

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

Integrated serum proteins and fatty acids analysis for putative biomarker discovery in inflammatory bowel disease

This is a pre print version of the following article:

Original Citation:

Availability:

This version is available <http://hdl.handle.net/2318/1682992> since 2018-12-03T12:58:44Z

Published version:

DOI:10.1016/j.jprot.2018.10.017

Terms of use:

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

Integrated Plasma Proteins and Fatty Acids Analysis for Biomarker Discovery in Inflammatory Bowel Disease

Marcello Manfredi^{1,2,3}, Eleonora Conte², Elettra Barberis^{1,2}, Arianna Buzzi¹, Elisa Robotti¹, Valeria Caneparo^{3,4,5}, Daniela Cecconi⁶, Jessica Brandl⁶, Ester Vanni⁷, Marco Astegiano⁷, Marisa Gariglio^{3,5}, Emilio Marengo^{1,2,3}, Marco De Andrea^{3,4,5}

¹ Università del Piemonte Orientale, Dipartimento di Scienze ed Innovazione Tecnologica, Alessandria, Italy; ² ISALIT, Spin-off del DISIT, Università del Piemonte Orientale, Italy; ³Center for Translational Research on Autoimmune and Allergic Disease - CAAD, University of Piemonte Orientale, Novara, Italy; ⁴University of Turin, Medical School, Department of Public Health and Pediatric Sciences, Viral Pathogenesis Unit, Turin, Italy; ⁵University of Piemonte Orientale, Department of Translational Medicine, Virology Unit, Novara, Italy; ⁶Department of Biotechnology, University of Verona; ⁷AOU Città della Salute e della Scienza - Molinette Hospital, Gastroenterology and Hepatology Unit, Turin, Italy,
E-mail: marcello.manfredi@uniupo.it

Inflammatory bowel disease (IBD) is a chronic inflammatory condition of unknown etiology that can develop in any portion of the digestive tract but which has a high incidence especially in the terminal ileum and/or in the colon. The diagnosis and management of IBD, including Crohn's disease (CD) and ulcerative colitis (UC), is still challenging, and the presence of intestinal inflammation is, to date, a primary criterion for its diagnosis and differentiation from other diseases. Moreover, there is no decisive gold standard diagnostic test based on patient history and physical examination, supplemented with objective laboratory, radiological, endoscopic and histological findings.

This study analyzed plasma proteins and fatty acids using mass spectrometry-based techniques. Quantitation of low-abundance plasma proteins was performed by depleting 14 high-abundance proteins, followed by tryptic digestion and LC-MS analysis, while fatty acids were analyzed using GC-MS.

We used bioinformatic tools to identify several new potential biomarkers for an early and non-invasive diagnosis of IBD, and to differentiate CD from UC. Moreover, the diagnostic power of the MS-identified biomarkers was also corroborated by Western Blot and ELISA assays. Hence, we identified the biological functions and pathways involved in the various subsets of IBD.

Coagulation, fibrinolysis and acute phase response processes were found to be strongly involved in the condition. The involvement of several fatty acids, such as anti-inflammatory mediators, was also identified. Finally, proteomic and lipidomic data were integrated by using combinatorial and multivariate analyses to discover new combined biomarkers and to study the molecular pathways involved in IBD.

KEYWORDS: Inflammatory bowel disease, Proteomics, Lipidomics, Biomarkers

Introduction

Inflammatory bowel disease (IBD) is a chronic inflammatory condition of unknown etiology that can develop in any portion of the digestive tract but which has a high incidence especially in the terminal ileum and/or the colon [1]. IBD can be sub-grouped into two common forms, precisely Crohn's disease (CD) and ulcerative colitis (UC). Previous studies reported that the disease pathogenesis involves genetic factors, an abnormal immune response to microbial infections and gut-microbiota imbalance [2-5].

Current treatment focuses on reducing and controlling inflammation. There is no cure, and the majority of IBD patients remain under medical care and management for life [6]. Diagnosis and management of IBD is still a challenge because the presence of intestinal inflammation is a primary criterion for diagnosis and differentiation from other diseases, but no decisive diagnostic tests are available. The diagnosis is based on patient history and physical examination, along with laboratory, radiological, endoscopic and histological findings [7].

Discovering new biomarkers is a challenging process, and researchers have not been entirely successful at uncovering biomarkers that can reliably differentiate IBD patients from healthy controls, CD from UC, or predict disease outcome or response to treatment [8]. Recently, Piras et al. investigated the serum proteomic profile of early and advanced CD in order to identify differentially expressed proteins in acute CD and during the course of the disease. Inflammatory proteins and complement C3 chain (C3c) were over-represented during early CD, while clusterin, retinol-binding protein, α 1-microglobulin and transthyretin were under-represented [9]. Moreover, C-reactive protein (CRP) correlates with disease activity and prognosis of IBD, but its specificity is insufficient [10, 11]. Although fecal markers, such as calprotectin or lactoferrin, are currently being tested or used as diagnostic tools for the evaluation of IBD, results have been largely unsatisfactory, especially in terms of clinical outcome and response to therapy. For instance, calprotectin, as biomarker of IBD, presents several limitations, such as poor specificity, no validated cut off value to define active disease/clinical remission, and dependence on physiological factors [12].

Lipids play a pivotal role in IBD pathogenesis since they are a static metabolic energy reserve that is employed in cellular signal transduction pathways. Lipids are involved in chronic intestinal inflammation, and several studies have focused on dietary lipids, the role of lipids in mucosal immunity, lipids in inflammatory signal transduction, alterations in lipids associated with the emergence of chronic inflammation, and lipid-based therapies [13-16].

This study adopted a multi-omics approach to analyze plasma proteins and lipids in UC and CD patients at a very early stage of diagnosis or presenting with mild symptoms. Interest in the quantitation of large proteomes across multiple samples has rapidly increased in recent years, especially stimulated by the development of new instruments and tools. Moreover, low-abundance human plasma proteins are considered the most promising biomarkers for disease diagnosis and therapeutic monitoring.

We used an immunodepletion method to reduce the levels of the 14 high-abundance proteins, followed by shotgun proteomics with liquid chromatography mass spectrometry (LC-MS/MS). We also investigated the role of serum fatty acids, which are very important because their expression can be modulated during inflammatory processes. Several fatty acids were simultaneously quantified using gas chromatography mass spectrometry (GC-MS). The bioinformatic and multivariate analysis of the proteomic and lipidomic fingerprint of patients' serum samples allowed to identify several potential biomarkers and pathways involved in the pathology.

Materials and Methods

Selection of Subjects

Fifteen patients with Crohn's disease (CD: 7 men and 8 women, mean age 47 years, range 30-65 years) and 13 patients with ulcerative colitis (UC: 8 men and 5 women, mean age 56 years, range 39-75 years) from the Gastroenterology and Hepatology Department, Città della Salute e della Scienza, Molinette Hospital (Turin, Italy) were enrolled in this study. All diagnoses had been previously confirmed by clinical, endoscopic, and histological criteria. At the time of enrollment, serum samples were collected and stored at -80°C until use, and clinical and serological features were recorded. All patients selected for this study had a low or moderate disease activity (Harvey-Bradshaw Index (HBI) for patients with CD; Mayo score for patients with UC). The demographic, clinical, and laboratory data obtained at the time of blood sampling are reported in Table 1.

Written informed consent was obtained from all participants as established by the Declaration of Helsinki, and approval of the study protocol was obtained from the local ethics committees.

Chemicals

All chemicals and materials, if not specifically stated otherwise, were purchased from Sigma-Aldrich, St Louis, MO, USA.

Sample preparation procedure for serum proteome and fatty acid profiles

12 μL of sera were depleted of high-abundance proteins using the Seppro IgY14 spin column kit (Sigma-Aldrich Inc., St. Louis, MO, USA) according to the manufacturer's procedure. The method was used to bind human serum HSA, IgG, fibrinogen, transferrin, IgA, IgM, haptoglobin, alpha 2-macroglobulin, alpha 1-acid glycoprotein, alpha 1-antitrypsin, Apo A-I HDL, Apo A-II HDL, complement C3 and LDL (ApoB) in order to increase low-abundance protein identification. The samples were transferred into an Amicon Ultra-0.5 mL 3 kDa centrifugal filter (Millipore, Billerica, MA, USA) following the manufacturer's procedure to collect high molecular weight proteins. The samples were then subjected to denaturation with TFE, to reduction with DTT 200 mM, to alkylation with IAM 200 mM and to complete protein digestion with 2 μg of Trypsin/Lys-C (Promega, Madison, WI, USA). The peptide digests were desalted on the Discovery® DSC-18 solid phase extraction (SPE) 96-well plate (25 mg/well) (Sigma-Aldrich Inc., St. Louis, MO, USA). The SPE plate was preconditioned with 1 mL of acetonitrile and 2 mL of water. After loading the sample, the SPE was washed with 1 mL of water. The adsorbed proteins were eluted with 800 μL of acetonitrile:water (80:20) [17]. After the desalting process, the sample was vacuum-evaporated and reconstituted with 20 μL of 0.05% formic acid in water. 2 μL of stable-isotope-labeled peptide standard (DPEVRPTSAVAA, Val- $^{13}\text{C}515\text{N}1$ at V10, Cellmano Biotech Limited, Anhui, China) was spiked into the samples before the LC-MS/MS analysis and used for instrument quality control.

