

## Review Article

# A Metabolomic Perspective on Coeliac Disease

**Antonio Calabrò,<sup>1,2</sup> Ewa Gralka,<sup>3</sup> Claudio Luchinat,<sup>3,4</sup>  
Edoardo Saccenti,<sup>5,6</sup> and Leonardo Tenori<sup>3,7</sup>**

<sup>1</sup> Department of Experimental and Clinical Biomedical Sciences, University of Florence, Viale Pieraccini 6, 50139 Florence, Italy

<sup>2</sup> Tuscany Referral Center for Adult Coeliac Disease, Viale Pieraccini 6, 50139 Florence, Italy

<sup>3</sup> Magnetic Resonance Center (CERM), University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

<sup>4</sup> Department of Chemistry, University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

<sup>5</sup> Biosystems Data Analysis Group, University of Amsterdam, 904 Science Park, 1098 XH Amsterdam, The Netherlands

<sup>6</sup> Laboratory of Systems and Synthetic Biology, Wageningen University and Research Center, Dreijenplein 10, 6703 HB Wageningen, The Netherlands

<sup>7</sup> FiorGen Foundation, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy

Correspondence should be addressed to Claudio Luchinat; [claudioluchinat@cerm.unifi.it](mailto:claudioluchinat@cerm.unifi.it)

Received 8 August 2013; Revised 30 September 2013; Accepted 20 October 2013; Published 9 February 2014

Academic Editor: Kalle Kurppa

Copyright © 2014 Antonio Calabrò et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Metabolomics is an “omic” science that is now emerging with the purpose of elaborating a comprehensive analysis of the metabolome, which is the complete set of metabolites (i.e., small molecules intermediates) in an organism, tissue, cell, or biofluid. In the past decade, metabolomics has already proved to be useful for the characterization of several pathological conditions and offers promises as a clinical tool. A metabolomics investigation of coeliac disease (CD) revealed that a metabolic fingerprint for CD can be defined, which accounts for three different but complementary components: malabsorption, energy metabolism, and alterations in gut microflora and/or intestinal permeability. In this review, we will discuss the major advancements in metabolomics of CD, in particular with respect to the role of gut microbiome and energy metabolism.

## 1. Introduction

Coeliac disease (CD, MIM 212750), first described in 1887, is a common complex chronic immune-mediated disorder with a known (gluten) environmental trigger. Recent surveys indicate that it may affect 1 in 105 subjects in the United States [1], 1 in 67 Finnish school children [2], and 1 in 230 in Italian school age children [3], with seroprevalence of about 1% in subjects of white European origin [4, 5].

Coeliac disease has a strong genetic component with multiple contributing genes: the most important and best characterized genetic risk factors are the HLA class II genes DQ2 and/or DQ8 which are located on chromosome 6p21. More than 97% of patients have at least one of the two genes: most patients (>90%) carry the DQ2 gene, while the rest expresses the DQ8 gene. HLA-DQ2 is encoded by the

HLA-DQA1\*05 allele ( $\alpha$  chain) and the HLA-DQB1\*02 ( $\beta$  chain) [6, 7]. Common to many other autoimmune disorders, the two alleles are often present in the *cis* conformation on the DR3 haplotype [8]; HLA-DQ2 and HLA-DQ8 are necessary but not sufficient for the development of CD.

Genome wide association studies indicated 39 non-HLA loci to be predisposing to CD [9–11]. Altogether, the nonHLA loci explain only 5% of the risk for CD [6], while the HLA loci account for 35% of the risk [7].

Several of these nonHLA CD susceptibility genes are associated with other diseases/traits [6] such as type 1 diabetes [12, 13], rheumatoid arthritis [14], and systemic lupus erythematosus [15] indicating a possibly shared genetic background with other diseases [7].

The environmental trigger of coeliac disease is gluten, a protein complex formed by gliadin and glutenin, which is

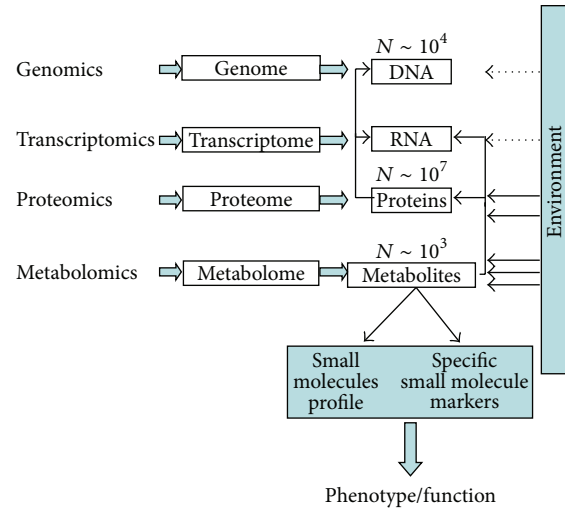


FIGURE 1: Relationships between the omics sciences.

found in wheat and related grain species like barley and rye and also in processed food where it is used to enhance food texture and as a stabilizing agent.

The active disease component is gliadin [16] with the  $\alpha$ -,  $\beta$ -,  $\gamma$ -, and  $\omega$ -fractions. These fractions are rich in proline and glutamine and resistant to enzymatic digestion; large proline/glutamine-rich peptides accumulate in the smallest intestine, triggering an abnormal innate and adaptive immune response in genetically predisposed subjects.

The response of the adaptive immune system is due to the gliadin-reactive CD4+ T cells; HLA-DQ molecules bind to these peptides which are deamidated by the intestinal brush border enzyme tissue transglutaminase; these complexes interact with the T-cell receptor on T cells leading to T-cell activation with subsequent release of proinflammatory cytokines and the production by B-cells of specific antibodies (anti-tissue transglutaminase and endomysial antibodies) [17, 18].

The role of the innate immune systems in CD is less clear [7]. Increased expression of interleukin-15 has been observed [19]; enhanced intestinal permeability has been also observed [20], induced by zonulin [21], whose release is mediated by gluten activated CXCR-3 [22, 23]. Loss of functionality in the intestinal barrier permits the passage of immunoreactive peptides and other antigens from the gut lumen to the lamina propria, with subsequent triggering of the innate immune system.

## 2. Metabolomics

The advent of high-throughput techniques led to a rapid expansion of data sets originated from the analysis of gut microbiota and currently several ongoing projects are aimed at the study and definition of the microbiome [24, 25]. In this framework, metabolomics is playing a crucial role.

Since the systematic genome sequencing of the first free-living microbe [26], we have seen the rising of genome-wide expression profiling methods, aimed to understand complex biological systems on a large scale [27]. The fast

development of genomics, transcriptomics, proteomics, and the other *omics* disciplines is the consequence of this new scientific paradigm.

In this framework, metabolomics has already proved in the past decade to be a useful complement for the characterization of several physiological and pathological conditions and offers promises as a clinical tool [28]. Metabolomics is based on the analysis of the measured dynamic changes of a living organism in response to genetic modifications or physiological stimuli such as nutrients, drugs treatment, or toxic insults [29].

The metabolome, the complete collection of all metabolites contained in a biological specimen, can be considered the downstream end-product of the complex interaction of genome, transcriptome, proteome, and the environment: it can be regarded as a cascade linking genome to the phenotype [30] (Figure 1). The metabolome, consisting of low-molecular weight chemical intermediates [31], can be considered as an amplified version of gene expression. While changes in gene expression levels (and thus proteins) will have only small effects on metabolic fluxes, they must have large effects on metabolic pathways (and thus metabolites concentrations) [27]. From this point of view, the metabolite space represents the optimal level at which changes in biological systems are analyzed with optimal sensitivity [32] under conditions of negligible effects on the global phenotype [33].

Metabolomics does not rely on the measurement of a single metabolite but considers the spectrum of (possibly) all metabolites as a whole, taking a holistic approach; this offers evident advantages with respect to a targeted search of metabolites; indeed, no assumption is required on the identity of the metabolites that are or may be relevant for the biological phenomenon under investigation.

The main analytical techniques employed in metabolomics (Boxes 1 and 2) are nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS). Both MS and NMR methods provide information on the relative and absolute concentrations of different classes of metabolites in a single measurement (see Box 3) and can be also used

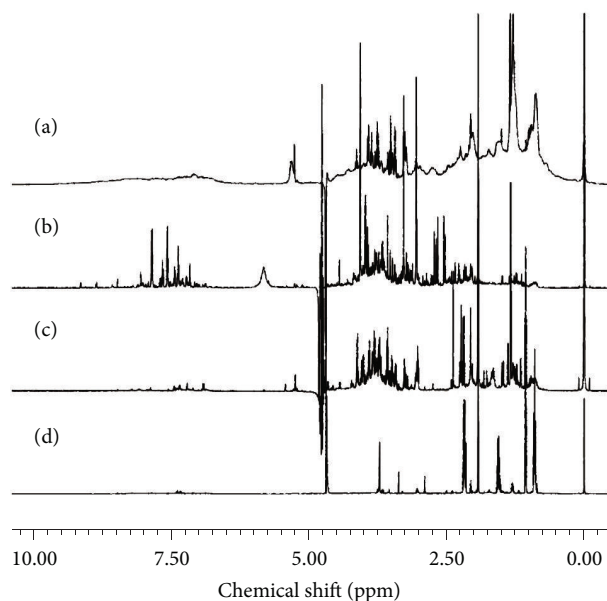


FIGURE 2: Examples of NMR profiles of (a) serum, (b) urine, (c) saliva, and (d) faecal extract.

to determine metabolite structures providing mechanistic insights.

