Computers in Biology and Medicine 104 (2019) 183–196

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Equations defined using gene expression and histological data resolve coeliac disease biopsies within the Marsh score continuum

Richard P.G. Charlesworth[a,*], Linda L. Agnew[a], David R. Scott[b], Nicholas M. Andronicos[a]

[a] Discipline of Biomedical Sciences, School of Science and Technology, University of New England, Armidale, NSW, 2351, Australia
[b] Gastroenterologist, Hunter New England Area Health Service, Tamworth, NSW, 2340, Australia

ARTICLE INFO

Keywords:
Coeliac disease
Marsh score
Gene expression
Discriminant analysis

ABSTRACT

Background/Aim: The gold standard diagnostic for coeliac disease (CD) is subjective histological assignment of biopsies into the Marsh score categories. It is hypothesized that discrete Marsh score categories can be quantitatively resolved into a continuum using discriminant equations defined using histological and gene expression data. Therefore, the aim of this study was to use a combination of histological and gene expression data to develop equations that classify CD patient biopsies into a quantitative Marsh score continuum which could be used by clinicians to monitor CD treatment efficacy.

Methods: Both empirical and simulated gene expression and histological data were used to define predictive Marsh score equations. The distances of treated sample biopsies from the Marsh score standards were determined using the Mahalanobis distance calculation.

Results: Three function, high resolution discriminant equations derived from simulated data were used to accurately classify 99.6% of simulated and empirically derived biopsy data. The first function resolved active (Marsh type 3) CD from mild (Marsh type 1) CD. The second function resolved normal (no specific pathology) biopsies from mild CD. The third function resolved active Marsh score 3 into a and b subcategories. Finally, measuring the Mahalanobis distance enabled the conversion of discrete Marsh score categories into a continuum.

Conclusions: This proof-of-concept study successfully demonstrated that the discrete Marsh score scale can be converted into a quantitative continuum capable of high resolution monitoring of patient treatment efficacy using equations defined by gene expression and histology data.

1. Introduction

Coeliac disease (CD) is a chronic intestinal condition [1,2] that currently affects approximately 1% of the United States and European populations [3] and if left untreated may lead to an increased risk of further complications, including carcinoma [4]. The American Gastroenterological Association recommends a biopsy to confirm the diagnosis of individuals suspected of having CD [5,6].

The histological assessment of duodenal biopsies and assignment into a Marsh score category is qualitative and subjective however. Therefore patients with mild or equivocal histological damage can be assigned into a lesser Marsh score category more ambiguously than patients with severe histological damage. It should also be noted that some patients with mild histological damage may display severe symptomology and vice versa. This ambiguity in diagnosis may impact treatment or the ability to effectively monitor treatment progression in patients [7–10].

Several studies have attempted to define high resolution, objective equations that can be used to classify CD biopsies into discrete Marsh score categories. IgA/IgG serology [11,12]; IEL counts with crypt/villous ratios [13]; capsule endoscopy images [14]; immunohistochemistry data [15] and histological analysis [16] have all been used to define classification equations. However, these parameters all resulted in the definition of low resolution classification equations capable of discriminating only between healthy tissue and pathological tissue, and were deemed uninformative for classifying mild or equivocal cases of CD. In contrast, gene expression analysis using a CD qRT-PCR array defined the expression of genes associated with discrete Marsh score categories [17] and have been used to define discriminant equations capable of classifying patient biopsies into discrete Marsh score categories. It is therefore hypothesized that the definition of discriminant equations using parameters derived from both quantitative duodenal histological parameters and duodenal gene expression data will enable the high-resolution classification of CD patient biopsies along the Marsh score continuum.

* Corresponding author. McClymont Building (W34), University of New England, Elm Avenue, Armidale, NSW, 2351, Australia.
E-mail address: rcharle3@une.edu.au (R.P.G. Charlesworth).

https://doi.org/10.1016/j.compbiomed.2018.10.036
Received 12 July 2018; Received in revised form 31 October 2018; Accepted 31 October 2018

Table 1
Definition of four functions to discriminate Marsh score on the basis of significant histological and gene expression variables for the original study cohort.

	Function 1	Function 2	Function 3	Function 4
Eigenvalue	449.783	50.360	3.442	1.156
% of Variance	89.112	9.977	0.682	0.229
Cumulative Variance (%)	89.112	99.089	99.771	100.000
Canonical Correlation	0.999	0.990	0.880	0.732
Wilks' Lambda*	4.59×10^{-6}	0.002	0.104	0.464
$\chi 2$	123.093	61.983	22.594	7.684
DF	68	48	30	14
P-value	4×10^{-5}	0.085	0.832	0.905

2. Methods

2.1. Experimental data

This study was reviewed by the institutional human research ethics committee and was performed in accordance with the ethical standards laid down by the Declaration of Helsinki (2013). All patients gave written, informed consent prior to their inclusion in the study (approval number HE13-184). Charlesworth et al., 2017 and 2018 studies consisted of duodenal biopsy material sourced from patients who were diagnosed with varying stages of CD (positive serology and positive histology confirmed by an independent histopathologist working at a commercial pathology laboratory), patients with treated CD and healthy control patients. The duodenal biopsy histological parameters used to formulate discriminant equations were derived from Charlesworth et al., 2017 [16]. As stated in Charlesworth et al., 2018 the genes selected for the pPCR array were analysed using Cytoscape from published expression datasets. The gene expression data used to define the discriminant equations have been reported in Charlesworth et al., 2018.