The preparation for plasma derivatization of fatty acids consists in collecting 30 μL of serum in PTFE screw-capped Pyrex tubes. Derivatization was performed with 2 mL of methanol acetyl chloride (10%) and 500 μL of n-hexane vigorously shaken in a water bath at 95°C for 1 h. Cooling to RT was followed by the addition of 5 mL of 6% potassium carbonate solution. 100 μL of the n-hexane top layer was transferred into a 500 μL auto-sampler vial and crimped [18].

LC-MS/MS analyses

Sera proteins were analyzed with the micro-LC Eksigent Technologies (Eksigent, Dublin, USA) system that included a micro LC200 Eksigent pump with flow module 5-50 μL , interfaced with a 5600+ TripleTOF system (AB Sciex, Concord, Canada) equipped with

DuoSpray Ion Source and CDS (Calibrant Delivery System). The stationary phase was a Halo C18 column (0.5 x 100 mm, 2.7 μm ; Eksigent Technologies Dublin, USA). The mobile phase was a mixture of 0.1% (v/v) formic acid in water (A) and 0.1% (v/v) formic acid in acetonitrile (B), eluting at a flow-rate of 15.0 $\mu\text{L min}^{-1}$ with an increasing concentration of solvent B from 2% to 40% in 30 min. Injection volume was 4.0 μL and oven temperature was set at 40°C. For identification purposes the samples were subjected to data-dependent acquisition (DDA): mass spectrometer analysis was performed using a mass range of 100–1500 Da (TOF scan with an accumulation time of 0.25 s), followed by an MS/MS product ion scan from 200 to 1250 Da (accumulation time of 5.0 ms) with the abundance threshold set at 30 cps (35 candidate ions can be monitored during every cycle). The ion source parameters in electrospray positive mode were set as follows: curtain gas (N_2) at 25 psig, nebulizer gas GAS1 at 25 psig, and GAS2 at 20 psig, ion spray voltage floating (ISVF) at 5000 V, source temperature at 450°C and declustering potential at 25 V.

For the label-free quantification process the samples were subjected to cyclical data-independent analysis (DIA) of the mass spectra, using a 25-Da window: the mass spectrometer was operated so that a 50-ms survey scan (TOF-MS) was performed and subsequent MS/MS experiments were performed on all precursors. These MS/MS experiments were carried out in a cyclical manner using an accumulation time of 40 ms per 25-Da swath (36 swaths in total) for a total cycle time of 1.5408 s. The ions were fragmented for each MS/MS experiment in the collision cell using the rolling collision energy. The MS data were acquired with Analyst TF 1.7 (AB SCIEX, Concord, Canada). Two DDA and three DIA acquisitions were performed [19, 20].

Protein database search

The DDA files were searched using Protein Pilot software v. 4.2 (SCIEX, Concord, Canada) and Mascot v. 2.4 (Matrix Science Inc., Boston, USA). Trypsin as digestion enzyme was specified for both software. For Mascot we used 2 missed cleavages, set the instrument to ESI-QUAD-TOF and specified the following modifications for the assay: carbamidomethyl cysteine as fixed modification and oxidized methionine as variable modification. An assay tolerance of 50 ppm was specified for peptide mass tolerance, and 0.1 Da for MS/MS tolerance. The peptide charges to be detected were set to 2+, 3+ and 4+, and the assay was set on monoisotopic mass [21]. The UniProt Swiss-Prot reviewed database containing human proteins (version 2015.07.07, containing 42,131 sequence entries) was used and a target-decoy database search was performed. False Discovery Rate was fixed at 1%.

Protein quantification

Quantification was performed by integrating the extracted ion chromatogram of all the unique ions for a given peptide. SwathXtend was employed to build an integrated assay library with the DDA acquisitions, using a protein FDR threshold of 1% [22].

Quantification was carried out with PeakView 2.0 and MarkerView 1.2. (ABSCIEX, Concord, Canada). Six peptides per protein and six transitions per peptide were extracted from the SWATH files. Shared peptides were excluded as well as peptides with modifications. Peptides with FDR lower than 1.0% were exported in MarkerView for the t-test.

GC-MS analysis

Gas chromatography–time of flight mass spectrometry (GC-TOF/MS) was performed using the Agilent 7890B GC (Agilent Technologies, USA) and Pegasus (BT) TOF-MS system (Leco Corporation, USA) equipped with an Rxi-5ms column (30m×0.25mm×0.25µm, RESTEK, USA), stationary phase 5% diphenyl-95% dimethyl polysiloxane. High-purity helium (99.999%) was used as the carrier gas at a flow rate of 1.20 mL/min⁻¹. Samples were injected in splitless mode at 250°C. The chromatographic conditions were: initial temperature 50°C, 1 minute isothermal, 25°C/min up to 175°C, 4°C/min up to 250°C, 11.25 minutes isothermal. MS parameters: electron impact ionization source temperature (EI, 70 eV) was set at 250°C; scan range 40/630 m/z, with an extraction frequency of 30 kHz. The chromatograms were acquired in TIC (total ion current) mode. Mass spectral assignment was performed by matching with NIST MS Search 2.2. Libraries, implemented with the MoNa Fiehns Libraries.

Western Blot analysis

Equal amounts of proteins extracted from serum samples were resolved on a 4-20% SDS-polyacrylamide gel and electroblotted onto PVDF membranes (Roche Diagnostic). Membranes were then blocked in 5% non-fat dried milk in Tris-buffered saline (TBS-T, 50 mM Tris pH 7.5, 0.9% NaCl, 0.1% Tween 20) for 1 h at room temperature. Zinc-alpha-2-glycoprotein (ZAG) was detected by using anti-ZAG (1D4) (1:500, Santa Cruz Biotechnology, #sc-13585) primary antibody, diluted in 1% non-fat dried milk in TBS-T for 3 h at room temperature. Blots were then incubated for 1 h at room temperature with anti-mouse IgGκ-Horseradish peroxidase (HRP)-conjugated secondary antibody (1:2000, Santa Cruz Biotechnology, #sc-516102). The immunocomplexes were visualized by

chemiluminescence using the Chemidoc MP imaging system (Bio-Rad Laboratories). Signal intensity of the bands was measured by using Image Lab software (Bio-Rad Laboratories), and mean protein quantification was performed by two independent experiments, designed to contain three serum samples from Crohn's Disease (CD), Ulcerative Colitis (UC) and randomly selected healthy patients per blot.

ELISA assay

Kallistatin protein was analyzed with the DuoSet ELISA Human Serpin A4/Kallistatin kit (cod. DY1669, R&D Systems, Minneapolis, MN, USA), according to the manufacturer's instructions.

Bioinformatics and Statistics Software

The identified proteins were classified with Funrich using a Gene ontology database [23]. The regulated proteins were analyzed by using STRING software (<http://string-db.org>), which is a database of known and predicted protein-protein interactions. Cytoscape 3.1.0 plug-ins ClueGO v2.0.8 were also used for functional annotation clustering and network analysis of proteins.

The fatty acid identification analysis was performed with the ChromaTOF BT software with the NIST library implemented with MoNa Fiehns Libraries. Statistical and pathway studies were carried out by MetaboAnalyst 4.0 tool.

The Mann-Whitney test analysis was performed by GraphPad Prism 5 software. P values <0.01 are indicated as (**). CombiROC software was employed for ROC analysis [24].

Multivariate statistical analysis and data pre-treatment were performed by The Unscrambler X (Camo Inc., Norway), Matlab 2014b (The MathWorks Inc., Natick, USA) and the Classification Toolbox for Matlab (Milano Chemometrics group, University of Milano Bicocca, Italy) [25]. Graphical representations were performed by Statistica v7.1 (Statsoft, Tulsa, OK, USA). Details on multivariate methods employed can be found in supplementary materials.

Results

The plasma proteomic and lipidomic profiles of 28 patients with IBD were studied using a shotgun proteomic approach and GC-MS, respectively. The plasma analysis of 15 patients with Crohn's disease (CD) and 13 patients with ulcerative colitis (UC) were compared to 17 healthy subjects. Fold change analysis of proteins and lipids was used to identify potential

biomarkers for IBD, CD and UC, to study the pathways involved in the pathology and to better understand the biological mechanisms of the diseases. Moreover, proteomics and lipidomics data were integrated by using chemometric and bioinformatic tools to identify new combined biomarkers. Figure 1 reports the experimental workflow of this research.

Plasma protein analysis

The presence of high-abundance plasma proteins, such as albumin, may mask the under represented ones, which are usually of diagnostic utility since their level of expression and regulation may change depending on the diseases. All plasma samples were depleted of the 14 high-abundance proteins, digested and then analyzed with SWATH-MS to obtain a proteomic fingerprint of each patient. The LC-MS analysis allowed the identification of 538 proteins, 367 of which were quantified across all the samples.

Gene ontology classification was carried out to characterize the plasma proteome profiles of the dataset. Identified proteins for each type of sample were categorized into biological process and molecular function. As reported in Supplementary Figure 1, the most represented biological processes were complement activation (12%), complement regulation (12%), proteolysis (11%) and innate immune response (11%), whereas the most represented molecular functions were: protein binding (58%), serine-type endopeptidase activity (11%) and antigen binding (8%).