The most common biological specimens used in metabolomics are serum/plasma and/or urines, firstly, because they can be collected with low invasiveness, and, secondly because, as they contain thousands of metabolites, they are rich in biological information at the systemic level; a number of other biofluids such as saliva [34], tissue extract [35], cerebrospinal fluid [36], bile [37], seminal fluid [38], amniotic fluid [39], synovial fluid [40], exhaled breath condensate [41, 42], and faecal extracts [43] can also be studied. Figure 2 shows typical NMR spectra of four different biofluids (Box 3).

Targeted and untargeted approaches are possible in metabolomics, the former focusing on the analysis of a subset of known compounds or class thereof (targeted MS and NMR) and the latter focusing on the whole array of metabolites within the detection limit of the technique employed (untargeted MS and NMR). Using both approaches, hundreds to thousands of metabolites are measured. Data are usually analyzed following the classical metabolomics pipeline (Box 4), and information is extracted using state-of-the-art statistical tools (Box 5).

Metabolomics has provided significant information on a wide range of pathologies, such as cancer [44], meningitis [45], neurological disorders [46], cardiovascular diseases [47], inborn errors of metabolism [48], and CD [49–51]. The first metabolomics investigation of CD revealed that a metabolic fingerprint for coeliac disease can be defined [49], which accounts for three different but complementary components: malabsorption, energy metabolism, and alterations in gut microflora and/or intestinal permeability.

In this review, we will discuss the major advancements in metabolomics of CD with respect to the role of gut microbiome and energy metabolism.

### 3. Gut Microbiota and CD

Recent studies [52–56] pointed to the possible role of intestinal microbiota (faecal and duodenal species) in the development of coeliac disease. A summary of the most relevant findings in this research area is reported in Table 1 together with the associated bacteria strains involved.

Nistal et al. [57] compared the differences between gut microorganisms in the upper small intestinal mucosa in adults and in children. A two-time higher number of microbial genera have been identified in adults compared to children, although the dominant genera were very similar: *Firmicutes*, *Proteobacteria*, *Bacteroidetes*, *Actinobacteria*, and *Fusobacteria*. Differences in the amount of bacterial communities between adult and juvenile groups can be also directly connected with the age of investigated subjects.

A comparison between healthy, diseased, and treated coeliac adults showed a lower number of *Streptococcus* spp. and *Prevotella* spp. families in untreated coeliac adults. Interestingly, similar patterns were also observed in CD children, suggesting that these bacterial populations may have been modified by changes in the intestine environment caused by active CD.

In the study by Nadal et al. [58], the bacterial species present in faeces and duodenum of children with active and treated CD were compared with a healthy control group. The ratio of harmless Gram-positive bacteria (*Lactobacillus* + *Bifidobacterium*) to potentially harmful Gram-negative (*Bacteroides/Prevotella* + *E. coli*) bacteria was significantly lower in CD patients than in controls, while no distinction was possible between active and inactive CD.

Sánchez et al. [59] applied denaturing gradient gel electrophoresis (DGGE) to analyze intestinal microbiota from biopsy specimens obtained from three groups of children, investigating the composition of *Bacteroides*, *Bifidobacterium*, and Lactic acid bacteria in duodenal biopsies of patients with active and treated coeliac disease. Dysbiosis in *Bacteroides* (the most abundant intestinal bacterial group) was observed, with a significant reduction in coeliac and coeliac treated patients in comparison with the control group. Moreover, it was observed that a treatment with gluten-free diet did not restore the balance of the *Bacteroides* composition. Interestingly, it was observed that the lactic acid bacteria (*Lactobacillus*) and the *Weissella* family were more abundant and diverse in treated coeliac and control patients than in patients with untreated coeliac disease. The authors suggested that some of the changes in duodenal bacterial community could be due to the inflammatory consequences of the active phase of the disease; nonetheless, the influence of different dietary habits could not be discarded.

Biopsies from treated coeliac children were analyzed in a study by di Cagno et al. [63], that also confirmed that a gluten-free diet lasting two or more years is not able to restore completely the microbiota. In addition, a higher diversity of the *Eubacteria* community was observed in the duodenum

TABLE 1: Most relevant findings, and associated references, for studies linking gut microbiota and CD.

References	Type of sample	Technique	Microbiota phylum/class	Relevant findings
Wacklin et al. (2013) [55]	Mucosa biopsy	PCR-DGGE (real-time polymerase chain reaction, denaturing gradient gel electrophoresis), 16S rRNA sequencing	<i>Firmicutes Bacteroides Proteobacteria Actinobacteria</i>	Diversity in mucosal microbiota of celiac disease patients is associated with the symptoms of the disease.
Nistal et al. (2012) [57]	Duodenal biopsies	PCR (polymerase chain reaction)	<i>Firmicutes Proteobacteria Bacteroidetes Actinobacteria Fusobacteria</i>	Composition of small intestinal microbiota is similar between adults and children; there is higher number of <i>Streptococcus</i> and <i>Prevotella</i> in healthy subjects.
Nadal et al. (2007) [58]	Duodenal biopsy	FISH (Fluorescent in situ hybridization), Flow cytometry detection.		In faeces and duodenum of CD children, smaller amount of harmless bacteria ( <i>Lactobacillus</i> and <i>Bifidobacterium</i> ) and higher number of harmful bacteria are found ( <i>Bacteroides/Prevotella</i> + <i>E. coli</i> ) compared to healthy children.
Sánchez et al. (2010) [59]	Duodenal biopsy	PCR-DGGE	<i>Bacteroidetes</i>	Reduced number of intestinal microbiota in CD children but also in treated CD children was noticed. Treatment with GFD does not restore the bacteria composition.
Sánchez et al. (2011) [60]	Faeces samples	PCR-DGGE		Studies were carried out on stools of infants with high/low risk of CD and different types of milk feeding. High-risk infants have higher prevalence of <i>Bacteroides vulgatus</i> , whereas low-risk infants have higher population of <i>B. uniformis</i> , <i>B. ovatus</i> , and <i>B. plebeius</i> considering the subgroup of either breast-fed or formula-fed infants.
Cheng et al. (2013) [61]	Duodenal biopsy	qRT-PCR (quantitative real-time PCR)	<i>Bacilli Bacteroides Clostridium Proteobacteria</i>	Overall microbiota composition in the duodenal mucosa is comparable between healthy and CD children, but studied groups differ regarding bacteria subpopulation profile.
Sellitto et al. (2012) [51]	Faeces samples	qPCR (quantitative PCR)	<i>Bacteroidetes Firmicutes</i>	Lack of microflora maturation during first 2 years of life in infants at risk of CD. Moreover, there was observed absence of <i>Bacteroidetes</i> and abundance of <i>Firmicutes</i> .
Sanz et al. (2007) [54]	Faeces samples	PCR-DGGE	<i>Actinobacteria Firmicutes</i>	<i>Lactobacillus</i> and <i>Weissella</i> are more abundant and diverse in treated CD patients and control subjects than in active CD individuals. Composition of lactic bacteria and <i>Bifidobacterium</i> differs between celiac children and age-matched controls.
Kaufman and Rousseeuw (2009) [62]	Intestine biopsies	PCR	<i>Proteobacteria</i>	There observed no statistical differences in bacteria composition between healthy and CD children. Nevertheless, <i>Haemophilus</i> was more common in CD patients and <i>Neisseria polysaccharea</i> in control individuals.
di Cagno et al. (2011) [63]	Faeces sample, duodenal biopsy	RAPD (random amplification of polymorphic DNA) -PCR	<i>Eubacteria</i>	Higher number of different <i>Eubacteria</i> classes was found in duodenum of coeliac children under gluten-free diet than in healthy children.

TABLE I: Continued.

References	Type of sample	Technique	Microbiota phylum/class	Relevant findings
Medina et al. (2008) [64]	Faeces sample	PBMC (peripheral blood mononuclear cell) phenotyping and flow cytometric analyses	<i>Actinobacteria</i>	Studies regarding interaction between faecal bacteria and immune system response of coeliac disease patients. It appeared that Gram-positive bacteria such as <i>Lactobacillus</i> and <i>Bifidobacterium</i> may act as inhibitors of inflammation.

of coeliac children under a gluten-free diet with respect to healthy children. Compared to that of duodenal biopsies, the faecal bacterial population was found to be more diverse. PCR-DGGE faecal profiles of *Lactobacillus* and *Bifidobacterium* differ between treated coeliac children and healthy controls. The ratio between *Lactobacillus/Bifidobacterium* and *Bacteroides/Enterobacteria* was lower in coeliac children under treatment compared to healthy children. Some of the differences could be related to both coeliac disease and dietary variations.

Surprisingly, a study by Ou et al. [65], based on biopsies collected from the distal duodenum/proximal jejunum of 45 children, did not reveal significant differences between the microbiota in the small intestine of diseased and healthy children, although bacteria from the *Haemophilus* family were more abundant in CD patients, while *Neisseria polysaccharea* were more widespread in the control group. However, at the *genus* level, no differences between the two groups were observed. The authors commented that differences at the species level could not be excluded because complete 16 S rDNA were not sequenced. Similar results, pointing to a lack of significant differences in global composition of duodenal microbiota between healthy controls and CD patients, were obtained also by Cheng et al.; on the other hand, a subpopulation profile, containing eight genus-like bacterial groups, was found to distinguish healthy controls from CD patients [61].