2.2. Generation of simulated histological and gene expression data

This study used two datasets, one consisting of histological data [16] and one consisting of gene expression data [17]. For the eight histological [16] and twenty-six gene expression [17] variables which had a significant association with CD, large random Gaussian datasets were generated using the mean and standard deviation of the original study data for each variable (Random.org, 2010). This particular platform used a Box-Müller transform; a technique for selecting independent, normally distributed random numbers from a source of uniformly distributed random numbers [18,19]. This process has been reported to be more computationally efficient than previous methods and results in the randomness in a sample coming from the inherent noise of the distribution [18,19]. Values which returned a negative number for this technique were treated as zero (i.e. equivalent to being not detected by gene expression analysis or unable to be measured histologically). The quality of the simulated Gaussian datasets were determined using Kolmogorov-Smirnov (K-S) normality and ANOVA analyses (SPSS v22.0, IBM, USA) by comparing the simulated datasets back to the original empirically-derived data.

2.3. Discriminant equation definition and testing

To generate predictive equations from histological and gene expression data, MANOVA and discriminant analysis was performed using SPSS (v22.0, IBM, USA). The validity of the discriminant equations was examined using a holdout analysis. Mahalanobis testing (SPSS v22.0, IBM, USA; XLSTAT 2017, Addinsoft, USA) was performed to assess potential alterations in centroid groupings and to quantify the distance of individual biopsies from Marsh score standards.

3. Results

3.1. Definition of discriminant equations using both histological and gene expression parameters

Definition of discriminant equations using empirical histological data [16] and gene expression data [17] resulted in four discriminant equations (Table 1). It was observed that function 1 was highly statistically significant ($p = 4 \times 10^{-5}$) and accounted for approximately 89% of the total observed experimental variation. The canonical correlation of function 1 was 0.999, which is extremely strong and indicated that the patient biopsies were separated into discrete groups by function 1 (Fig. 1A).

In contrast, function 2 accounted for 10% of the total observed variation but did not make a significant ($p = 0.085$) contribution to discrimination capacity of the equation. However, when function 2 was also included in the classification equation the various Marsh score grades were resolved into Cartesian quadrants (Fig. 1B). Definition of duodenal biopsies using functions 1 and 2 provided moderate resolution between Marsh type 3a and Marsh type 3b patient biopsies. Marsh type 3a biopsies were defined exclusively by the histological parameters mean crypt goblet cells, mean villous length and mean lymphocytes, which were all associated with active CD (Fig. 1B). Marsh type 3b biopsies were defined by the histological parameter mean villous width (an indirect measure of leukocyte infiltration) and the expression of FGF7 and the immune genes HLA-E, CD117 (Fig. 1B). Recovery from active CD (M3) towards a less pathological state (M1) were defined by the histological parameters mean villous goblet cells, mean plasma cells and mean monocytes and the expression of S100a8, C5A, TGM2, CD11c (a monocyte marker) and CD19 (a B lymphocyte marker) genes. Finally, duodenal biopsies from subjects with no specific pathology were defined by large negative mean infiltrative leukocytes and S100a9 gene expression parameters. The negative coefficient of the mean infiltrative leukocyte parameter indicated that the absence of leukocyte infiltration correlated with a non-pathological (normal) duodenal biopsy (Fig. 1B). The Marsh score category was predicted by the two functions with 100% accuracy using a standard function calculations (Table 2). However, a more stringent holdout analysis was unable to be performed due to the limited patient numbers in the pilot studies.

A two-function discriminant equation was used to classify the Marsh score of patient samples as well as the samples from patients that had undergone treatment with a gluten-free diet (TF1, TF7, TF10 and TF21). Two other patients (TM1 and TF6) that had biopsies before and after treatment (gluten-free diets) were also superimposed on the Marsh score categories as well as patients given equivocal diagnoses (Fig. 1C). Analysis of the histology and gene expression parameters from these treated CD patients by the two-function equation produced a wide distribution extending beyond the Marsh score category scale and suggested that the treatment outcomes were heterogeneous for these patients.

The before and after biopsies from (TM1 and TF6) demonstrated the recovery of these patients from an active CD state (Marsh score 3) towards Marsh score 1. The tracking of patients with before and after treatment biopsies clearly demonstrated the potential high resolution utility of these classification equations to map patient recovery progression after treatment imposition.

3.2. Generation of simulated histological and gene expression data using Gaussian extrapolation of empirical patient biopsy data

Gaussian modelling of empirical CD histological [16] (Fig. 1) and gene expression [17] (Fig. 2) data were used to generate simulated data from either 100 hypothetical patients from each Marsh score category (Marsh 0, 1, 3a and 3b; total of 400 data points) or 1000 patients from each Marsh score category (Marsh 0, 1, 3a and 3b; total of 4000 data points).