Fold change analysis of IBD patients versus healthy subjects revealed the presence of 32 modulated proteins, with a p -value ≤ 0.05 and fold change ≥ 1.5 . 15 proteins were under-expressed in IBD, while 17 proteins were overexpressed (Figure 2a). STRING analysis of the regulated proteins showed that the disease affects proteins that are related to the acute phase response and to complement activity (up-regulation), and to the regulation of protease function, to blood coagulation and to oxygen transport (down-regulation). Gene ontology classification based on biological processes confirmed the STRING results and highlighted the down-regulation of proteins linked to lipoprotein metabolic processes (Figures 2b and 2c). In addition, the pathway analysis performed with the Cytoscape software and the ClueGO plug-in showed an increased secretion of acute phase response protein, complement activation pathway, and a negative regulation of coagulation and oxidative processes in IBD patients (Supplementary Figure 2).

The plasma protein expression was also evaluated to highlight the differences between CD and UC, and to identify potential biomarkers for the two sub-classes of IBD. Figure 3 shows the heat map of log fold changes of regulated proteins in IBD, CD and UC patients,

compared to healthy subjects. 42 proteins were modulated in CD patients, 17 were up-regulated and 25 were down-regulated, while UC patients were characterized by 24 up-regulated and 19 down-regulated proteins. Some proteins were modulated only for CD (i.e., KAIN, PRCC and GELS) or UC (i.e., LPPRC, SURF4 and CHADL), suggesting that they may be used as potential biomarkers for the sub-classification of IBD.

Validation of the proteomic results

In order to corroborate the results obtained with shotgun mass spectrometry, we selected zinc-alpha-2-glycoprotein (ZAG) and kallistatin (KAIN) for Western Blot and ELISA validation, respectively.

The expression levels of ZAG protein in CD and UC serum samples was compared to healthy subjects. As reported in Figure 4a, the relative abundance of sera levels of ZAG2 protein were significantly lower in CD and UC, in comparison to controls. Then we performed Western Blot validation by using a polyclonal antibody specific for ZAG2. As shown in figure 4b, the optical density measurements of the bands revealed an approximately 2-fold statistically significant decrease in the amount of ZAG protein in CD and UC samples, compared to healthy individuals.

The expression level of KAIN protein in CD, UC and healthy subjects was measured using the ELISA method. Briefly, MS-results revealed that the median values of the sera levels of KAIN protein were significantly lower in CD than in UC and in healthy controls (Figure 4c). As reported in figure 4d, the ELISA analysis performed in the same sera revealed that KAIN protein is down-regulated in CD patients, while it is not modulated in UC patients, confirming our MS-based results (Figure 4d).

Plasma fatty acid analysis

As previously reported, the role of lipids in mucosal immunity and in chronic inflammation makes their study pivotal to understand IBD pathogenesis. A gas chromatographic analysis of plasma samples was also performed to investigate changes in the fatty acid profile between UC and CD patients. After the derivatization process and the instrumental analysis, the eluted compounds were identified by matching their raw mass spectra and retention index with previously existing databases. The fatty acids were then analyzed using bioinformatic tools to reveal modulated molecules and to underscore their main biological processes.

GC-MS analysis allowed to identify 37 non-redundant and common fatty acids for UC, CD and healthy patients. The fold change calculation was carried out by the MetaboAnalyst software: eleven fatty acids were modulated in CD patients, while only two were regulated in UC. Table 1 reports the fold changes of differentially regulated fatty acids in UC, CD and IBD (UC+CD) patients, compared to healthy subjects (fold change ≥ 1.5 and p-value ≤ 0.05).

Data integration and multivariate analysis

Integrating multi-omics data offers an unprecedented and original opportunity to assess molecular interactions and to provide new combined biomarkers for the diagnosis and prognosis of diseases. In this study we used combined ROC curves and multivariate data analysis of multi-omics data to integrate and evaluate the performance of biomarkers and models of the two IBD subclasses.

ROC curves and combined ROC for marker identification

To determine if a multi-model could improve performance over single biomarkers, we employed CombiROC software that uses combinatorial analysis to select the best combinations and ROC curves for automatic comparisons. Data were not subjected to any transformation step. All the possible combinations of modulated biomarkers (both proteins and lipids) were obtained. Table 2 reports the area under the curve (AUC), the sensitivity (SE), the specificity (SP) and the optimal cut off value for the best markers for CD and UC. CombiROC generated ROC curves for single markers but also for a marker combo with even higher sensitivity and specificity. For UC patients, the combo LAMP2_HUMAN, PTGDS_HUMAN, SURF4_HUMAN, tridecanoic acid was found in those with the highest AUC, while for CD patients the best combo was CBG_HUMAN, KPCA_HUMAN, LITD1_HUMAN, 13-Docosenoic acid, -4,7,10,13,16,19-Docosahexaenoic acid.

CombiROC analysis revealed the best combination of markers, made up of markers that would not achieve sufficient AUC, SE and SP when individually considered.

Multivariate data analysis

Proteomic and lipidomic studies, combined with multivariate data analysis methods, allowed to perform a highly efficient multifactorial biomarker analysis. Several hundreds of molecules (proteins and lipids) were measured in every single sample, obtaining an individual

biochemical fingerprint. These complex profiles were converted into a data matrix, which was then analyzed using multivariate methods.

The dataset was arranged in a matrix of 40 samples (17 controls, 15 patients affected by CD and 13 patients affected by UC) and 399 variables (362 corresponding to the signal of the quantified proteins, and 37 corresponding to the signal of the quantified fatty acids). Regarding proteins, data were first row-scaled to eliminate variations caused by minor differences in the total protein content of the initial sample, while fatty acid signals were normalized for the internal standard to eliminate random variations. Then, the complete dataset was autoscaled before Partial Least Square – Discriminant Analysis (PLS-DA). PLS-DA was applied to perform four different comparisons:

- CD samples vs. control and vs. all other samples (UC and controls together);
- UC samples vs. controls and vs. all other samples (CD and controls together).

A variable selection procedure was applied for backward elimination in all four cases: one variable was eliminated at a time according to the lowest error in cross-validation. The variable characterized by the least importance was eliminated at each iteration. Leave-more-out cross-validation was applied with 5 cancellation groups. The final results presented in cross-validation were calculated by random sampling, taking out 20% of the samples of each class at each iteration, for a total of 1,000 iterations.

Significance of the information about fatty acids. First of all, PLS-DA models were built based on the variables selected by the univariate approach, both including and excluding fatty acids, to evaluate if fatty acids contain significant information about the pathology, and if they potentially include interesting markers. This evaluation was limited to the comparison of UC and CD samples vs. controls, since these two cases were the only ones for which significant markers could be identified with the univariate approach. Table 3 reports the results in terms of classification performance of the PLS-DA models built including the proteins identified as significant by the univariate approach, both including or excluding fatty acids. The different number of latent variables included guarantees that the smallest cross-validation error is achieved for each model. The results show that fatty acids include interesting information about sample classification, since classification performance of the models is always better both in terms of fit and of cross-validation when fatty acids are included.

Multivariate models were, therefore, calculated by always including both proteins and fatty acids.

Biomarkers for CD. CD samples were compared to both control samples and UC plus control samples. Specific markers for CD are identified in both cases; however, in the first case they allow to differentiate CD patients from healthy subjects, while in the second case the identified markers allow to differentiate CD patients not only from healthy subjects but also from patients presenting another IBD pathology (UC).

In both cases, PLS-DA was carried out between the X-block variables and a Y variable coded to assign -1 to non-CD samples (controls or control plus UC samples) and +1 to patients with CD. The final PLS-DA models retain 112 identified variables (90 proteins and 12 fatty acids) discriminating CD vs. control samples, and 97 variables (91 proteins and 6 fatty acids), discriminating CD vs. all other samples. In both cases 3 LVs were included in the final model.

Figures 5 a and b present the score plot of the first three LVs: in both cases the samples are correctly separated into the two classes.

The regression coefficients of each variable on the final models are reported in Supplementary Information 1: a positive coefficient corresponds to variables with a larger signal in CD patients and a smaller signal in controls or in all other samples, while variables with a negative coefficient show the opposite behavior. A rise in the absolute value of the regression coefficient corresponds to an increase in the influence of the corresponding variable on the final model.

Table 4 reports the classification performance of PLS-DA models with backward elimination variable selection (BE-PLS-DA), and of PLS-DA models calculated including the proteins. Fatty acids were statistically significant according to the univariate approach. The different number of LVs included guarantees that the smallest cross-validation error is achieved for each model. The 70% of variables identified as relevant by the univariate approach are also included in the PLS-DA models (73% in the case of CD vs. all other samples), showing a good overlap between the two approaches (mono- and multi-variate) for what regards the identification of biomarkers; however, BE-PLS-DA ensures a more exhaustive search for biomarkers.

