeng. Biotechnol.* 6:31. doi: 10.3389/fbioe.2018.00031
- Muñoz, A., and Costa, M. (2013). Nutritionally mediated oxidative stress and inflammation. *Oxid. Med. Cell. Longev.* 2013:610950. doi: 10.1155/2013/610950
- Newgard, C. B. (2017). Metabolomics and metabolic diseases: where do we stand? *Cell Metab.* 25, 43–56. doi: 10.1016/j.cmet.2016.09.018
- Pallares-Méndez, R., Aguilar-Salinas, C. A., Cruz-Bautista, I., and del Bosque-Plata, L. (2016). Metabolomics in diabetes, a review. *Ann. Med.* 48, 89–102. doi: 10.3109/07853890.2015.1137630
- Prentice, K. J., Luu, L., Allister, E. M., Liu, Y., Jun, L. S., Sloop, K. W., et al. (2014). The furan fatty acid metabolite CMPF is elevated in diabetes and induces β cell dysfunction. *Cell Metab.* 19, 653–666. doi: 10.1016/j.cmet.2014.03.008
- Rappez, L., Stadler, M., Triana, S., Phapale, P., Heikenwalder, M., and Alexandrov, T. (2019). Spatial single-cell profiling of intracellular metabolomes *in situ*. *BioRxiv* 510222. doi: 10.1101/510222
- 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, 201–209. doi: 10.1159/000459635
- Sartipy, P., and Loskutoff, D. J. (2003). Monocyte chemoattractant protein 1 in obesity and insulin resistance. *Proc. Natl. Acad. Sci. U.S.A.* 100, 7265–7270. doi: 10.1073/pnas.1133870100
- Toda, A., Yokomizo, T., and Shimizu, T. (2002). Leukotriene B4 receptors. *Prostaglandins Other Lipid Mediat.* 68, 575–585. doi: 10.1016/S0090-6980(02)00056-4
- Velloso, L. A., Folli, F., and Saad, M. J. (2015). TLR4 at the crossroads of nutrients, gut microbiota, and metabolic inflammation. *Endocr. Rev.* 36, 245–271. doi: 10.1210/er.2014-1100
- Wang, S., Gustafson, E., Pang, L., Qiao, X., Behan, J., Maguire, M., et al. (2000). A novel hepatointestinal leukotriene B4 receptor cloning and functional characterization. *J. Biol. Chem.* 275, 40686–40694. doi: 10.1074/jbc.M004512200
- Williams, I., Wheatcroft, S., Shah, A., and Kearney, M. (2002). Obesity, atherosclerosis and the vascular endothelium: mechanisms of reduced nitric oxide bioavailability in obese humans. *Int. J. Obes.* 26:754. doi: 10.1038/sj.ijo.0801995
- World Health Organization (1995). *Physical Status: The Use of and Interpretation of Anthropometry*. Report of a WHO Expert Committee.
- World Health Organization (2018). *Obesity and Overweight, Fact Sheets*. World Health Organization. Available online at: <http://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight> (accessed July 7, 2018).
- Zhang, A., Sun, H., and Wang, X. (2017). Emerging role and recent applications of metabolomics biomarkers in obesity disease research. *RSC Adv.* 7, 14966–14973. doi: 10.1039/C6RA28715H
- Zhong, F., Xu, M., Bruno, R. S., Ballard, K. D., and Zhu, J. (2017). Targeted high performance liquid chromatography tandem mass spectrometry-based metabolomics differentiates metabolic syndrome from obesity. *Exp. Biol. Med.* 242, 773–780. doi: 10.1177/1535370217694098

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.

Copyright © 2020 Dias-Audibert, Navarro, de Oliveira, Delafiori, Melo, Guerreiro, Rosa, Petenuci, Watanabe, Velloso, Rocha and Catharino. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). 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.