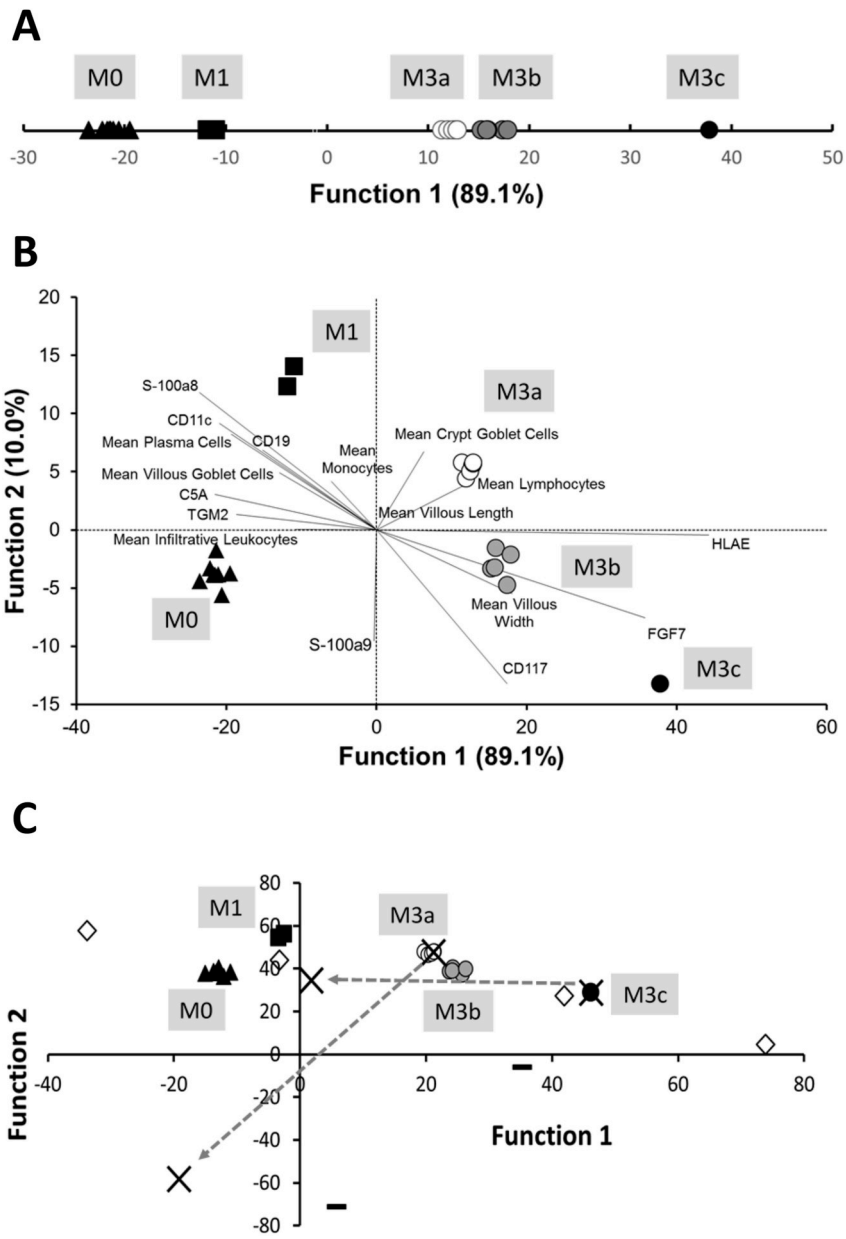

Fig. 1. Discriminant functions defined using empirical histological and gene expression data. A. The distribution of patient biopsy data into discrete Marsh score categories defined using a single discriminant function. B. The resolution of patient biopsies into discrete Marsh score categories using two discriminant functions. C. The use of two discriminant functions to define discrete Marsh score categories overlayed with biopsy data from treated patients. Symbols: M0 (▲); M1 (■); M3a (◯); M3b (●) Patient biopsy data before and after treatment imposition are indicated by crosses (×). The arrows highlight patient trajectories after treatment. Post-treatment only samples (◇). Samples from patients that received an equivocal diagnosis (−).

For the simulated gene expression dataset K-S testing indicated that a normal distribution of data was achieved for the majority of the parameters with 100 patients/study group with the exception of ALPI, DEFA5, CD86, IFNG and CD11c. A normal distribution was also achieved for all parameters in the 1000 patients/study group with the exception of only TAP1 and CD11c genes. ANOVA testing of these parameters demonstrated that there no significant differences between the group means for any of these gene expression measures (Table 6).

Table 7 showed the results of K-S normality and ANOVA analyses for the original empirically-derived treated CD histological data and the data generated for 100 simulated treated patients and 1000 simulated treated patients. K-S testing showed that a normal distribution of data was achieved for all variables excluding mean monocytes for the 100 patient group and all variables excluding mean crypt goblet cells for the 1000 patient group, although the empirically-derived dataset was also shown to not be normally distributed for both of these cases. No statistically significant differences were observed between the empirical dataset means and the simulated data means for any of the parameters except for mean villous goblet cells. Thus, these simulated gene expression and histological datasets mimicked the structure of the original empirical data [16,17] and can therefore be used to investigate the performance of the discriminant equations used to classify patient biopsies along the Marsh score scale within larger datasets.