The possible effect on microbiota of different types of milk feeding in early life and the link to the risk of CD development were investigated by Sánchez et al. [60]. The study was carried out on stools of breast-fed infants with different genetic risk of CD (low and high); it showed that high-risk infants had a higher prevalence of *Bacteroides vulgatus*, whereas low-risk infants had higher population of *B. uniformis*, *B. ovatus*, and *B. plebeius*. In the study group of formula-fed infants, *B. ovatus* and *B. plebeius* were increased in subjects with lower genetic risk, while *B. vulgatus* had higher prevalence in those subjects with higher genetic risk. The authors concluded that both types of milk feeding in conjunction with HLA-DQ genotype can influence the *Bacteroides* colonization, increasing the risk of coeliac disease onset. Also, the time of exposure to milk feeding was found to be relevant in prompting coeliac disease development [66].

Sellitto et al. [51] reported the impact on the intestinal tract of two different patterns of gluten introduction. A delay in gluten exposure of at least 6 to 12 months was found to have a positive effect on gluten tolerance: it caused a delay

in CD autoimmunity onset in infants that were genetically susceptible to CD. Instead, the early exposure to gluten of infants at risk of coeliac disease was found to induce an immune response and led to a more frequent development of CD. Moreover, a lack of gut microflora maturation during the first 2 years of life in infants at risk of CD was also noted. The gut metabolome of the first 6 months of infant's life reflects mainly the milk diet (rich in polysaccharides and other sugars) and is very similar in all infants; once the solid food is introduced, a shift occurs and a group of short-chain fatty acids are found in faeces. By the end of the second year of life, *Bacteroides* are the main bacteria group found in the metabolome of healthy infants. Conversely, in infants with a genetic risk of celiac disease, an overall lack of bacteria of the phylum *Bacteroides* and abundance of *Firmicutes* were observed.

Recent studies [67] suggest that the colonization of gastrointestinal tract is very important in the development of autoimmune disorders and food-related disease. Furthermore, possible interaction between the intestinal bacteria and the mammalian immune system in the direct differentiation of both pro- and anti-inflammatory T-cells population has been suggested [68]. To clarify whether the gut microflora present in the faeces of CD patients is involved in the proinflammatory activity of coeliac disease, *Bifidobacterium* from healthy subjects was co-incubated together with the faecal microflora or the peripheral blood mononuclear cell culture of coeliac subjects [64]. It appeared that certain strains of *Bifidobacterium* are able to suppress and reverse the proinflammatory effect by increasing IL-10 cytokine production. These results may suggest the use of selected strains of *Bifidobacterium* as probiotics for treatment of CD.

It has been also suggested that gluten intolerance may be also triggered by environmental factors like viruses or bacteria showing molecular mimicking with gluten proteins, causing an autoimmune response that may last even after infection [69]. Several studies pointed to infections by human adenovirus [70], hepatitis C virus [71], rotaviruses [72], or *Campylobacter jejuni* [73] that could induce allergic reactions similar to that induced by gluten exposure, causing the onset of CD.

#### 4. Body Composition and Energy Expenditure in CD Patients

Patients with the classic form of coeliac disease are always characterized by weight loss directly connected with

malabsorption and subsequent risk of malnutrition. Often coeliac disease results in a general lack of energy and strength that can create abnormal conditions described as (chronic) fatigue. Appearance of fatigue and fatigue-related problems seems to be more frequent in nontreated coeliac patients than in patients on a gluten-free diet [74].

Body composition, resting metabolic rate (RMR), and substrate oxidation rates were investigated in [75, 76]. The results showed that untreated and treated CD patients had a lower body weight, lower levels of fat-free mass (FFM), and lower fat mass (FM) in comparison to the healthy controls. In [77], the analysis of body composition at the diagnosis time and after one year of treatment with a gluten-free diet was carried out. The analysis showed a significant increase of body weight and FM but only a slight increase of FFM after treatment with gluten-free diet. Additionally, RMR values were higher in CD patients (treated and untreated) than in controls. Moreover, untreated CD patients showed a higher  $\text{npRQ}$  (nonprotein respiratory quotient); this may indicate that untreated patients oxidize larger amounts of carbohydrate under resting metabolite conditions than treated CD and healthy subjects.

Interestingly ghrelin, one of the hormones responsible for energy balance regulation, is also changed in CD patients. Ghrelin is a 28-amino acid-peptide produced by the enteroendocrine cells of the gastric mucosa and the intestine [78]. Recent studies have shown that ghrelin is able to increase food intake, decrease fat use, and reduce energy expenditure [79]. While serum ghrelin concentration was increased in CD patients, body mass was decreased [80, 81]. Lower levels of circulating ghrelin were found in CD patients after gluten-free treatment in comparison with CD and control subjects [82]. These results suggest that low amounts of ghrelin in the blood may be partially responsible for the slight increase in body weight and FM in CD patients after treatment with a gluten-free diet.

## 5. Metabolomic Signature of CD

In many cases, the diagnosis of CD is not an easy task, mainly because CD has a variable clinical picture due to its intertwined genetic, immunological, and environmental components. The presence of the HLA genetic factor, together with a positive biopsy and serological antibodies upon gluten-containing diet, is used to diagnose coeliac disease at any age. In order to better understand the processes underlying the activation and development of coeliac disease, it is important to examine the mechanisms from the early beginning.

To date, a limited number of metabolomics studies of coeliac disease are available, but they clearly show that metabolic differences between healthy individuals and coeliac patients exist. In the first (to our knowledge) metabolomic study on CD, Bertini et al. [49], examined adult healthy controls and coeliac patients by  $^1\text{H}$  NMR profiling of their serum and urine profiles before and after GFD, showing that a metabolic fingerprint for CD can be defined. This fingerprint was found to be made up by three components, one related to malabsorption, one related to energy metabolism, and the third related to alterations in gut microflora

and/or intestinal permeability. Using this metabolic fingerprint, it was possible to make predictions about the coeliac status with a very good accuracy (ca. 84%). One of the most interesting findings was that the metabolic profile of CD patients reverts to normality after 12 months of a strict gluten-free diet; interestingly, a similar behavior was not found in CD patients when analyzing them from a gut microflora prospective [58, 59, 63].

The main observed differences in serum spectra between CD patients and controls were lower levels of several amino acids (asparagine, isoleucine, methionine, proline, and valine), methylamine, pyruvate, creatinine, choline, methylglutarate, lactate, lipids, and glycoproteins and higher levels of glucose and 3-hydroxybutyric acid. Notably, the best discrimination is obtained from CPMG spectra (Carr-Purcell-Meiboom-Gill spin echo sequence) [83], that is, from spectra in which signals arising from large macromolecules such as lipidic components are suppressed [49]. So, although it is known that coeliac patients usually appear to be hypocholesterolemic, lipids do not contribute significantly to the metabolomic signature of coeliac disease. A decrease in the level of pyruvate and lactate and a higher level of glucose in the blood of coeliac patients were observed, probably as a consequence of an impaired glycolysis process. Glycolysis impairment can cause a lowering of pyruvate and lactate levels and an increase of glucose levels in blood. If this metabolic way is reduced,  $\beta$ -oxidation is probably increased. Enhanced  $\beta$ -oxidation and malabsorption can then explain lower levels of lipids in serum [49]. In these conditions, the authors hypothesized an increase of the use of ketonic bodies as a source of energy in coeliac patients, consistently with the higher observed levels of 3-hydroxybutyric acid in blood and acetoacetate in urines [49].

Energy conversion from lipids and catabolism of ketonic bodies are far less efficient than that from glucids. Untreated coeliac subjects often report symptoms of fatigue. In patients on a gluten-free diet, fatigue tends to be reduced and, in fact, it has been proposed that this condition is gluten-related [74]. In [49], the authors found that in CD patients on a gluten-free diet the levels of glucose and 3-hydroxybutyric acids revert to normality.

Further, the authors found that CD patients are characterized by higher urine levels of some metabolites related to gut microbiota: indoxyl sulfate (IS), meta-[hydroxyphenyl] propionic acid (m-HPPA), and phenylacetylglutamine (PAG). M-HPPA mostly originates from gut microflora, being one of the several products of the microbially mediated breakdown of larger plant phenolic compounds such as caffeic acid and its conjugate chlorogenic acids [84]. IS is a harmful uremic toxin produced in the liver from indole through indoxyl. Indole is a subproduct of tryptophan metabolism by intestinal bacteria [85]. Modulation of PAG excretion in urine has been attributed to gut microflora, and increases of PAG have been reported in cases of drug-induced phospholipidosis; nonetheless, the contribution of mammalian and microbial sources to PAG excretion is not yet fully characterized [86]. All these findings are consistent with the hypothesis that in CD patients the gut microflora of the small bowel is

altered or presents peculiar species with their own microbial metabolome.

In a following investigation [50], the same research group highlighted again the existence of a metabolic fingerprint for coeliac disease, confirming most of the previously discussed metabolites with the additional finding of higher levels of *p*-cresolsulfate in the urines of CD patients. Interestingly, *p*-cresolsulfate, a metabolite of bacterial origin, is associated with several gastric-related disease [87], including bowel cancer [88]. In the same study, the analysis of the so-called “potential coeliac patients” (i.e., subjects who have a positive antibody test but no evidence of intestinal damage) showed that the metabolic patterns of overt and potential coeliac patients are similar [50] indicating that CD-related dysmetabolism precedes the intestinal damage. Only a few serum metabolites differentiate between potential and overt CD, and none of these metabolites are related to the energy metabolism [50]. It appears that, as in overt CD patients, glycolysis is somehow impaired also in potential CD patients. Impairment of glycolysis explains both the observed lower lactate levels and the higher glucose levels in blood of potential CD patients. In urine, there are more metabolites that discriminate potential and overt CD. The key differences lie in the concentration of metabolites originating from the gut microflora (m-HPPA, IS, and PAG) which in potential coeliac subjects are similar to those of controls, suggesting a relationship between overt CD, villous atrophy, and bacterial consortia of the host [50].