Regarding the comparison between CD and control samples, both the multivariate and the univariate models can correctly classify all samples in the fit but, in this case, the multivariate approach could provide a better performance during cross-validation and a more exhaustive identification of candidate biomarkers, improving the ability to identify interesting up- and/or down-regulated pathways as a result of the pathology. The

multivariate model guarantees a perfect classification of the samples in the comparison between CD and other samples, and very good cross-validation performance, while the univariate approach shows a worse performance both in terms of fit and of cross-validation. Again, the multivariate model provides a more exhaustive identification of candidate biomarkers, including not only proteins but also 6 fatty acids, while no fatty acids were considered significantly expressed in CD vs. other samples by the univariate approach. Other classification parameters (specificity, precision and sensitivity) are reported in Supplementary Information 2.

Biomarkers for UC. As for CD samples, UC patients were compared to both control and CD plus control individuals. In both cases, PLS-DA was carried out between the X-block variables and a Y variable coded to assign -1 to non-UC samples (controls or control plus CD samples) and +1 to patients affected by UC. The final PLS-DA models retain 81 identified variables (75 proteins and 6 fatty acids) discriminating UC vs. control samples, and 106 variables (100 proteins and 6 fatty acids) discriminating UC vs. all other samples. 3 LVs were included in the final model for the comparison between UC and other samples, while just one LV was sufficient for the discrimination of UC and control samples.

Figures 5 c and d present the score plot of the two models calculated: the score plot reports the first LV in the comparison UC vs. control samples, and three LVs in the comparison UC vs. all other samples. In both cases the patients are correctly separated into the two classes.

The regression coefficients of each variable on the final model are reported in Supplementary Information 1: as for CD samples, a positive coefficient corresponds to variables with a larger signal in UC patients and smaller signals in controls or in other samples, while variables with a negative coefficient present the opposite behavior.

Table 4 reports the classification performance of the PLS-DA models with backward elimination variable selection (BE-PLS-DA), and of the PLS-DA models calculated including the proteins. Finally, the fatty acids were statistically significant according to the univariate approach, and the number of LVs indicates very small cross-validation errors. In this case too, the % of variables identified as relevant by the univariate approach included in the PLS-DA models is about 70%, thus presenting a good overlap between the two approaches.

Comparing UC and control samples, both the multivariate and the univariate models provide 100% of correct classifications in terms of fit, with the multivariate approach providing better performance in cross-validation tests and a more exhaustive identification

of candidate biomarkers, with a larger number of proteins and fatty acids identified as significant. The multivariate model's classification performance is much better both regarding fit and cross-validation, and also for the comparison between UC and other samples, while guaranteeing a more exhaustive identification of candidate biomarkers, including not only proteins but also 6 fatty acids. Other classification parameters (specificity, precision and sensitivity) are reported in Supplementary Information 2.

Discussion

Accurate classification of UC and CD patients is essential in the management of IBD due to their different response to medication, surgical indications and prognosis. Currently, the diagnosis is still not precise, and several clinical tests are required, including radiology, endoscopy and histopathology investigations [26]. Though many non-invasive biomarkers of IBD disease activity have been described, none of them has been fully validated for early diagnosis and sub-classification between CD and UC [27].

Our proteomic results confirmed the presence in IBD sera of proteins involved in coagulation and fibrinolysis, such as complement system components, C-reactive protein (CRP), serum amyloid A (SAA) and fibronectin (FINC). This is typical of an acute-phase reaction. CRP, which is one of the most important acute-phase proteins, has been shown to correlate with disease activity and prognosis of IBD, but its specificity is unsatisfactory, and it may indicate a general inflammation [10]. Serum amyloid A, which is involved in binding and removing cholesterol from inflammatory lesions, increases in inflammatory conditions, such as CD and UC [28]. Fibronectin too can be included in this class of molecules as it is essential for wound healing and plays a central role in cell adhesion, growth, migration, and differentiation. Alterations in this protein can also be observed in other conditions, such as cancer and fibrosis [29]. Our study observed that CRP and SAA are highly expressed in IBD patients, while fibronectin is up-regulated only in CD patients.

Extensive evidence underscores the connection between inflammation, coagulation, and fibrinolysis [30]. Coagulation factor XIII A chain (F13A) and histidine-rich glycoprotein (HRG) are strongly involved in these mechanisms, and their concentrations show a tendency to decrease in our cohort. Factor XIII is the final enzyme in the coagulation cascade and is essential for hemostasis and fibrinolysis. During an inflammatory flare in CD, consumption of this factor increases in digestive mucosa, with a subsequent decrease in its serum levels [31].

Histidine-rich glycoprotein (HRG), which regulates several physiological processes, such as tissue remodeling, angiogenesis and tumor metastasis [32], decreases only in our UC patients: usually its levels decline significantly during acute inflammation and surgery. Circulating HRG levels have been found to be significantly lower in patients with systemic lupus erythematosus, suggesting the active involvement of HRG during acute as well as chronic inflammation [33, 34].

The expression of protease inhibitors in IBD is quite variable, reporting increased, decreased or stable levels of expression for serpins [35]. Kallistatin (KAIN), a human serine proteinase inhibitor, is a newly identified tissue kallikrein inhibitor. It strongly binds tissue kallikrein but weakly binds other serine proteinases, such as chymotrypsin and elastase. Down-regulation of plasma kallistatin has already been described in IBD tissues, and our data confirm this trend [36]. In fact, our CD patients showed down-regulation of KAIN, compared to healthy subjects. This was also confirmed by ELISA.

Anemia, a common systemic complication and/or extra-intestinal manifestation in IBD, is usually defined as a decrease in the total amount of red blood cells (RBCs) or hemoglobin in blood. Anemia has already been reported in both CD and UC patients with percentages of 21% and 27%, respectively [37, 38]. Consistently, hemoglobin levels in our IBD patients are low in comparison to controls, suggesting the presence of anemia.

Plasma lipid and lipoprotein alterations in inflammatory bowel disease have already been described: there is an increase in triglycerides (TGs) due to an accumulation of very low density lipoprotein (VLDL), and a decrease in high-density lipoprotein (HDL) and apolipoprotein A-I levels, along with apolipoprotein C-I (APOC1) [39].

Fatty acids have been historically considered simple membrane components that serve as structural elements and energy storing entities, but today they are increasingly recognized as potent signaling molecules involved in many metabolic processes. Fatty acids, particularly n-3 and n-6 polyunsaturated fatty acids (PUFAs), mediate a number of key biological processes, including eicosanoid production, inflammation, cell membrane physiological processes, and gene regulation and expression [40]. Long-chain dietary n-3 PUFAs have been shown to inhibit the expression of genes that activate inflammatory processes [41]. n-6 PUFAs are metabolized into lipid inflammatory mediators, such as prostaglandin E2 and leukotriene B4. Arachidonic acid (AA) is mobilized by phospholipase A2 family members to synthesize lipid mediators at sites of cellular damage or inflammation. This study observed that three fatty acids are down-regulated in IBD patients, namely eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA) and linoleic acid (LA), which are

well-known anti-inflammatory mediators in biological processes. The signaling functions of several classes of lipids discovered in the past decade resolve rather than promote inflammation [42]. E-series resolvins originate from EPA, while DHA generates the family of D-series resolvins (RvD), protectin D1, and maesins [43]. Resolvins are biosynthesized endogenously during the resolution phase of inflammation. Animal studies have revealed that RvE1, RvE2, and RvE3 inhibit leukocyte/neutrophil migration associated with a reduction in the release of pro-inflammatory cytokines and an increase in the phagocytic activity of macrophages that, in turn, cause tissue “cleaning”. These results are consistent with our data: EPA was slightly lower in CD patients, compared to the controls. A reduction in the weight percentage of EPA may contribute to the exacerbation of inflammation in CD patients. Moreover, eicosapentaenoic acid is a precursor to docosahexaenoic acid (DHA): lower levels of EPA have a negative effect on DHA levels. We also found calcium metabolism dysregulation related to nervonic acid, which is an omega-9 monounsaturated fatty acid. This compound is a regulator of the Ca²⁺ ion channel in the cell membrane. Patients with gut disorders are unable to retain calcium properly, and they need dietary supplements to increase its levels. Down-regulation of omega-9 MUFA is thus consistent with the onset of the disease [44].

Down-regulated fatty acids include palmitic acid (PA), which has low levels in CD patients. PA belongs to the omega-7 monounsaturated fatty acid family. Recent studies have reported a correlation between MUFA and anti-inflammatory effects. This novel fatty acid called omega-7 can help break the cycle of high blood sugar, elevated lipid levels, inflammation, and excess fat gain, as well as enhance insulin sensitivity. Moreover, dietary intake of omega-7 caused a reduction in C-reactive protein levels, which are closely related to this kind of pathologies [45]. Negative regulation of saturated fatty acids can be related to an inability to retain certain substances in the inflamed intestine: tridecylic, caprylic, behenic, lignoceric, arachidic and tricosylic acid showed negative fold change values.

Concluding, the use of mass spectrometry and bioinformatic techniques allows rapid and robust analysis of hundreds of biological molecules in serum. Moreover, they enable investigators to extend the analysis to proteins and metabolites that were previously not suspected of being involved in IBD pathogenesis.

It is widely deemed that serum levels of the majority of biomarkers for disease diagnosis and prognosis are low. Depletion of the major serum proteins before the final analysis allowed us to identify several potential biomarkers and a proteomic signature of CD and UC.