3.3. Using empirical data-derived discriminant equations to classify simulated histological and gene expression data

The discriminant functions (Table 1), derived using empirical gene expression and histological data were used to classify simulated histological and gene expression data defined for 100 and 1000 hypothetical patients (Fig. 2). The empirically-derived equations produced low resolution separation of the Marsh score categories thereby suggesting that these empirically-derived discriminant functions were unable to successfully classify larger patient numbers and suggest that a larger

Table 2
Actual and Predicted Marsh scores for each patient using a standard calculation.

Actual Group	Predicted Group (Standard)					
	M0	M1	M3a	M3b	M3c	Correctly Classified
M0	13	0	0	0	0	1.000
M1	0	2	0	0	0	1.000
M3a	0	0	6	0	0	1.000
M3b	0	0	0	5	0	1.000
M3c	0	0	0	0	1	1.000
Overall Correct Classification Rate						**1.000**

number of patient biopsies are required to generate discriminant equations before these equations can be used to classify patients in a clinical setting. Therefore, the larger simulated datasets were used to define discriminant functions for Marsh score classifications.

3.4. Simulated data-derived discriminant equations

Redefinition of the discriminant equations using 100 patient simulated gene expression and histological data (Table 3) generated a three function equation. All three functions made a significant ($p < 0.00001$) contribution to the discriminant power of the equation and accounted for 100% of the observed variation (function 1: 76.645%, function 2: 17.210% and function 3: 7.145%). Each of the functions had a large canonical correlation indicating that all of the functions contributed significantly to the separation of the Marsh score categories. Similarly, the 1000 patient simulated dataset also generated a discriminant equation consisting of three significant ($p < 0.00001$) functions each with a high canonical correlation (Table 3). The summation of the three functions accounted for all of the observed variance.

The distribution of the parameters that contributed to each Marsh score category classification for the 100 and 1000 parameter datasets are shown in Fig. 3A and B respectively. The three-function discriminant equation defined by the 100 patient and 1000 patient simulation data sets essentially used the same. However, slightly different parameters were used to classify the Marsh score categories. The major parameters used by the three-function equation to define normal (no specific pathology) duodenal biopsies were mean villous goblet cells, mean villous length, mean monocyte and C5A gene expression. In addition to these parameters, DEFA5 was also used by the 1000 patient simulated data to define discriminant equations that classified normal (no specific pathology) duodenal biopsies. For both the 100 and 1000 patient simulated data the Marsh score 1 category was defined by gene expression parameters in the discriminant equation such as epithelial structural genes such as CD326 (E-cadherin), cytokines such as IL15 and IL18 and immune marker genes such as CD203c and CD55. Active CD (Marsh score 3) was clearly defined by a combination of immunological histology parameters (mean lymphocytes, mean infiltrative leukocytes, mean plasma cells, mean villous width) and mean crypt goblet cells. Immune gene (IFNG, HLAE, CD117, CD19, CD11c, CD54) and non-immune gene (TGM2, TAP1, FGF7, S100a9, s100a8) expression parameters also contributed to the definition of the Marsh score 3 category. However, using only the first two functions of the equation insufficiently resolved the active CD into its subtypes (Fig. 3).

When the third significant function was included in the equation both the 100 and 1000 patient simulated datasets active CD were resolved into the Marsh score 3a and 3b sub-categories (Fig. 4). Function 1 was responsible for separating Marsh score 0 and 1 from 3 (active CD). Function 2 was responsible for further separating Marsh score 0 from 1 and Function 3 resolved active CD (Marsh score 3) into the subclassifications of 3a and 3b. Thus, all three functions were necessary for high resolution classification of biopsies along the Marsh score scale. The holdout analysis reported in Table 4 demonstrated that the three function discriminant equation was able to correctly classify the data into Marsh score 0, 1, 3a and 3b with 99% accuracy.

As a final verification of the validity of the simulated 100 and 1000 patient datasets a Mahalanobis analysis was performed to determine if there were significant differences between the centroids of the Marsh