The authors concluded that, although free from intestinal injury, placing potential CD subjects on a gluten-free diet could be justified because they are experiencing most of the pathological alterations experienced by overt coeliac patients [50]. Figure 3 shows the discrimination between overt CD patients and healthy controls and the statistical prediction of the potential CD patients: almost all potential CD patients are predicted as overt CD. The plot shown in Figure 3(a) was obtained using a training set composed of the serum CPMG spectra of 34 overt CD patients, 34 healthy controls, and 13 (out of the 34) CD patients after 12 months of gluten-free diet. It clearly appears that all but one patient on gluten-free diet were classified as healthy. Similarly, the plot in Figure 3(b) was built using the CPMG spectra of 61 overt CD patients, 51 healthy controls, and 29 potential CD patients. Almost all the potential CD patients fall in the CD group, underlining the affinity between the metabolic fingerprints of these two dissimilar clinical conditions.

Differences between the metabolic profiles of faeces and urine of CD and healthy children using a combination of  $^1\text{H-NMR}$  and GC-MS/SPME techniques were reported by di Cagno et al. [63]. The analysis allowed the identification of a group of compounds that were significantly changed in the treated coeliac children group. A set of volatile organic compounds and short fatty acids were identified using MS, whereas amino acids were identified using NMR [63]. Faecal and urine samples of treated CD children showed elevated levels of free amino acids (proline, methionine, histidine, and tryptophan) and lowered levels of some short fatty acids (butyric, isocaproic, and propanoic acids) compared to healthy children [63]. The authors suggested that these

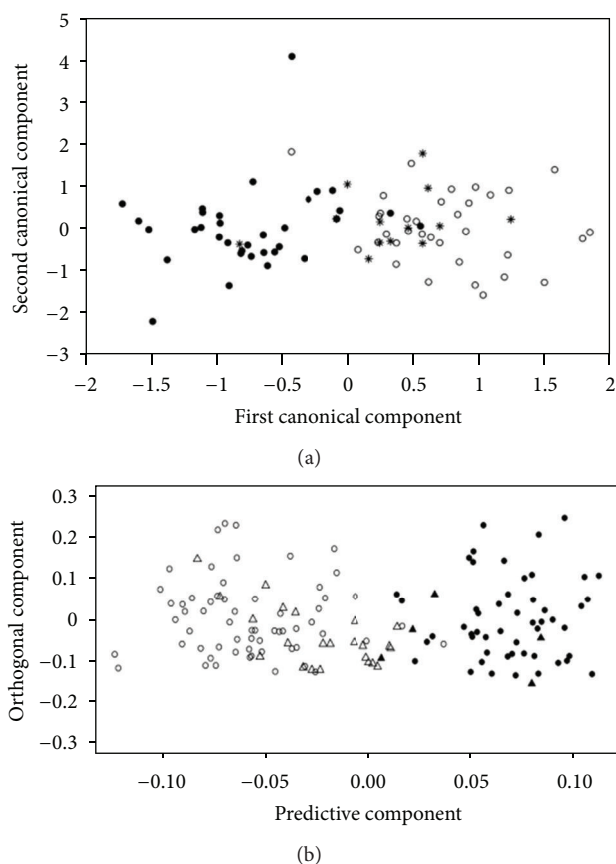


FIGURE 3: (a) Clustering of CPMG (Carr-Purcell-Meiboom-Gill spin echo sequence) [83] serum spectra of CD patients (filled circles) and controls (open circles). The discriminant model between the two groups was calculated using a combination of partial least square [89] and (regularized) canonical analysis [90] (PLS-RCC) and was validated using cross-validation. The CPMG spectra of 13 (out of the 34) CD patients after 12 months of gluten-free diet were then projected into the discriminant space of the model (stars) and were assigned to the CD or the healthy group applying a support vector machine [91] classifier (SVM). (b) Clustering of overt CD patients (open circles) and healthy controls (filled circles) obtained with CPMG serum spectra. The discriminant model was calculated using orthogonal partial least square [92] (OPLS) and validated using double cross-validation [93]. The CPMG spectra of 29 potential CD patients were then projected in the model (triangles) and filled or not according to the results of an SVM classifier. Adapted with permission from [49, 50]. Copyright (2009 and 2011) American Chemical Society.

changes may be associated with intestinal and faecal bacteria modifications that could induce a nonspecific inflammation and a reduction of the absorptive surface of the intestinal mucosa; this may lead to a reduction of the absorption of amino acids which are subsequently lost with stool [63]. By combining microbiology and metabolomics, the authors showed that a gluten-free diet lasting at least two years did not completely restore the microbiota of the CD children. From that work, a broader picture seems to emerge that microbial indices (i.e., the ratio of faecal cell density of lactic

acid *bacteria-Bifidobacterium* to *Bacteroides-Enterobacteria*) and the levels of some metabolites (i.e., ethyl-acetate, octyl-acetate, SCFA, and glutamine) are characteristic of CD patients [63].

## 6. Perspectives

Metabolomics is a rapidly growing discipline bringing together analytical technologies, metabolite pathways evaluation, and information technology. A major advantage is the noninvasive or minimally invasive measurement of potentially useful biomarkers from biofluids such as urine and plasma. A great deal of validation work (both at the analytical and data analysis level) has been carried out to gain full acceptance to metabolomics in routine clinical practice. Challenges for the development of metabolomics still exist, including simplified systems to present data to end-users (such as interpretation of often complex statistical models), the coordination of multiple data streams, and the implementation of quality control programs [94]. We expect that in the next few years it will be clear whether or not metabolomics will take its place as a complementary or even an alternative tool in the clinical setting.

At the present time, only few applications devoted to the investigation of coeliac disease have been presented in the literature, but a complex picture of the interaction between energy metabolism and gut microbiota seems to emerge, providing new hints on the biochemistry of the disease. In our institutions, as a logical complement to the results obtained analyzing overt coeliac and potential coeliac subjects, we are currently applying metabolomics to the biomolecular investigation of a gluten-related condition defined as gluten sensitivity [95]. This condition is still not very well characterized and its pathogenesis is caused by unknown mechanisms; we believe that metabolomics is a useful tool to expand our current limited knowledge of this condition.

Metabolomics-based approaches are expected to enable diagnosis, prognosis, and prediction of response of individuals to treatment. We can expect that metabolomics will provide more accurate and less expensive biomarkers (obtained by means of proper statistical analysis and properly validated) than presently available, which could improve diagnostic accuracy and sensitivity. However, far more research is essential to reach such a goal, and a validation of the results on an epidemiological scale is indeed needed.

*Box 1* (MS and metabolomics). The main analytical techniques used in metabolomics are nuclear magnetic resonance spectroscopy (NMR) and mass spectrometry (MS) [96]. Both MS and NMR methods provide information on a wide range of metabolites in a single measurement. Furthermore, both can be used to identify the metabolites' structures and to measure the relative and absolute concentrations of the molecules (MS has higher sensitivity but NMR is more reliable for determining concentrations) [97].

Mass spectrometry is a technique to determine extremely accurate mass of molecules in a pure sample or in a mixture. The molecules in a sample are converted to ions by an electron

beam; the ions are accelerated by charged plates and then deflected by a magnetic field according to the mass-to-charge ratio of each ion. When the ions reach the detector, the mass-to-charge ratio is registered to provide a spectrum where series of peaks are shown reporting the intensity of each ion generated by the sample. MS is a destructive technique but requires a very low quantity of sample [98]. Over the last few years, its application to mammalian study increased, especially for its high sensitivity, and because it is a major technique for molecular identification [99]. As opposed to NMR, MS usually requires metabolites separation before detection, typically by using gas chromatography (GC) or liquid chromatography (LC). GC-MS is a robust technique for the analysis of volatile and semivolatile compounds suitable for chemical derivatization to increase their volatility [100]. Electron ionization (EI) in GC-MS is quite reproducible [100]. In contrast to GC-MS, LC-MS is especially suitable for the analysis of nonvolatile and/or thermally unstable metabolites. The introduction of UPLC (ultraperformance liquid chromatography) and capillary LC enabled better peak resolution and further increase in sensitivity and speed, and it is now successfully applied to metabolomics studies [101].

*Box 2* (NMR and metabolomics). NMR spectroscopy is an analytical technique that exploits the magnetic properties of certain atomic nuclei. It determines the physical and chemical properties of molecules by detecting the magnetically active nuclei. When placed in a magnetic field, an active nucleus (such as  $^1\text{H}$  or  $^{13}\text{C}$ ) absorbs electromagnetic radiation at a characteristic frequency and then reemits it. After absorbing electromagnetic radiation in the range of frequencies of  $^1\text{H}$  (or  $^{13}\text{C}$ , or  $^{31}\text{P}$ , . . .), the sample emits all frequencies of its active nuclei of that type, which constitute its  $^1\text{H}$  (or  $^{13}\text{C}$  or  $^{31}\text{P}$ , . . .) NMR spectrum. The resonance frequency and the corresponding intensity of each signal are dependent, respectively, on the chemical environment where that particular nucleus is located (i.e., molecular structure) and on the concentration of that molecule.