Moreover, since lipids play a pivotal role in IBD pathogenesis, we also investigated the lipidomic profiles of UC and CD, identifying new potential markers and lipid functions associated with IBD. We used chemometric and bioinformatic tools to identify several pathways and biological functions involved in IBD, but we also integrated the multi-omics information to discriminate between the two subclasses of CD and UC. Comparing the mono- and the multi-variate approaches, both models showed excellent classification abilities both in terms of fit and cross-validation, with the multivariate approach providing slightly better results and guaranteeing more exhaustive marker identification.

The results of this study identified potential candidate biomarkers that should be further validated in a bigger cohort. They deserve further evaluation and new hypotheses for a mechanistic investigation of IBD.

References

1. Baumgart DC, Carding SR. Inflammatory bowel disease: cause and immunobiology *The Lancet* 2007;369(9573): 1627–1640.
2. Presley LL, Ye J, Li X, LeBlanc J, Zhang Z, Ruegger PM, et al. Host–microbe relationships in inflammatory bowel disease detected by bacterial and metaproteomic analysis of the mucosal–luminal interface. *Inflammatory Bowel Diseases* 2012;18(3):409–417.
3. Anderson CA, Boucher G, Lees CW, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nature genetics* 2011;43:246–252
4. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;491:119–124.
5. O’Toole A, Korzenik J. Environmental triggers for IBD. *Current gastroenterology reports* 2014;16:396.
6. Wasinger VC, Yau Y, Duo X, Zeng M, Campbell B, Shin S, Lubner R, Redmond D, Leong RW. Low Mass Blood Peptides Discriminative of Inflammatory Bowel Disease (IBD) Severity: A Quantitative Proteomic Perspective. *Mol Cell Proteomics* 2016;15(1):256-65.

7. Travis S, Stange E, Lémann M, Öresland T, Chowers Y, Forbes A, et al. European evidence based consensus on the diagnosis and management of Crohn's disease: current management. *Gut* 2006;55:i16–i35.
8. Townsend P, Zhang Q, Shapiro J, Webb-Robertson BJ, Bramer L, et al. Serum Proteome Profiles in Stricturing Crohn's Disease: A Pilot Study. *Inflamm Bowel Dis.* 2015;21(8):1935-41
9. Piras C, Soggiu A, Greco V, et al. Serum protein profiling of early and advanced stage Crohn's disease. *EuPA Open Proteomics* 2014;3:48–59.
10. Vermeire S, Van Assche G, Rutgeerts P. C-reactive protein as a marker for inflammatory bowel disease. *Inflammatory Bowel Diseases* 2004;10(5):661–665.
11. Meuwis MA, Fillet M, Geurts P, De Seny D, Lutteri L, Chapelle JP, et al. Biomarker discovery for inflammatory bowel disease: using proteomic serum profiling. *Biochemical Pharmacology* 2007;73(9):1422–1433.
12. Lehmann FS, Burri E, Beglinger C. The role and utility of faecal markers in inflammatory bowel disease. *Therap Adv Gastroenterol* 2015;8(1):23-36.
13. Nee J, Feuerstein JD. Optimizing the care and health of women with inflammatory bowel disease. *Gastroenterol Res Pract* 2015;2015:435820.
14. Elia PP, Tolentino YF, Bernardazzi C, de Souza H.S. The role of innate immunity receptors in the pathogenesis of inflammatory bowel disease. *Mediators Inflamm* 2015;2015:936193.
15. Andrade ME, Araújo RS, de Barros PA, Soares AD, Abrantes FA, Generoso Sde V, Fernandes SO, Cardoso VN. The role of immunomodulators on intestinal barrier homeostasis in experimental models. *Clin Nutr* 2015;34:1080-1087.
16. Victoria CR, Sasaki LY, Nunes HRC. Incidence and prevalence rates of inflammatory bowel diseases, in midwestern of São Paulo State, Brazil. *Arq Gastroenterol* 2009;46:20-25.
17. Guo X, Kristal BS. The use of underloaded C(18) solid-phase extraction plates increases reproducibility of analysis of tryptic peptides from unfractionated human plasma. *Anal Biochem* 2012;426:86-90.
18. Ecker J, Scherer M, Schmitz G, Liebisch G. A rapid GC–MS method for quantification of positional and geometric isomers of fatty acid methyl esters. *Journal of Chromatography B* 2012;897:98– 104.
19. Martinotti S, Patrone M, Manfredi M, Gosetti F, Pedrazzi M, Marengo E, Ranzato E. HMGB1 Osteo-Modulatory Action on Osteosarcoma SaOS-2 Cell Line: An Integrated Study From Biochemical and -Omics Approaches. *J Cell Biochem* 2016;117:2559-2569.
20. Manfredi M, Martinotti S, Gosetti F, Ranzato E, Marengo E. The secretome signature of malignant mesothelioma cell lines. *J Proteomics* 2016;145:3-10.
21. Cvijetic S, Bortolotto V, Manfredi M, Ranzato E, Marengo E, Salem R, Canonico PL, Grilli M, *GLIA* 2017;65:169–181.
22. Jemma X. Wu, Xiaomin Song, Dana Pascovici, Thiri Zaw, Natasha Care, Christoph Krisp and Mark P. Molloy, SWATH mass spectrometry performance using extended peptide MS/MS assay libraries, *Molecular and Cellular Proteomics*, 2016;15(7):2501-14.

23. Pathan M, Keerthikumar S, Chisanga D, Alessandro R, Ang CS, et al. A novel community driven software for functional enrichment analysis of extracellular vesicles data. *J Extracellular Vesicles*. 2017;1:1321455.
24. Mazzara S, Rossi RL, Grifantini R, Donizetti S, Abrignani S, Bombaci M, CombiROC: an interactive web tool for selecting accurate marker combinations of omics data, *Sci Rep*. 2017;30;7:45477
25. Ballabio D, Consonni V, Classification tools in chemistry. Part 1: Linear models. PLS-DA. *Analytical Methods* 2013;5:3790-3798.
26. Chan, Patrick PY, Wasinger VC, Leong RW. Current Application of Proteomics in Biomarker Discovery for Inflammatory Bowel Disease. *World Journal of Gastrointestinal Pathophysiology* 2016;7.1:27–37.
27. Sands BE. Biomarkers of Inflammation in Inflammatory Bowel Disease. *Gastroenterology* 2015;149(5):1275-1285.
28. Chambers RE, Stross P, Barry RE, Whicher JT. Serum amyloid A protein compared with C-reactive protein, alpha 1-antichymotrypsin and alpha 1-acid glycoprotein as a monitor of inflammatory bowel disease. *Eur J Clin Invest*. 1987;17(5):460-7.
29. Riedera F, Fiocchi C. Intestinal fibrosis in inflammatory bowel disease: progress in basic and clinical science. *J Crohns Colitis* 2008;2(4):279-90
30. Devaraj S, Xu DY, Jialal I. C-reactive protein increases plasminogen activator inhibitor-1 expression and activity in human aortic endothelial cells: Implications for the metabolic syndrome and atherothrombosis. *Circulation* 2003;107:398–404.
31. Cougard PA, Desjeux A, Vitton V, Baumstarck-Barrau K, Lesavre N, Grimaud JC. The usefulness of factor XIII levels in Crohn's disease. *J Crohns Colitis* 2012;6(6):660-664.
32. Poon IK1, Olsson AK, Hulett MD, Parish CR., Regulation of histidine-rich glycoprotein (HRG) function via plasmin-mediated proteolytic cleavage, *Biochem J*. 2009 Oct 23;424(1):27-37.
33. Castel M, Horellou MH, Conard J, Samama M, Davidson JF, Bachmann F, Bouvier CA. Immunochemical determination of histidine-rich glycoprotein in healthy subjects and in a clinical population, EKO Kruithof. Churchill Livingstone, Edinburgh 1983;370–373, (Progress in fibrinolysis).
34. Badolaa S., Spurlinga H., Robisona K., Fedyka ER, Silvermanb GA, Straylec J, Kapellera R, Tsua CA. Correlation of serpin–protease expression by comparative analysis of real-time PCR profiling data. *Genomics* 2006;88(2):173-84.
35. Vergnolle N. Protease inhibition as new therapeutic strategy for GI diseases. *Gut* 2016;65(7):1215–1224.
36. Stadnicki A, Mazurek U, Gonciarz M, Plewka D, Nowaczyk G, Orchel J, Pastucha E, Plewka A, Wilczok T, Colman RW. Immunolocalization and expression of kallistatin and tissue kallikrein in human inflammatory bowel disease. *Dig Dis Sci*. 2003;48(3):615-23.
37. Filmann N, Rey J, Schneeweiss S, Ardizzone S, Bager P, Bergamaschi G, Koutroubakis I, Lindgren S, Morena Fde L, Moum B, Vavricka SR, Schröder O, Herrmann E, Blumenstein I. Prevalence of anemia in inflammatory bowel diseases in European countries: a systematic review and individual patient data meta-analysis. *Inflamm Bowel Dis* 2014;20(5):936-45.