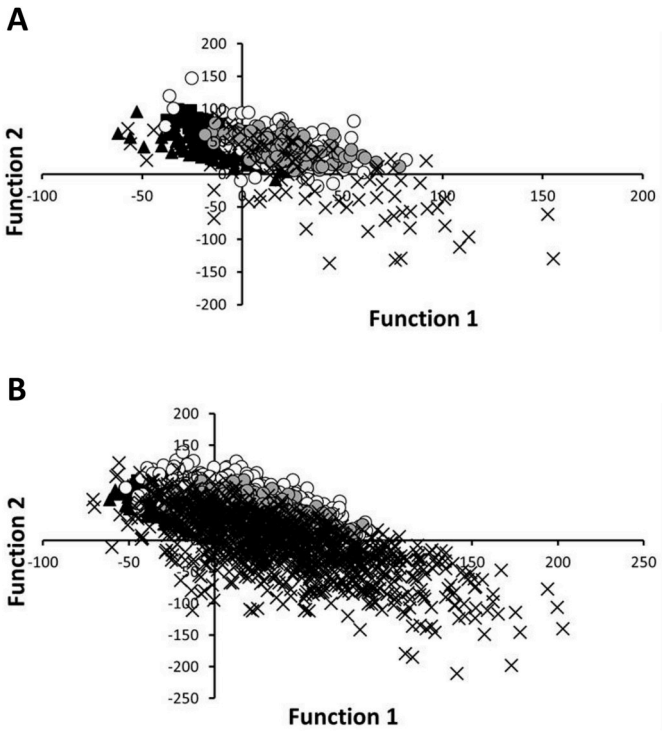

Fig. 2. Classification and resolution of (A) 100 simulated patients/group and (B) 1000 simulated patients/group datasets by discriminant functions derived from empirical histological and gene expression data. Symbols: M0 (▲); M1 (■); M3a (◯); M3b (●); Treated CD (X).

Table 3
Definition of three functions that discriminate Marsh score on the basis of significant histological and gene expression variables for 100 simulated patients/group or 1000 patients/group.

100 patients/group	Function 1	Function 2	Function 3
Eigenvalue	23.726	6.398	2.241
% of Variance	76.645	17.210	7.145
Cumulative Variance (%)	76.645	92.855	100.000
Canonical Correlation	0.980	0.919	0.832
Wilks' Lambda*	0.002	0.048	0.309
χ^2	2371.114	1152.125	446.858
DF	102	66	32
P-value	< 0.00001	< 0.00001	< 0.00001
1000 patients/group	**Function 1**	**Function 2**	**Function 3**
Eigenvalue	22.635	7.405	2.216
% of Variance	70.171	22.958	6.871
Cumulative Variance (%)	70.171	93.129	100.000
Canonical Correlation	0.979	0.939	0.830
Wilks' Lambda*	0.002	0.037	0.311
χ^2	25710.396	13122.724	4649.782
DF	102	66	32
P-value	< 0.00001	< 0.00001	< 0.00001