NMR spectroscopy is a nondestructive and highly reproducible technique and provides detailed information on the molecular structure of both pure compounds and complex mixtures [102]. In a typical biological fluid, all hydrogen-containing molecules will give a  $^1\text{H}$ -NMR spectrum as long as they are present in concentrations above the detection limit. The NMR spectrum of a biological fluid is therefore the superposition of the spectra of thousands of different small molecules (up to 2500 for urine and up to 200 for serum/plasma) present in the sample at concentrations  $>1\ \mu\text{M}$  [103]. An advantage of NMR is that the biological fluid requires only a mild treatment prior to the analysis.

The main disadvantage of NMR is its relatively low sensitivity. Another disadvantage of the NMR approach is the difficult identification of all metabolites in the samples:  $^1\text{H}$ -NMR spectra of biological fluids are very complex and often additional two-dimensional NMR experiments may be needed to assign metabolites in biofluids. The development of high-resolution  $^1\text{H}$  magic angle spinning (MAS) spectra made viable the acquisition of data on small slices of tissue without any pretreatment [104–106].



*Box 3* (biofluids and metabolomics). Most biofluids used in metabolomics can be collected noninvasively. The Human Metabolome Database (<http://www.hmdb.ca/>) lists 16 different biofluids investigated and up to 5000 identified or putative metabolites: amniotic fluid (17), aqueous humor (1), ascites fluid (1), bile fluid (18), blood (4297), breast milk (37), cellular cytoplasm (49), cerebrospinal fluid (436), faeces (0), lymph (1), menses (0), mucus (0), pericardial effusion (1), prostate tissue (13), saliva (70), sebum (0), semen (4), sweat (1), synovial fluid (0), tear fluid (1), urine (3873), and vaginal fluid (0). Of these, 694 have been associated with one or more diseases and pathologies.

Blood, urine, cerebrospinal fluid, and saliva are the richest in metabolites. The Human Serum Metabolome project [107] (<http://www.serummetabolome.ca/>) lists 4229 detectable metabolites (most of them lipids) obtained by enhanced NMR, MS, and other analytical platforms. NMR was able to measure 1.2% (49/4229) of the human serum metabolome, GC 2.13% (90/4229), ESI-MS/MS (lipid mediator profiling) 2.3% (96/4229), and TLC/GC-FID-MS (general lipidomics) 79.9% (3381/4229, mostly, however, components of the complex lipid fraction) and DFI MS/MS is able to access 3.3% (139/4229) of the serum metabolome. Some of the compounds identified by NMR are urea (6 mM), glucose (5 mM), lactic acid, (1.4 mM), glutamine (0.51 mM), and glycerol (0.43 mM). The least abundant compounds were carnitine (46  $\mu$ M), acetic acid (42  $\mu$ M), creatine (37  $\mu$ M), cysteine (34  $\mu$ M), propylene glycol (22  $\mu$ M), and aspartic acid (21  $\mu$ M), and the lowest concentration reliably detected using NMR was 12.3  $\mu$ M (for malonic acid) and 14.5  $\mu$ M (for choline).

The Human Urine Metabolome project [108] (<http://www.urinemetabolome.ca/>) lists up to 3100 metabolites identified in urine. Human urine contains many classes of compounds excreted as waste products, including organic acids, amino acids, purines, pyrimidines, sugars, sugar alcohols, sugar acids, amines, and other compounds, at a variety of concentrations. Fresh urine is also characterized by the presence of human cells (erythrocytes, leucocytes, urothelial cells, and epithelial cells), bacteria, fungi, sperms, and noncellular components (mucus filaments, cylinders, cylindroids, pseudocylinders, and crystals, urates).

Some of the metabolites identified in saliva using NMR are [109] glucose, propionate, acetate, taurine, glycine, alanine, sucrose, dimethylamine, formate, glycine, lactate, methanol, propionate, propylene glycol, pyruvate, succinate, and taurine.

A large panel of metabolites has been also identified in cerebrospinal fluid by using NMR and GC-MS [36]. Among those obtained by NMR, there are amino acids, sugars, 2-oxoglutarate, 2-oxoisovalerate, 3-hydroxybutyrate, 3-hydroxyisovalerate, xanthine, and pyruvate.

Up to 50 metabolites were identified in faecal extracts via NMR [43]: amino acids, n-butyrate, propionate, n-caproate, 3-(4'-hydroxyphenyl) propionate, 5-aminopentanoate, glucose, 5-N-acetylneuraminic acid, 5-aminosalicylate, N-acetyl-5-aminosalicylate, deoxycholate, and phenylacetate, many of which are of bacterial origin.

*Box 4* (the metabolomics pipeline). The workflow of a metabolomics study is complex and each step has its own criticalities that need to be addressed. The metabolomics workflow can be summarized as follows [96, 110–112].

*Biological Question.* It includes definition of the biological/biomedical problem to be addressed.

*Study Design.* It involves power analysis and treatment design.

*Data Acquisition.* It concerns quality control strategies, experimental setting (platform specific), Sampling, and measurement design.

*Data Preprocessing.* It is a fundamental step before analysis involving alignment, baseline correction (MR), phasing, alignment, bucketing (NMR), normalization, and scaling.

*Metabolite Identification.* It includes spectral matching (MS) and peak assignment (NMR).

*Statistical Analysis.* It includes explorative (i.e., PCA and clustering), predictive (regression, PLS-DA), and univariate analysis and model optimization and validation.

*Biological Interpretation.* It involves embedding the results within the framework of existing biological knowledge.

*Box 5* (statistical analysis of metabolomic data). Metabolomic data are high dimensional in nature. Tens, hundreds, or even thousands of (un) identified metabolites (relative concentrations are measured by means of NMR or MS platforms, usually on a limited number of samples. Biological information is retrieved from this data by means of univariate and multivariate statistical methods [27, 113, 114]. Multivariate methods make also use of covariances or correlations which reflect the extent of the relationships among the variables, in contrast to univariate methods that focus solely on the mean and variance of a *single* variable.

Commonly used univariate methods are *t*-test and ANOVA [115] together with their corresponding nonparametric versions [116] and with appropriate correction methods for multiple testing [117]. Multivariate methods are a broad category. When the interest centers on predicting or explaining one variable (either a group category like case/control or a continuous response) by the other variables, methods like multiple regression [118] or partial least squares regression and discriminant analysis (PLS-DA) [119] or its extensions like Multilevel PLS-DA [120], Orthogonal PLS-DA [121], and N-way PLS-DA [122] together with a proper optimization and validation of the models [93, 123, 124] are used. In other cases, the interest centers on providing insight into the underlying structure of the complete set of variables and other tools are used. Some examples are principal component analysis (PCA) [125], used to reduce the number of variables when there is correlation present and to explore relations between objects, or cluster analysis [62], used when objects have to be grouped to represent data structure. Hybrid methods like nearest shrunken centroids [126] or simplivariate methods [127] and machine-learning techniques like

artificial neural networks [128], random forest [129], and support vector machines [91] are also used in metabolomics [27, 112, 130].

## Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

## References

- [1] J. A. Murray, C. van Dyke, M. F. Plevak, R. A. Dierkhising, A. R. Zinsmeister, and L. J. Melton III, "Trends in the identification and clinical features of celiac disease in a North American community, 1950–2001," *Clinical Gastroenterology and Hepatology*, vol. 1, no. 1, pp. 19–27, 2003.
- [2] M. Mäki, K. Mustalahti, J. Kokkonen et al., "Prevalence of celiac disease among children in Finland," *The New England Journal of Medicine*, vol. 348, no. 25, pp. 2517–2524, 2003.
- [3] C. Catassi, I. M. Ratsch, E. Fabiani et al., "High prevalence of undiagnosed coeliac disease in 5280 Italian students screened by antigliadin antibodies," *Acta Paediatrica*, vol. 84, no. 6, pp. 672–676, 1995.
- [4] J. West, R. F. A. Logan, P. G. Hill et al., "Seroprevalence, correlates, and characteristics of undetected coeliac disease in England," *Gut*, vol. 52, no. 7, pp. 960–965, 2003.
- [5] P. J. Bingley, A. J. K. Williams, A. J. Norcross et al., "Undiagnosed coeliac disease at age seven: population based prospective birth cohort study," *The British Medical Journal*, vol. 328, no. 7435, pp. 322–323, 2004.
- [6] P. C. A. Dubois, G. Trynka, L. Franke et al., "Multiple common variants for celiac disease influencing immune gene expression," *Nature Genetics*, vol. 42, no. 4, pp. 295–302, 2010.
- [7] G. Trynka, C. Wijmenga, and D. A. van Heel, "A genetic perspective on coeliac disease," *Trends in Molecular Medicine*, vol. 16, no. 11, pp. 537–550, 2010.
- [8] P. Price, C. Witt, R. Allcock et al., "The genetic basis for the association of the 8.1 ancestral haplotype (A1, B8, DR3) with multiple immunopathological diseases," *Immunological Reviews*, vol. 167, pp. 257–274, 1999.
- [9] D. A. van Heel, L. Franke, K. A. Hunt et al., "A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21," *Nature Genetics*, vol. 39, no. 7, pp. 827–829, 2007.
- [10] K. A. Hunt, A. Zhernakova, G. Turner et al., "Newly identified genetic risk variants for celiac disease related to the immune response," *Nature Genetics*, vol. 40, no. 4, pp. 395–402, 2008.
- [11] G. Trynka, K. A. Hunt, N. A. Bockett et al., "Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease," *Nature Genetics*, vol. 43, no. 12, pp. 1193–1201, 2011.
- [12] D. J. Smyth, V. Plagnol, N. M. Walker et al., "Shared and distinct genetic variants in type 1 diabetes and celiac disease," *The New England Journal of Medicine*, vol. 359, no. 26, pp. 2767–2777, 2008.
- [13] J. C. Barrett, D. G. Clayton, P. Concannon et al., "Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes," *Nature Genetics*, vol. 41, no. 6, pp. 703–707, 2009.
- [14] A. Zhernakova, B. Z. Alizadeh, M. Bevova et al., "Novel association in chromosome 4q27 region with rheumatoid arthritis and confirmation of type 1 diabetes point to a general risk locus for autoimmune diseases," *The American Journal of Human Genetics*, vol. 81, no. 6, pp. 1284–1288, 2007.
- [15] J. W. Han, H. F. Zheng, Y. Cui et al., "Genome-wide association study in a Chinese Han population identifies nine new susceptibility loci for systemic lupus erythematosus," *Nature Genetics*, vol. 41, no. 11, pp. 1234–1237, 2009.
- [16] F. Koning, "Celiac disease: caught between a rock and a hard place," *Gastroenterology*, vol. 129, no. 4, pp. 1294–1301, 2005.
- [17] E. M. Nilsen, F. L. Jahnsen, K. E. A. Lundin et al., "Gluten induces an intestinal cytokine response strongly dominated by interferon gamma in patients with celiac disease," *Gastroenterology*, vol. 115, no. 3, pp. 551–563, 1998.
- [18] L. M. Sollid, "Coeliac disease: dissecting a complex inflammatory disorder," *Nature Reviews Immunology*, vol. 2, no. 9, pp. 647–655, 2002.
- [19] J.-J. Mention, M. B. Ahmed, B. Bègue et al., "Interleukin 15: a key to disrupted intraepithelial lymphocyte homeostasis and lymphomagenesis in celiac disease," *Gastroenterology*, vol. 125, no. 3, pp. 730–745, 2003.
- [20] I. Bjarnason, T. J. Peters, and N. Veall, "A persistent defect in intestinal permeability in coeliac disease demonstrated by a 51Cr-labelled EDTA absorption test," *Lancet*, vol. 1, no. 8320, pp. 323–325, 1983.
- [21] A. Fasano, T. Not, W. Wang et al., "Zonulin, a newly discovered modulator of intestinal permeability, and its expression in coeliac disease," *The Lancet*, vol. 355, no. 9214, pp. 1518–1519, 2000.
- [22] K. M. Lammers, R. Lu, J. Brownley et al., "Gliadin induces an increase in intestinal permeability and zonulin release by binding to the chemokine receptor CXCR3," *Gastroenterology*, vol. 135, no. 1, pp. 194–e3, 2008.
- [23] A. Tripathi, K. M. Lammers, S. Goldblum et al., "Identification of human zonulin, a physiological modulator of tight junctions, as prehaptoglobin-2," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, no. 39, pp. 16799–16804, 2009.
- [24] S. Schloissnig, M. Arumugam, S. Sunagawa et al., "Genomic variation landscape of the human gut microbiome," *Nature*, vol. 493, no. 7430, pp. 45–50, 2013.
- [25] H. Curtis, G. Dirk, K. Rob et al., "Structure, function and diversity of the healthy human microbiome," *Human Microbiome Project Consortium*, vol. 486, no. 7402, pp. 207–214, 2012.
- [26] R. D. Fleischmann, M. D. Adams, O. White et al., "Whose genome random sequencing and assembly of Haemophilus influenzae Rd," *Science*, vol. 269, no. 5223, pp. 496–521, 1995.
- [27] M. Brown, W. B. Dunn, D. I. Ellis et al., "A metabolome pipeline: from concept to data to knowledge," *Metabolomics*, vol. 1, no. 1, pp. 39–51, 2005.
- [28] C. J. Lindon, J. K. Nicholson, and E. Holmes, *The Handbook of Metabonomics and Metabolomics*, Elsevier, 2007.
- [29] J. K. Nicholson, J. C. Lindon, and E. Holmes, "Metabonomics: understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data," *Xenobiotica*, vol. 29, no. 11, pp. 1181–1189, 1999.
- [30] O. Fiehn, "Metabolomics—the link between genotypes and phenotypes," *Plant Molecular Biology*, vol. 48, no. 1–2, pp. 155–171, 2002.
- [31] S. G. Oliver, M. K. Winson, D. B. Kell, and F. Baganz, "Systematic functional analysis of the yeast genome," *Trends in Biotechnology*, vol. 16, no. 9, pp. 373–378, 1998.

- [32] J. Li, O. Brazhnik, A. Kamal et al., *Metabolic Profiling—Its Role in Biomarker Discovery and Gene Function Analysis*, Kluwer Academic, 2003.
- [33] L. M. Raamsdonk, B. Teusink, D. Broadhurst et al., “A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations,” *Nature Biotechnology*, vol. 19, no. 1, pp. 45–50, 2001.
- [34] M. Aimetti, S. Cacciatore, A. Graziano, and L. Tenori, “Metabonomic analysis of saliva reveals generalized chronic periodontitis signature,” *Metabolomics*, vol. 8, no. 3, pp. 465–474, 2012.
- [35] H. Wu, A. D. Southam, A. Hines, and M. R. Viant, “High-throughput tissue extraction protocol for NMR- and MS-based metabolomics,” *Analytical Biochemistry*, vol. 372, no. 2, pp. 204–212, 2008.
- [36] D. S. Wishart, M. J. Lewis, J. A. Morrissey et al., “The human cerebrospinal fluid metabolome,” *Journal of Chromatography B*, vol. 871, no. 2, pp. 164–173, 2008.
- [37] G. A. Nagana Gowda, N. Shanaiah, A. Cooper, M. Maluccio, and D. Raftery, “Bile acids conjugation in human bile is not random: new insights from 1H-NMR spectroscopy at 800 MHz,” *Lipids*, vol. 44, no. 6, pp. 527–535, 2009.
- [38] A. D. Maher, O. Cloarec, P. Patki et al., “Dynamic biochemical information recovery in spontaneous human seminal fluid reactions via 1H NMR kinetic statistical total correlation spectroscopy,” *Analytical Chemistry*, vol. 81, no. 1, pp. 288–295, 2009.
- [39] G. Graça, I. F. Duarte, B. J. Goodfellow et al., “Metabolite profiling of human amniotic fluid by hyphenated nuclear magnetic resonance spectroscopy,” *Analytical Chemistry*, vol. 80, no. 15, pp. 6085–6092, 2008.
- [40] L. Lacitignola, F. P. Fanizzi, E. Francioso, and A. Crovace, “1H NMR investigation of normal and osteoarthritic synovial fluid in the horse,” *Veterinary and Comparative Orthopaedics and Traumatology*, vol. 21, no. 1, pp. 85–88, 2008.
- [41] P. Montuschi, D. Paris, D. Melck et al., “NMR spectroscopy metabolomic profiling of exhaled breath condensate in patients with stable and unstable cystic fibrosis,” *Thorax*, vol. 67, no. 3, pp. 222–228, 2012.
- [42] I. Bertini, C. Luchinat, M. Miniati, S. Monti, and L. Tenori, “Phenotyping COPD by 1H NMR of exhaled breath condensate,” *Metabolomics*, 2013.
- [43] G. le Gall, S. O. Noor, K. Ridgway et al., “Metabolomics of fecal extracts detects altered metabolic activity of gut microbiota in ulcerative colitis and irritable bowel syndrome,” *Journal of Proteome Research*, vol. 10, no. 9, pp. 4208–4218, 2011.
- [44] C. Oakman, L. Tenori, W. M. Claudino et al., “Identification of a serum-detectable metabolomic fingerprint potentially correlated with the presence of micrometastatic disease in early breast cancer patients at varying risks of disease relapse by traditional prognostic methods,” *Annals of Oncology*, vol. 22, no. 6, pp. 1295–1301, 2011.
- [45] M. Coen, M. O’Sullivan, W. A. Bubb, P. W. Kuchel, and T. Sorrell, “Proton nuclear magnetic resonance-based metabonomics for rapid diagnosis of meningitis and ventriculitis,” *Clinical Infectious Diseases*, vol. 41, no. 11, pp. 1582–1590, 2005.
- [46] T. Bartsch, K. Alfke, S. Wolff, A. Rohr, O. Jansen, and G. Deuschl, “Focal MR spectroscopy of hippocampal CA-1 lesions in transient global amnesia,” *Neurology*, vol. 70, no. 13, pp. 1030–1035, 2008.
- [47] P. Bernini, I. Bertini, C. Luchinat, L. Tenori, and A. Tognaccini, “The cardiovascular risk of healthy individuals studied by NMR metabonomics of plasma samples,” *Journal of Proteome Research*, vol. 10, no. 11, pp. 4983–4992, 2011.
- [48] S. H. Moolenaar, U. F. H. Engelke, and R. A. Wevers, “Proton nuclear magnetic resonance spectroscopy of body fluids in the field of inborn errors of metabolism,” *Annals of Clinical Biochemistry*, vol. 40, part 1, pp. 16–24, 2003.
- [49] I. Bertini, A. Calabrò, V. de Carli et al., “The metabonomic signature of celiac disease,” *Journal of Proteome Research*, vol. 8, no. 1, pp. 170–177, 2009.
- [50] P. Bernini, I. Bertini, A. Calabrò et al., “Are patients with potential celiac disease really potential? The answer of metabonomics,” *Journal of Proteome Research*, vol. 10, no. 2, pp. 714–721, 2011.
- [51] M. Sellitto, G. Bai, G. Serena et al., “Proof of concept of microbiome-metabolome analysis and delayed gluten exposure on celiac disease autoimmunity in genetically at-risk infants,” *PLoS ONE*, vol. 7, no. 3, Article ID e33387, 2012.
- [52] Y. Sanz, G. D. Pama, and M. Laparra, “Unraveling the ties between celiac disease and intestinal microbiota,” *International Reviews of Immunology*, vol. 30, no. 4, pp. 207–218, 2011.
- [53] E. Sánchez, E. Donat, C. Ribes-Koninckx, L. Fernández-Murga, and Y. Sanz, “Duodenal-mucosal bacteria associated with celiac disease in children,” *Applied and Environmental Microbiology*, vol. 79, no. 18, pp. 5472–5479, 2013.
- [54] Y. Sanz, E. Sánchez, M. Marzotto, M. Calabuig, S. Torriani, and F. Dellaglio, “Differences in faecal bacterial communities in coeliac and healthy children as detected by PCR and denaturing gradient gel electrophoresis,” *FEMS Immunology and Medical Microbiology*, vol. 51, no. 3, pp. 562–568, 2007.
- [55] P. Wacklin, K. Kaukinen, E. Tuovinen et al., “The duodenal microbiota composition of adult celiac disease patients is associated with the clinical manifestation of the disease,” *Inflammatory Bowel Diseases*, vol. 19, no. 5, pp. 934–941, 2013.
- [56] M. Kalliomaki, R. Satokari, H. Lahteenoja et al., “Expression of microbiota, toll-like receptors, and their regulators in the small intestinal mucosa in celiac disease,” *Journal of Pediatric Gastroenterology*, vol. 54, no. 6, pp. 727–732, 2012.
- [57] E. Nistal, A. Caminero, A. R. Herrán et al., “Differences of small intestinal bacteria populations in adults and children with/without celiac disease: effect of age, gluten diet, and disease,” *Inflammatory Bowel Diseases*, vol. 18, no. 4, pp. 649–656, 2012.
- [58] I. Nadal, E. Donant, C. Ribes-Koninckx, M. Calabuig, and Y. Sanz, “Imbalance in the composition of the duodenal microbiota of children with coeliac disease,” *Journal of Medical Microbiology*, vol. 56, no. 12, pp. 1669–1674, 2007.
- [59] E. Sánchez, E. Donat, C. Ribes-Koninckx, M. Calabuig, and Y. Sanz, “Intestinal Bacteroides species associated with coeliac disease,” *Journal of Clinical Pathology*, vol. 63, no. 12, pp. 1105–1111, 2010.
- [60] E. Sánchez, G. de Palma, A. Capilla et al., “Influence of environmental and genetic factors linked to celiac disease risk on infant gut colonization by Bacteroides species,” *Applied and Environmental Microbiology*, vol. 77, no. 15, pp. 5316–5323, 2011.
- [61] J. Cheng, M. Kalliomaki, H. G. Heilig et al., “Duodenal microbiota composition and mucosal homeostasis in pediatric celiac disease,” *BMC Gastroenterology*, vol. 13, article 113, 2013.
- [62] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, vol. 344, John Wiley & Sons, 2009.
- [63] R. di Cagno, M. de Angelis, I. de Pasquale et al., “Duodenal and faecal microbiota of celiac children: molecular, phenotype and metabolome characterization,” *BMC Microbiology*, vol. 11, article 219, 2011.