38. Rogler G, Vavricka S. Anemia in inflammatory bowel disease: an under-estimated problem? *Front Med (Lausanne)* 2015;19:1-58.
39. Ripollés Piquer B, Nazih H, Bourreille A, Segain JP, Huvelin JM, Galmiche JP, Bard JM. Altered lipid, apolipoprotein, and lipoprotein profiles in inflammatory bowel disease: consequences on the cholesterol efflux capacity of serum using Fu5AH cell system. *Metabolism* 2006;55(7):980-8.
40. Elia PP, Tolentino YF, Bernardazzi C, de Souza HS. The role of innate immunity receptors in the pathogenesis of inflammatory bowel disease. *Mediators Inflamm* 2015;2015:936193.
41. Andrade ME, Araújo RS, de Barros PA, et al. The role of immunomodulators on intestinal barrier homeostasis in experimental models. *Clin Nutr* 2015;34:1080-1087.
42. Ito Z, Uchiyama K, Odahara S, Takami S, Saito K, et al. Fatty Acids as Useful Serological Markers for Crohn's Disease. *Dig Dis*. 2018;36(3):209-217.
43. Schwanke RC, Marcon R, Bento AF, Calixto JB. EPA- and DHA-derived resolvins' actions in inflammatory bowel disease, *European Journal of Pharmacology* 2016;785:156–164.
44. Vernia P, Loizos P, Di Giuseppantonio I, Amore B, Chiappini A, Cannizzaro S. Dietary calcium intake in patients with inflammatory bowel disease. *J Crohns Colitis*. 2014;8(4):312-7.
45. Bernstein AM, Roizen MF, Martinez L. Purified palmitoleic acid for the reduction of high-sensitivity C-reactive protein and serum lipids: a double-blinded, randomized, placebo controlled study. *J Clin Lipidol*. 2014;8(6):612-7.

Tables

	CD (n=15)	UC (n=13)
Sex ratio (M/F)	7/8	8/5
Age at inclusion, yrs, mean ± SD (range)	47 ± 11.0 (30-65)	56 ± 10.1 (39-75)
Location CD, n (%)		
Ileal (L1)	8 (53)	-
Colonic (L2)	1 (7)	-
Ileocolonic (L3)	6 (40)	-
Upper GI (L4)	0 (0)	-
Extent UC, n (%)		
Ulcerative proctitis (E1)	-	5 (38)

Distal colitis (E2)	-	5 (38)
Extensive colitis (E3)	-	3 (23)
Disease behavior CD, n (%)		
Non-stricturing and non-penetrating (B1)	2 (13)	-
Stricturing (B2)	8 (53)	-
Penetrating (B3)	5 (33)	-
Perianal disease (PD)	3 (20)	-
Disease activity indexes, n (%)		
Low activity	12 (80)	8 (62)
Moderate activity	3 (20)	5 (38)
Therapy, n (%)		
No therapy	11 (73)	6 (46)
Mesalazine	3 (20)	4 (31)
Topical steroids	1 (7)	3 (23)

Table 1: Demographical and Main Characteristics of the Study Patients.

Fold Change			Fatty Acid Name	Name	Description
IBD	CD	UC			
0.34	0.30	-	5,8,11,14,17-Eicosapentaenoic acid	Eicosapentaenoic acid	PUFA Omega-3, 20:5
0.39	0.31	-	4,7,10,13,16,19-Docosahexaenoic acid	Docosahexaenoic acid	PUFA Omega-3, 22:6
0.38	-	-	9,12-Octadecadienoic acid	Linoleic acid	PUFA Omega-6, 18:2
0.43	-	0.32	Tridecanoic acid	Tridecylic acid	Saturated fatty acid 13:0
-	0.32	-	Octanoic acid	Caprylic acid	Saturated fatty acid 8:0
-	0.45	-	Docosanoic acid	Behenic acid	Saturated fatty acid 21:0
-	0.44	-	Tetracosanoic acid	Lignoceric acid	Saturated fatty acid 24:0
-	0.54	-	Eicosanoic acid	Arachidic acid	Saturated fatty acid 20:0

-	0.24	-	Tricosanoic acid	Tricosylicacid	Saturated fatty acid 23:0
-	0.48	-	5.8.11.14- Eicosatetraenoic acid	Arachidonic acid	PUFA Omega-6, 20:4
-	0.36	-	cis-11.14-Eicosadienoic acid	Icosadienoic acid	PUFA Omega-6, 20:2
-	0.42	-	15-Tetracosenoic acid	Nervonic acid	MUFA Omega-9, 24:1
-	0.55	-	13-Docosenoic acid	Paulinic acid	MUFA Omega-7, 20:1
-	-	9.23	Octanedioic acid	Suberic acid	Saturated dicarboxylic acid 8:0

Table 2: differentially regulated fatty acids in UC, CD and IBD (UC+CD) patients compared against healthy subjects (fold change ≥ 1.5 and p-value ≤ 0.05). Results indicated four down-regulated compounds in IBD plasma samples: eicosapentaenoic acid (EPA), docosahexaenoic acid (DHA) and linoleic acid (LA), while eleven and two fatty acids were modulated for CD and UC, respectively.

Disease	Biomarker	AUC	SE	SP	Cutoff
UC	Tridecanoic acid	0.67	0.923	0.5	0.285
	LAMP2_HUMAN	0.758	0.923	0.643	0.257
	PTGDS_HUMAN	0.703	0.692	0.786	0.32
	SURF4_HUMAN	0.794	0.615	0.929	0.495
	LAMP2_HUMAN, PTGDS_HUMAN, SURF4_HUMAN, Tridecanoic acid	0.929	0.846	0.889	0.49
CD	13-Docosenoic acid	0.72	0.786	0.731	0.398
	4,7,10,13,16,19-Docosahexaenoic acid	0.74	0.786	0.769	0.413
	CBG_HUMAN	0.779	0.667	0.923	0.468

	LITD1_HUMAN	0.818	0.933	0.615	0.269
	KPCA_HUMAN	0.81	0.733	0.808	0.41
	CBG_HUMAN, KPCA_HUMAN, LITD1_HUMAN,13-Docosenoic acid,- 4,7,10,13,16,19-Docosahexaenoic acid	0.97	1	0.846	0.153

Table 3: Area Under the Curve (AUC), sensitivity (SE), specificity (SP) and optimal cut off for individual and combo markers for UC and CD patients.

			Fit			Cross-validation
		n° LV	Acc%	NER%	Acc%	NER%
UC vs. Healthy	With FA	2	100	100	95.16	95.13
	Without FA	1	96.15	96.15	94.50	94.52
CD vs. Healthy	With FA	2	100	100	99.70	99.71
	Without FA	2	100	100	97.26	97.34

Table 4: performance classification of PLS-DA models calculated from the significant variables selected according to the monivariate approach, either including fatty acids (with FA) or excluding them (without FA). The table reports the number of latent variables included

in the models (n° LV), accuracy % (Acc%) and non-error-rate % (NER%) calculated by both fit and cross-validation tests (5 cancellation groups with a random selection of the samples to be included in each cancellation group for tests repeated 1,000 times).

						Fit		Cross-validation	
		n° LVs	n° prot.	n° FA	% overlapping variables	Acc.%	NER%	Acc.%	NER%
CD vs. Healthy	BE-PLS-DA	3	90	12	70% (37 over 53)	100	100	100	100
	Monov.	2	42	11		100	100	99.70	99.71
CD vs. all	BE-PLS-DA	3	91	6	73% (19 over 26)	100	100	97.44	97.17
	Monov.	1	26	-		90.00	90.66	85.93	87.50
UC vs. Healthy	BE-PLS-DA	1	75	6	72% (31 over 43)	100	100	100	100
	Monov.	2	42	1		100	100	95.16	95.13
UC vs. all	BE-PLS-DA	3	100	6	69% (20 over 29)	100	100	98.90	99.17
	Monov.	1	29	-		82.50	85.04	81.60	84.11

Table 5: performance classification of BE-PLS-DA models and of PLS-DA models calculated including all proteins and fatty acids that were statistically significant, according to the monovariate approach, for each of the comparisons investigated separately. The results of cross-validation tests repeated 1,000 times are presented for 5 cancellation groups with a random selection of the samples to be included in each cancellation group. The table reports the number of latent variables (n° LVs), the number of proteins and fatty acids included in the models (n° prot., n° FA), the % of variables selected by the monovariate approach, which are also selected by BE-PLS-DA (% overlapping variables), besides accuracy (Acc%) and non-error-rate (NER%), by both fit and cross-validation tests.

Figure captions

Fig. 1: Experimental workflow of the study. Twelve microliters of plasma were analyzed using LC-MS to obtain a proteomic profile of plasma proteins, while 30 microliters of sample were used to obtain the plasma lipid profile through GC-MS analysis.

Fig. 2: (a) over-expressed and under-expressed proteins in IBD patients (Log FC ≥ 0.176 or Log FC ≤ 0.13 ; p-value < 0.05); (b) STRING network analysis of regulated proteins; (c) FUNRICH Gene Ontology classification of up and down-regulated proteins based on biological process.

Fig. 3: Log fold changes in regulated proteins for IBD, CD and UC patients (FC ≥ 1.50 or FC ≤ 0.667 ; p-value < 0.05).