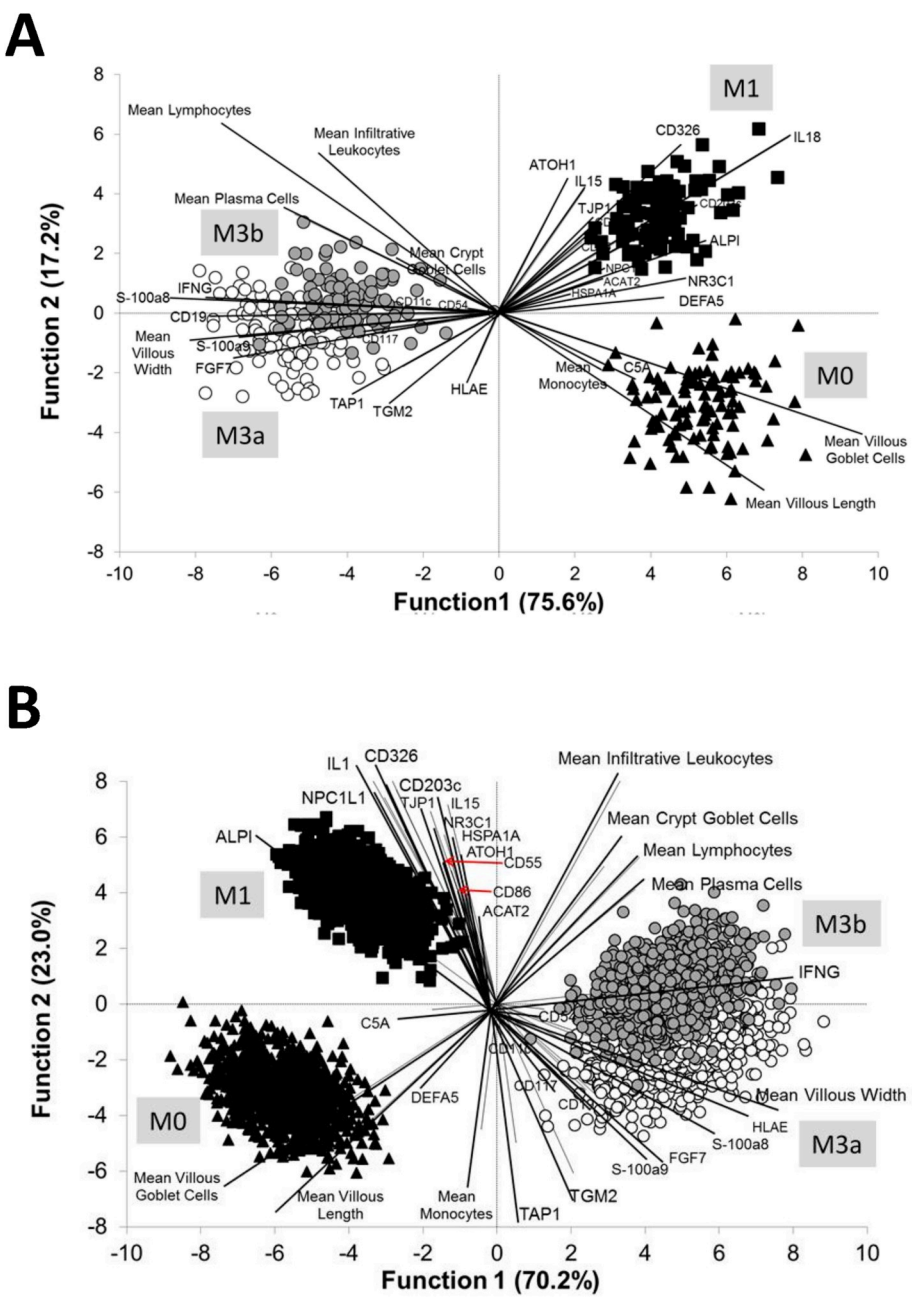

Fig. 3. Marsh score category two function discriminant plots. Discriminant functions were defined using simulated data from (A) 100 patient/group dataset or (B) 100 patient/group dataset. The plots was labelled with the eigenvectors of significant predictive parameters for each Marsh score category with the discriminant strength of each gene parameter indicated by the length of the line. Symbols: M0 (▲); M1 (■); M3a (◯); M3b (●).

score classifications derived from equations generated using the empirical data compared to the discriminant equations derived using the simulated datasets (Table 5). There were no significant differences between the centroid values using ANOVA analysis. Thus, simulated patient clustering was not significantly altered with increasing data points for generating the discriminant equations based on histological and gene expression parameters.

3.5. Classification of simulated CD patient biopsies after treatment

Fig. 5 demonstrates the separation of the Marsh score categories (M0, M1, M3a and M3b) overlayed with either the 100 or 1000 (Fig. 5A and B, respectively) simulated treated CD patients, by all three gene expression and histologically derived discriminant functions (functions 1–3). Interestingly, the majority of these simulated treated patients were plotted between the groups of Marsh type 3a/3b and Marsh type 1. As expected there was an absence of simulated treated data that intersected with the simulated non-specific pathology duodenal biopsy cohort. This suggested that removal of gluten from the diet does not result in the complete restoration of healthy mucosa.

Because most of the treated data did not coincide with the Marsh score standards, the Mahalanobis distances between empirical before and after treatment biopsies (TM1 and TF6) and the Marsh score category standards (i.e. M0, M1, M3a and M3b) were calculated (Fig. 6). After treatment, the patient biopsies had shorter distances to the M1 standard compared to the active CD standards (M3a, M3b). Thus, by determining the distance of the before and after treatment biopsies from each of the Marsh score category standards treatment efficacy has been quantified within the Marsh score continuum.