- [64] M. Medina, G. de Palma, C. Ribes-Koninckx, M. Calabuig, and Y. Sanz, "Bifidobacterium strains suppress in vitro the pro-inflammatory milieu triggered by the large intestinal microbiota of coeliac patients," *Journal of Inflammation*, vol. 5, article 19, 2008.
- [65] G. Ou, M. Hedberg, P. Hörstedt et al., "Proximal small intestinal microbiota and identification of rod-shaped bacteria associated with childhood celiac disease," *The American Journal of Gastroenterology*, vol. 104, no. 12, pp. 3058–3067, 2009.
- [66] J. M. Norris, K. Barriga, E. J. Hoffenberg et al., "Risk of celiac disease autoimmunity and timing of gluten introduction in the diet of infants at increased risk of disease," *Journal of the American Medical Association*, vol. 293, no. 19, pp. 2343–2351, 2005.
- [67] J. Kranich, K. M. Maslowski, and C. R. Mackay, "Commensal flora and the regulation of inflammatory and autoimmune responses," *Seminars in Immunology*, vol. 23, no. 2, pp. 139–145, 2011.
- [68] J. L. Round and S. K. Mazmanian, "The gut microbiota shapes intestinal immune responses during health and disease," *Nature Reviews Immunology*, vol. 9, no. 5, pp. 313–323, 2009.
- [69] L. Plot and H. Amital, "Infectious associations of celiac disease," *Autoimmunity Reviews*, vol. 8, no. 4, pp. 316–319, 2009.
- [70] M. F. Kagnoff, R. K. Austin, and J. J. Hubert, "Possible role for a human adenovirus in the pathogenesis of celiac disease," *Journal of Experimental Medicine*, vol. 160, no. 5, pp. 1544–1557, 1984.
- [71] C. Ruggeri, A. T. la Masa, S. Rudi et al., "Celiac disease and non-organ-specific autoantibodies in patients with chronic hepatitis C virus infection," *Digestive Diseases and Sciences*, vol. 53, no. 8, pp. 2151–2155, 2008.
- [72] L. C. Stene, M. C. Honeyman, E. J. Hoffenberg et al., "Rotavirus infection frequency and risk of celiac disease autoimmunity in early childhood: a longitudinal study," *The American Journal of Gastroenterology*, vol. 101, no. 10, pp. 2333–2340, 2006.
- [73] E. F. Verdu, M. Mauro, J. Bourgeois, and D. Armstrong, "Clinical onset of celiac disease after an episode of *Campylobacter jejuni* enteritis," *Canadian Journal of Gastroenterology*, vol. 21, no. 7, pp. 453–455, 2007.
- [74] F. C. Jordá and J. L. Vivancos, "Fatigue as a determinant of health in patients with celiac disease," *Journal of Clinical Gastroenterology*, vol. 44, no. 6, pp. 423–427, 2009.
- [75] E. Capristo, G. Addolorato, A. V. Greco et al., "Body composition and energy metabolism in celiac disease [2] (multiple letters)," *The American Journal of Gastroenterology*, vol. 93, no. 9, pp. 1596–1597, 1998.
- [76] S. Bode, C. Hassager, E. Gudmand-Hoyer, and C. Christiansen, "Body composition and calcium metabolism in adult treated coeliac disease," *Gut*, vol. 32, no. 11, pp. 1342–1345, 1991.
- [77] E. Capristo, G. Addolorato, G. Mingrone et al., "Changes in body composition, substrate oxidation, and resting metabolic rate in adult celiac disease patients after a 1-y gluten-free diet treatment," *The American Journal of Clinical Nutrition*, vol. 72, no. 1, pp. 76–81, 2000.
- [78] M. Kojima, H. Hosoda, Y. Date, M. Nakazato, H. Matsuo, and K. Kangawa, "Ghrelin is a growth-hormone-releasing acylated peptide from stomach," *Nature*, vol. 402, no. 6762, pp. 656–660, 1999.
- [79] A. M. Wren, C. J. Small, H. L. Ward et al., "The novel hypothalamic peptide ghrelin stimulates food intake and growth hormone secretion," *Endocrinology*, vol. 141, no. 11, pp. 4325–4328, 2000.
- [80] A. Lanzini, P. Magni, M. L. Petroni et al., "Circulating ghrelin level is increased in coeliac disease as in functional dyspepsia and reverts to normal during gluten-free diet," *Alimentary Pharmacology and Therapeutics*, vol. 23, no. 7, pp. 907–913, 2006.
- [81] M. A. Selimoglu, S. Altinkaynak, V. Ertekin, and F. Akcay, "Serum ghrelin levels in children with celiac disease," *Journal of Clinical Gastroenterology*, vol. 40, no. 3, pp. 191–194, 2006.
- [82] E. Capristo, S. Farnetti, G. Mingrone et al., "Reduced plasma ghrelin concentration in celiac disease after gluten-free diet treatment," *Scandinavian Journal of Gastroenterology*, vol. 40, no. 4, pp. 430–436, 2005.
- [83] S. Meiboom and D. Gill, "Modified spin-echo method for measuring nuclear relaxation times," *Review of Scientific Instruments*, vol. 29, no. 8, pp. 688–691, 1958.
- [84] A. N. Phipps, J. Stewart, B. Wright, and I. D. Wilson, "Effect of diet on the urinary excretion of hippuric acid and other dietary-derived aromatics in rat. A complex interaction between diet, gut microflora and substrate specificity," *Xenobiotica*, vol. 28, no. 5, pp. 527–537, 1998.
- [85] D. H. Bueschgens and M. E. Stiles, "Escherichia coli variants for gas and indole production at elevated incubation temperatures," *Applied and Environmental Microbiology*, vol. 48, no. 3, pp. 601–605, 1984.
- [86] J. K. Nicholson, E. Holmes, and I. D. Wilson, "Gut microorganisms, mammalian metabolism and personalized health care," *Nature Reviews Microbiology*, vol. 3, no. 5, pp. 431–438, 2005.
- [87] M. Duran, D. Ketting, P. K. de Bree, C. van der Heiden, and S. K. Wadman, "Gas chromatographic analysis of urinary volatile phenols in patients with gastro intestinal disorders and normals," *Clinica Chimica Acta*, vol. 45, no. 4, pp. 341–347, 1973.
- [88] E. Bone, A. Tamm, and M. Hill, "The production of urinary phenols by gut bacteria and their possible role in the causation of large bowel cancer," *The American Journal of Clinical Nutrition*, vol. 29, no. 12, pp. 1448–1454, 1976.
- [89] S. de Jong, "SIMPLS: an alternative approach to partial least squares regression," *Chemometrics and Intelligent Laboratory Systems*, vol. 18, no. 3, pp. 251–263, 1993.
- [90] I. González, S. Déjean, P. Martin, and A. Baccini, "CCA: an R package to extend canonical correlation analysis," *Journal of Statistical Software*, vol. 23, no. 12, pp. 1–14, 2008.
- [91] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
- [92] J. Trygg and S. Wold, "Orthogonal projections to latent structures (O-PLS)," *Journal of Chemometrics*, vol. 16, no. 3, pp. 119–128, 2002.
- [93] E. Szymańska, E. Saccenti, A. K. Smilde, and J. A. Westerhuis, "Double-check: validation of diagnostic statistics for PLS-DA models in metabolomics studies," *Metabolomics*, vol. 8, pp. 3–16, 2012.
- [94] W. M. Claudino, A. Quattrone, L. Biganzoli, M. Pestrin, I. Bertini, and A. di Leo, "Metabolomics: available results, current research projects in breast cancer, and future applications," *Journal of Clinical Oncology*, vol. 25, no. 19, pp. 2840–2846, 2007.
- [95] R. Vanga and D. A. Leffler, "Gluten sensitivity: not celiac and not certain," *Gastroenterology*, vol. 145, no. 2, pp. 276–279, 2013.
- [96] R. Madsen, T. Lundstedt, and J. Trygg, "Chemometrics in metabolomics—a review in human disease diagnosis," *Analytica Chimica Acta*, vol. 659, no. 1–2, pp. 23–33, 2010.
- [97] J. K. Nicholson and J. C. Lindon, "Systems biology: metabolomics," *Nature*, vol. 455, no. 7216, pp. 1054–1056, 2008.