Fig. 4: Western blot and ELISA validation. Box-plot of ZA2G protein abundance in healthy subjects (CTRL), Crohn's Disease (CD) and Ulcerative Colitis (UC) patients (a) and western blot validation (b). Box-plot of KAIN protein abundance in healthy subjects (CTRL), Crohn's Disease (CD) and Ulcerative Colitis (UC) patients (c), and ELISA validation (d).

Fig. 5: score plot of significant LVs calculated for all the comparisons: CD vs CTRL (a); CD vs ALL (b); UC vs CTRL (c); UC vs. ALL (d); UC vs CD (e). The samples are clearly separated into the two classes in all four cases.

Figure 1

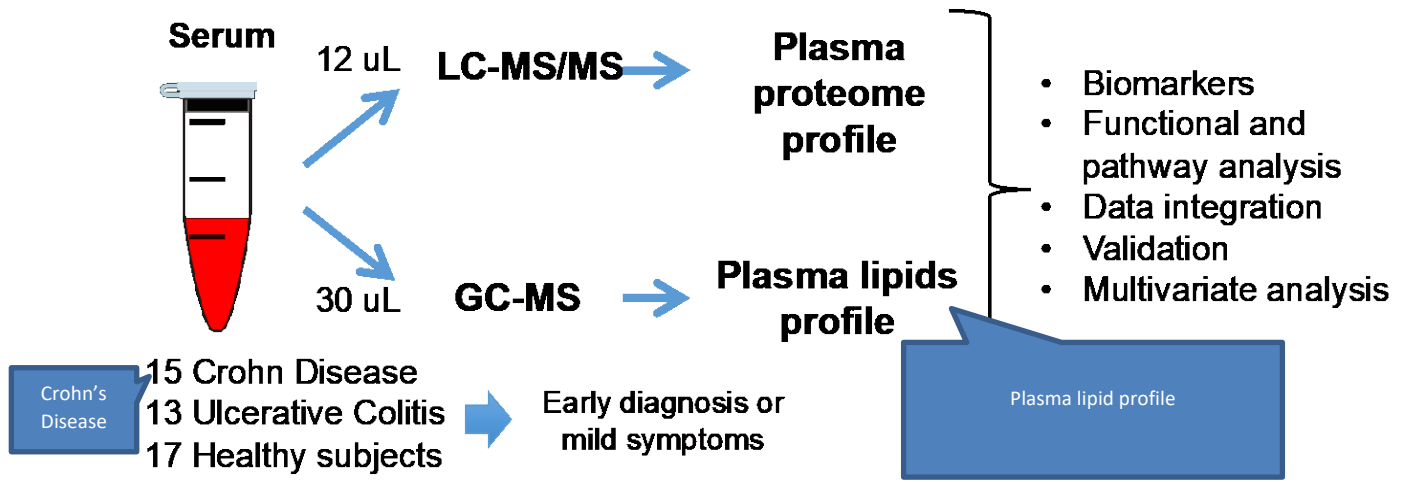


Figure 2

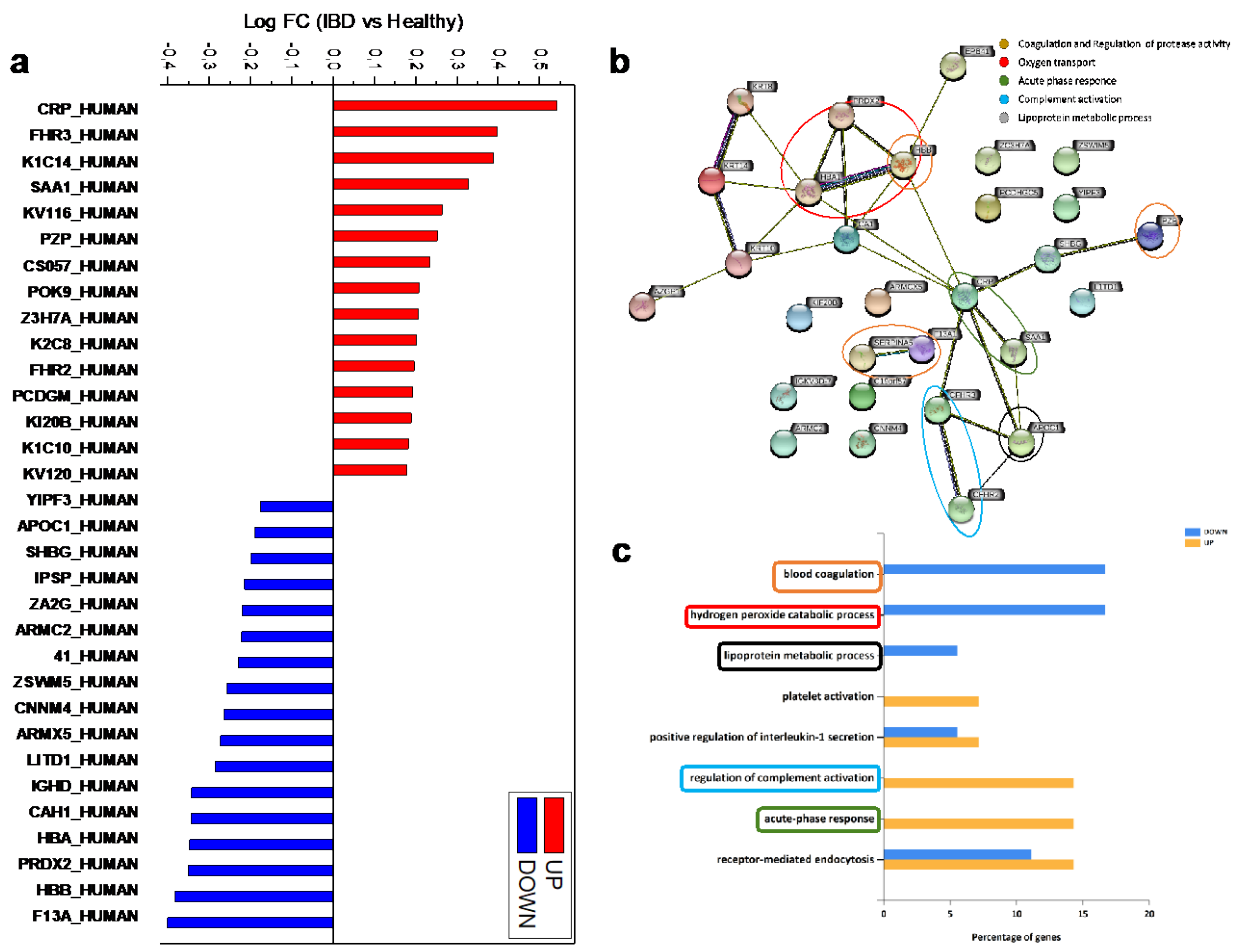


Figure 3

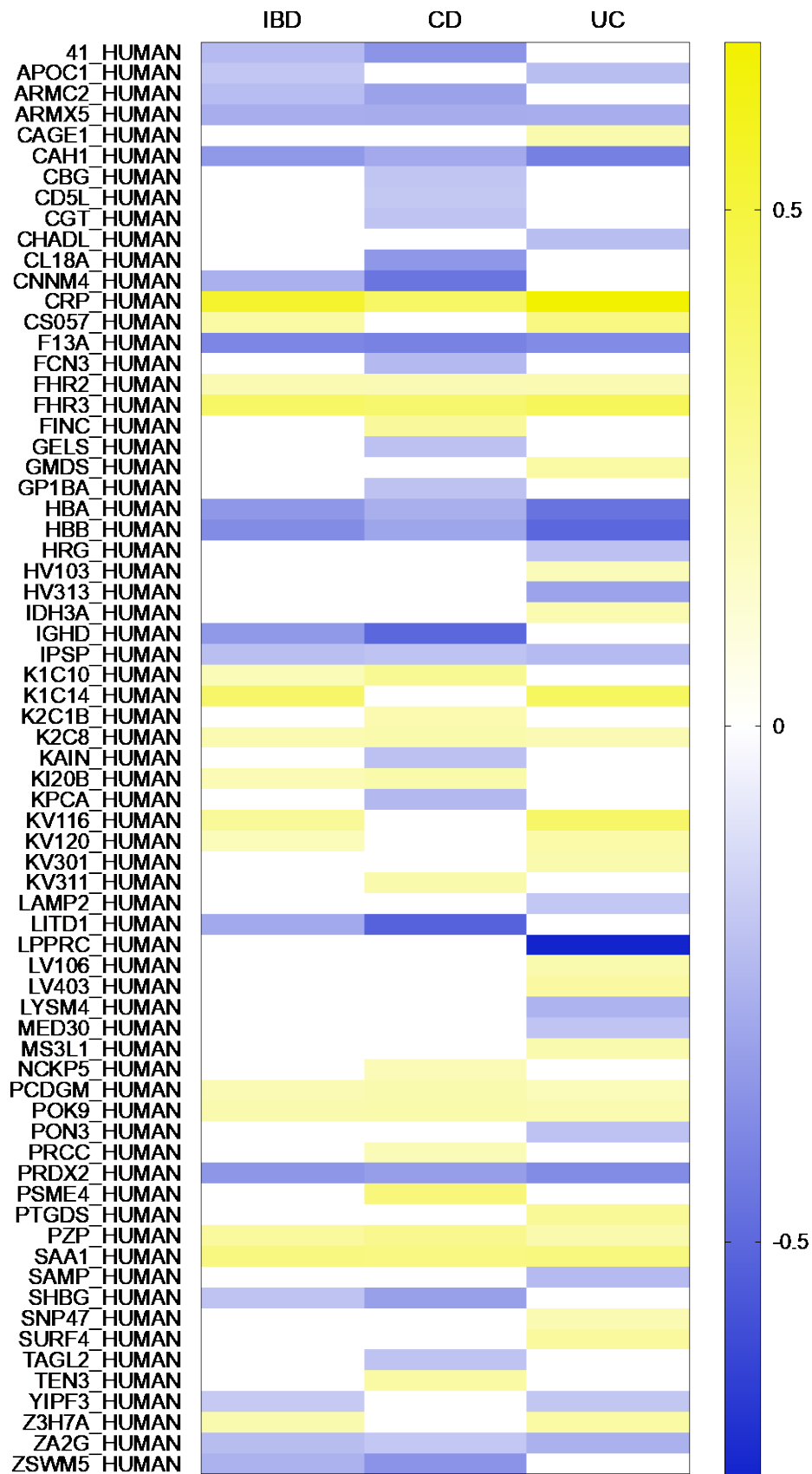


Figure 4

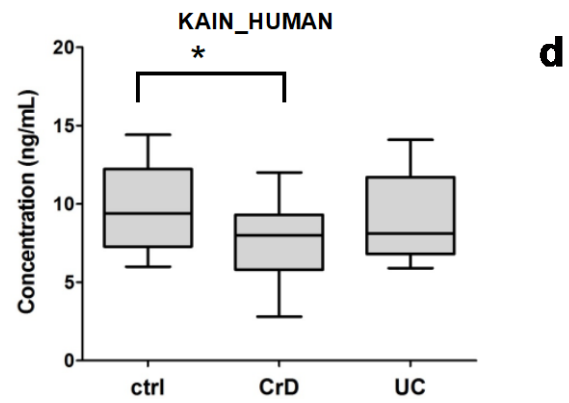
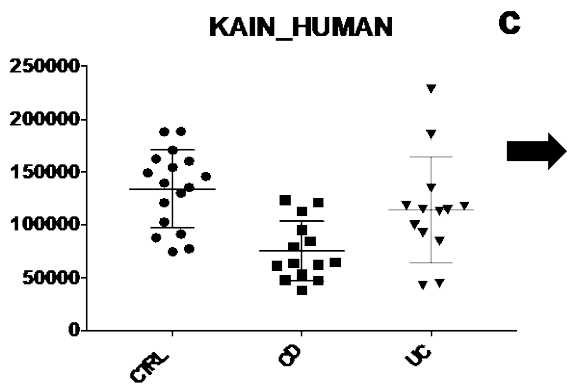
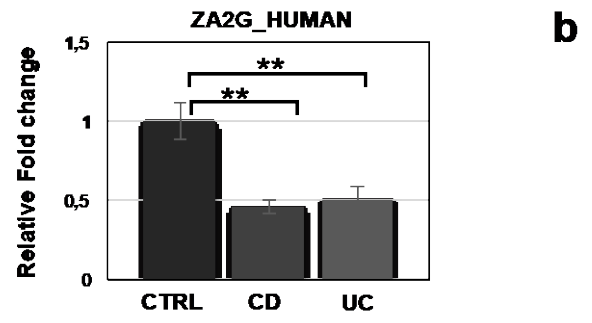
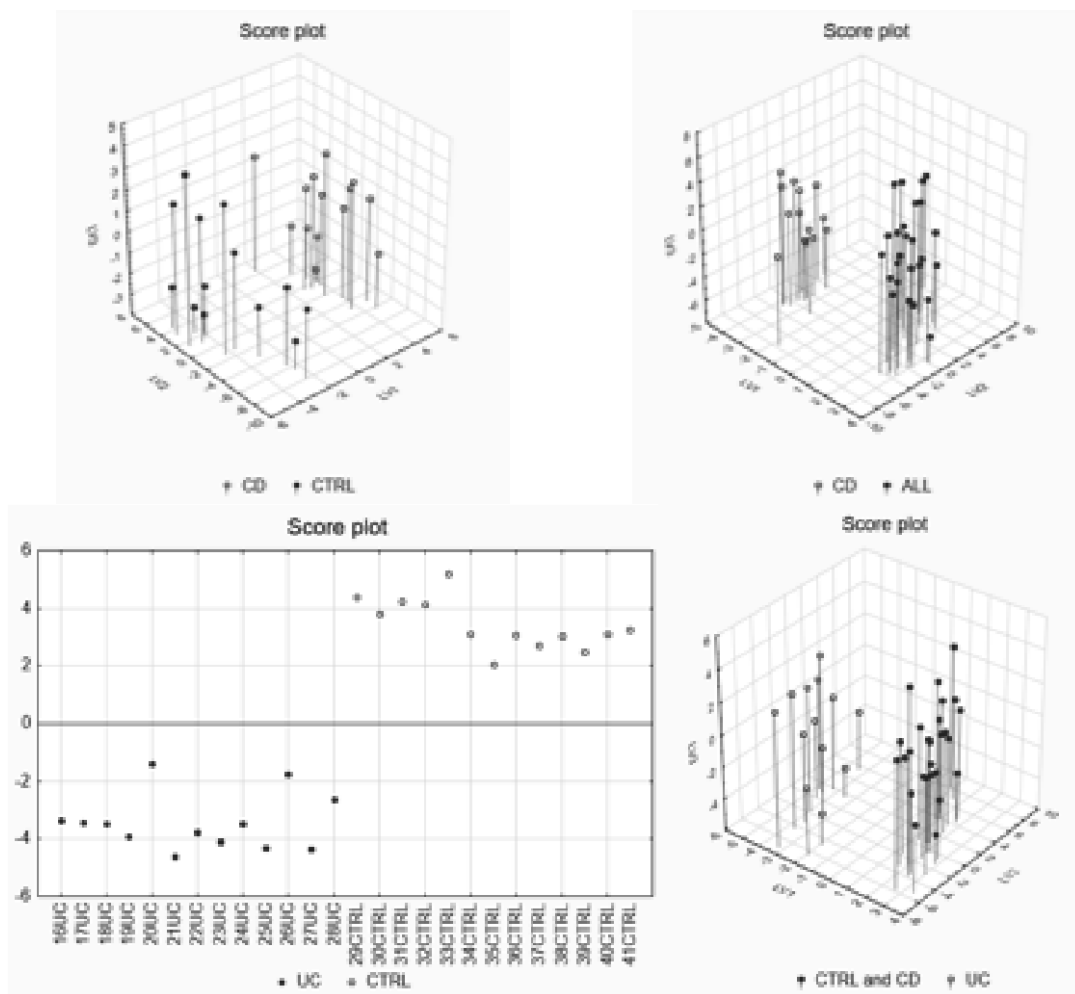
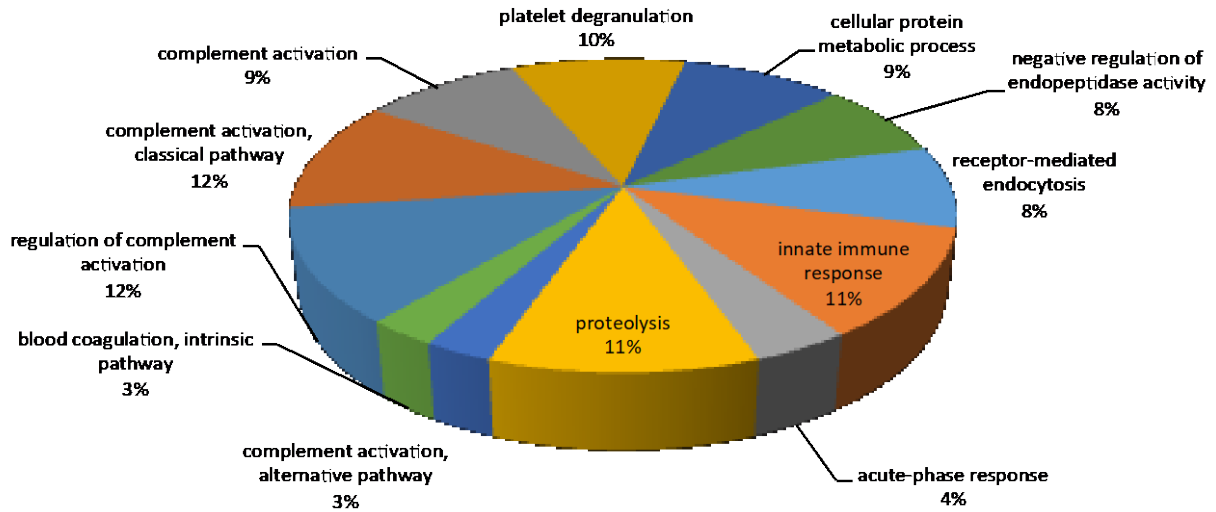


Figure 5



Supplementary figure 1

Biological process



Molecular function