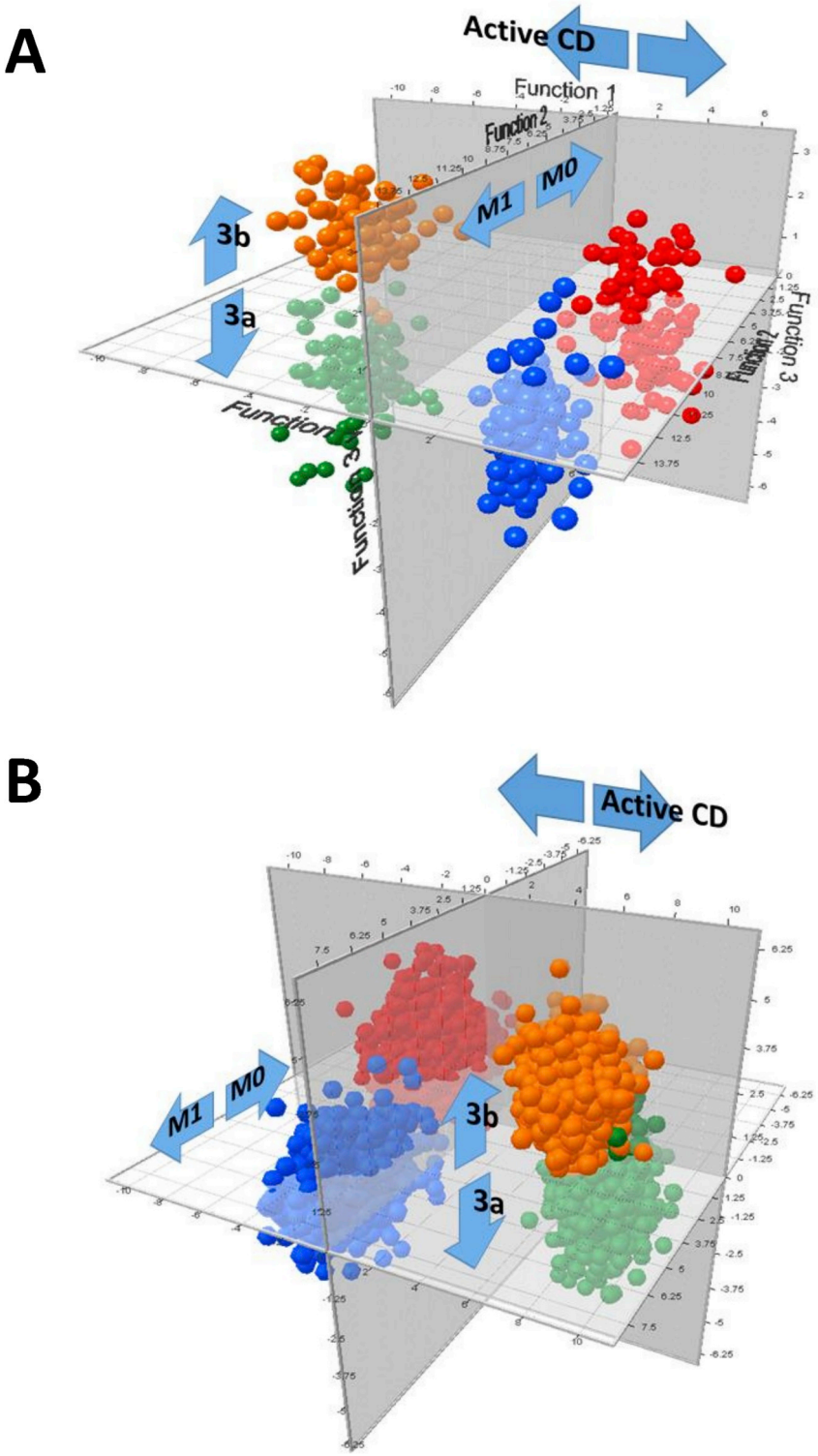

Fig. 4. Marsh score category three function discriminant plots. Discriminant functions were defined using simulated data from (A) 100 patient/group dataset or (B) 100 patient/group dataset. Symbols: M0 (red); M1 (blue); M3a (green); M3b (orange). Arrows indicate the Marsh score categories resolved by the specific discriminant functions.

4. Discussion

This study has used histological and gene expression data to define high resolution discriminant equations that correctly classified 99.6% of duodenal biopsies from CD patients into discrete Marsh score categories to create a continuum that was used to quantitatively follow patient treatment progression.

Histology and gene expression data defined discriminant equations with a small number of actual patient biopsies per Marsh score category resulted in the definition of a single function that accounted for

Table 4
Actual and predicted Marsh scores of the simulated 100 patients and 1000 patient populations using both standard and holdout calculations for the histological and gene expression equations.

Actual Group	100 Predicted Group (Standard)					100 Predicted Group (Holdout)				
	M0	M1	M3a	M3b	Correctly Classified	M0	M1	M3a	M3b	Correctly Classified
M0	100	0	0	0	100%	99	1	0	0	99%
M1	0	100	0	0	100%	0	100	0	0	100%
M3a	0	0	99	1	99%	0	0	98	2	98%
M3b	0	0	0	100	100%	0	0	1	99	99%
Overall Correct Classification Rate					**99.8%**	**Overall Correct Classification Rate**				**99%**
Actual Group	**1000 Predicted Group (Standard)**					**1000 Predicted Group (Holdout)**				
	M0	**M1**	**M3a**	**M3b**	**Correctly Classified**	**M0**	**M1**	**M3a**	**M3b**	**Correctly Classified**
M0	1000	0	0	0	100%	1000	0	0	0	100%
M1	0	1000	0	0	100%	0	1000	0	0	100%
M3a	0	0	974	26	97.4%	0	0	971	29	97.1%
M3b	0	0	1	999	99.9%	0	0	1	999	99.9%
Overall Correct Classification Rate					**99.3%**	**Overall Correct Classification Rate**				**99.3%**

Table 5
Mahalanobis testing of centroid values from the equations derived from the empirical data to the equations generated for 100 simulated patients and 1000 simulated patients based on histological and gene expression data. Overall, no significant differences were detected with ANOVA analysis between any of the centroid values ($F = 1.2$, $p = 0.417$). Please note: p-value was computed by comparing the Mahalanobis values to a $\chi 2$ distribution with the same degrees of freedom as the number of parameters, i.e. 3.

Function	Centroid (Empirical)	Centroid (100 Patients)	Centroid (1000 Patients)	Mahalanobis Distance (MD)	P-value (DF = 3)
M0 Func.1 (X)	−21.45	5.32	−5.77	3.1	0.38
M0 Func.2 (Y)	−3.79	−3.09	−3.23	2.87	0.41
M1 Func.1 (X)	−11.41	4.27	−3.62	1.22	0.75
M1 Func.2 (Y)	13.19	3.34	4.06	4.41	0.22
M3a Func.1 (X)	12.39	−5.64	4.73	3.11	0.37
M3a Func.2 (Y)	5.37	−0.70	−1.49	3.31	0.35
M3b Func.1 (X)	16.46	−3.95	4.66	1.46	0.69
M3b Func.2 (Y)	−3.01	0.45	0.67	1.5	0.68

approximately 89% of the observed variation of the experimental data. Once the position of the discrete Marsh score categories were defined, biopsies from patients after various periods of treatments were overlayed onto the Marsh score scale. Empirical data from the treated biopsies spanned the Marsh score plot and extended beyond the extremities of the Marsh score classification standards (M0, M1, M3a and M3b). Therefore, this single function was considered inadequate for classification of a single biopsy from patients undergoing treatment. However, it was clearly demonstrated that treatment progression could be monitored using a discriminant equation if before and after treatment biopsies were interrogated. Therefore, it was concluded that to monitor treatment progression, higher resolution classification equations were required to accurately assign patient biopsies into Marsh score categories.