- [98] R. P. Bo Zhang, "Analysis of bacterial biofilms using NMR-based metabolomics," *Future Medicinal Chemistry*, vol. 4, no. 10, pp. 1273–1306, 2012.
- [99] W. Wolfram, *Metabolomics Methods and Protocols*, Humana Press Inc., 2006.
- [100] Z. Wu, Z. Huang, R. Lehmann, C. Zhao, and G. Xu, "The application of chromatography-mass spectrometry: methods to metabolomics," *Chromatographia*, vol. 69, supplement 1, pp. S23–S32, 2009.
- [101] I. D. Wilson, J. K. Nicholson, J. Castro-Perez et al., "High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies," *Journal of Proteome Research*, vol. 4, no. 2, pp. 591–598, 2005.
- [102] T. D. W. Claridge, *High-Resolution NMR Techniques in Organic Chemistry*, Elsevier, 2008.
- [103] D. S. Wishart, D. Tzur, C. Knox et al., "HMDB: the human metabolome database," *Nucleic Acids Research*, vol. 35, no. 1, pp. D521–D526, 2007.
- [104] A. M. Tomlins, P. J. D. Foxall, J. C. Lindon et al., "High resolution magic angle spinning 1H nuclear magnetic resonance analysis of intact prostatic hyperplastic and tumour tissues," *Analytical Communications*, vol. 35, no. 3, pp. 113–115, 1998.
- [105] L. L. Cheng, I.-W. Chang, D. N. Louis, and R. G. Gonzalez, "Correlation of high-resolution magic angle spinning proton magnetic resonance spectroscopy with histopathology of intact human brain tumor specimens," *Cancer Research*, vol. 58, no. 9, pp. 1825–1832, 1998.
- [106] S. Garrod, E. Humpfer, M. Spraul et al., "High-resolution magic angle spinning 1H NMR spectroscopic studies on intact rat renal cortex and medulla," *Magnetic Resonance in Medicine*, vol. 41, no. 6, pp. 1108–1118, 1999.
- [107] N. Psychogios, D. D. Hau, J. Peng et al., "The human serum metabolome," *PLoS ONE*, vol. 6, no. 2, Article ID e16957, 2011.
- [108] S. Bouatra, F. Aziat, R. Mandal et al., "The human urine metabolome," *PLoS ONE*, vol. 8, no. 9, Article ID e73076, 2013.
- [109] I. Takeda, C. Stretch, P. Barnaby et al., "Understanding the human salivary metabolome," *NMR in Biomedicine*, vol. 22, no. 6, pp. 577–584, 2009.
- [110] W. B. Dunn and D. I. Ellis, "Metabolomics: current analytical platforms and methodologies," *Trends in Analytical Chemistry*, vol. 24, no. 4, pp. 285–294, 2005.
- [111] R. Goodacre, D. Broadhurst, A. K. Smilde et al., "Proposed minimum reporting standards for data analysis in metabolomics," *Metabolomics*, vol. 3, no. 3, pp. 231–241, 2007.
- [112] M. M. W. B. Hendriks, F. A. V. Eeuwijk, R. H. Jellema et al., "Data-processing strategies for metabolomics studies," *Trends in Analytical Chemistry*, vol. 30, no. 10, pp. 1685–1698, 2011.
- [113] K. H. Liland, "Multivariate methods in metabolomics—from pre-processing to dimension reduction and statistical analysis," *Trends in Analytical Chemistry*, vol. 30, no. 6, pp. 827–841, 2011.
- [114] E. Saccenti, H. C. J. Hoefsloot, A. K. Smilde, J. A. Westerhuis, and M. M. W. B. Hendriks, "Reflections on univariate and multivariate analysis of metabolomics data," *Metabolomics*, 2013.
- [115] R. G. Miller and B. W. Brown, *Beyond ANOVA: Basics of Applied Statistics*, Chapman & Hall/CRC, 1997.
- [116] J. Fletcher, "Non-parametric tests," *The British Medical Journal*, vol. 338, 2009.
- [117] D. I. Broadhurst and D. B. Kell, "Statistical strategies for avoiding false discoveries in metabolomics and related experiments," *Metabolomics*, vol. 2, no. 4, pp. 171–196, 2006.
- [118] N. R. Draper and H. Smith, *Applied Regression Analysis*, Wiley Series in Probability and Statistics, 1998.
- [119] S. Wold, M. Sjöström, and L. Eriksson, "PLS-regression: a basic tool of chemometrics," *Chemometrics and Intelligent Laboratory Systems*, vol. 58, no. 2, pp. 109–130, 2001.
- [120] E. J. J. van Velzen, J. A. Westerhuis, J. P. M. van Duynhoven et al., "Multilevel data analysis of a crossover designed human nutritional intervention study," *Journal of Proteome Research*, vol. 7, no. 10, pp. 4483–4491, 2008.
- [121] H. Stenlund, A. Gorzsás, P. Persson, B. Sundberg, and J. Trygg, "Orthogonal projections to latent structures discriminant analysis modeling on in situ FT-IR spectral imaging of liver tissue for identifying sources of variability," *Analytical Chemistry*, vol. 80, no. 18, pp. 6898–6906, 2008.
- [122] R. Bro, "Multiway calibration. Multilinear PLS," *Journal of Chemometrics*, vol. 10, no. 1, pp. 47–61, 1996.
- [123] J. A. Westerhuis, H. C. J. Hoefsloot, S. Smit et al., "Assessment of PLS-DA cross validation," *Metabolomics*, vol. 4, no. 1, pp. 81–89, 2008.
- [124] R. G. Brereton, "Consequences of sample size, variable selection, and model validation and optimisation, for predicting classification ability from analytical data," *Trends in Analytical Chemistry*, vol. 25, no. 11, pp. 1103–1111, 2006.
- [125] I. T. Jolliffe, "Principal component analysis: a beginner's guide-I. Introduction and application," *Weather*, vol. 45, no. 10, pp. 375–382, 2012.
- [126] R. Tibshirani, T. Hastie, B. Narasimhan, and G. Chu, "Diagnosis of multiple cancer types by shrunken centroids of gene expression," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 10, pp. 6567–6572, 2002.
- [127] E. Saccenti, J. A. Westerhuis, A. K. Smilde, M. J. van der Werf, J. A. Hageman, and M. M. W. B. Hendriks, "Simplivariate models: uncovering the underlying biology in functional genomics data," *PLoS ONE*, vol. 6, no. 6, Article ID e20747, 2011.
- [128] D. Lowe and D. Broomhead, "Multivariable functional interpolation and adaptive networks," *Complex Systems*, vol. 2, pp. 321–355, 1988.
- [129] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [130] M. M. W. B. Hendriks, S. Smit, W. L. M. W. Akkermans et al., "How to distinguish healthy from diseased? Classification strategy for mass spectrometry-based clinical proteomics," *Proteomics*, vol. 7, no. 20, pp. 3672–3680, 2007.