Thus, larger random Gaussian datasets were generated based on CD histological and gene expression empirical data [16,17]. The structures of these larger simulated datasets were demonstrated to be statistically similar to the empirical data that were used to produce them. Hence, these larger simulated datasets were used to define the Marsh score classification equations. Definition of discriminant equations using simulated gene expression and histological parameters produced three significant functions that accounted for 100% of the observed variation in the simulated datasets and correctly classified 99% of the data. Even though these equations were able to classify non-CD and mild (M1) CD data correctly, they only had enough resolution power to correctly classify Marsh score 3a data from 3b in 97% of cases.

Interestingly, each of the three discriminant functions stratified the Marsh score categories in a binary fashion. For example, function 1 resolved active CD (M3 and 3b) from mild (M1) and non-CD (M0) biopsies. Other studies that have attempted to quantitatively classify CD were only able to classify between active CD and normal (no specific pathology) intestinal tissue [11–16]. Function 2 was responsible for separating mild CD (M1) from non-CD (M0) biopsies. This was achieved by using both histological and gene expression parameters. Histological parameters were the major determinants for resolving non-CD biopsies with no specific pathology (M0), whereas, gene expression parameters including epithelial structural genes, epithelial differentiation genes and innate immune genes were responsible for defining the mild CD (M1) category. Finally, function 3 was responsible for resolving the subclassifications of active CD (M3) into a and b subcategories, respectively.

The three discriminant functions were able to resolve biopsy data into the discrete Marsh score categories, which when overlayed with simulated treated patient data can potentially be used to categorise the treated patient samples into discrete Marsh score categories. However, only a minor proportion of treated samples were co-localised with the discrete Marsh score categories and suggested that biopsies from treated patients existed as a continuum. Therefore, the calculation of the mathematical distances of each treatment biopsy data point from the centroid of each of the Marsh score category standards enabled the quantitative assignment of the Marsh score indices to specific treatment biopsies. Calculation of the Mahalanobis distance seemed most appropriate since it is a data driven metric that can ease the distance distortion caused by a linear combination of attributes [20–22]. Therefore, by deploying this analysis methodology the potential exists to convert the semi-quantitative, discrete Marsh score CD scale into a high resolution continuum that can be used to more accurately assess treatment efficacy of CD patients.

However, before this can be considered as an objective quantitative scale several important caveats must be addressed. First, definition of these discriminant equations was based on gene expression and histological data reported in studies with extremely small sample sizes for each of the Marsh score categories. Analysis of the simulated Gaussian datasets reported in this current study has demonstrated that a larger number of biopsy samples for each Marsh score category is required to

Table 6
Kolmogorov-Smirnov (K/S) analysis of normality and ANOVA testing for the original treated CD and the hypothetical treated CD histological data generated for 100 patients and 1000 patients. Non-normally distributed parameters were shaded.

Parameter	Patient Number	KS Statistic (D)6	P-value	Normally Distributed?	ANOVA F Statistic	ANOVA p value
Mean Infiltrative Leukocytes	6	0.155	0.200	Yes	0.969	0.380
	100	0.294	0.115	Yes		
	1000	0.221	0.200	Yes		
Mean Villous Length	6	0.203	0.200	Yes	2.307	0.100
	100	0.219	0.200	Yes		
	1000	0.225	0.200	Yes		
Mean Villous Width	6	0.274	0.181	Yes	2.838	0.059
	100	0.321	0.053	Yes		
	1000	0.217	0.200	Yes		
Mean Crypt Goblet Cells	6	0.405	0.003	No	2.209	0.110
	100	0.227	0.200	Yes		
	1000	0.331	0.039	No		
Mean Villous Goblet Cells	6	0.268	0.200	Yes	5.992	0.003
	100	0.173	0.200	Yes		
	1000	0.157	0.200	Yes		
Mean Lymphocytes	6	0.252	0.200	Yes	0.024	0.977
	100	0.193	0.200	Yes		
	1000	0.195	0.200	Yes		
Mean Monocytes	6	0.365	0.012	No	0.695	0.499
	100	0.370	0.010	No		
	1000	0.280	0.154	Yes		
Mean Plasma Cells	6	0.253	0.200	Yes	0.279	0.756
	100	0.237	0.200	Yes		
	1000	0.292	0.121	Yes		

Table 7
Kolmogorov-Smirnov (K/S) analysis of normality and ANOVA testing for the original and the simulated gene expression data generated for 100 patients/group (n = 400 total) and 1000 patients/group (n = 4000 total). Non-normally distributed parameters were shaded.

Parameter	Patient Number	KS Statistic (D)26	P-value	Normally Distributed	ANOVA F Statistic	ANOVA p value
	26	0.145	0.167	Yes		
CD203c	400	0.111	0.200	Yes	0.164	0.849
	4000	0.082	0.200	Yes		
	26	0.152	0.123	Yes		
IL18	400	0.095	0.200	Yes	0.161	0.852
	4000	0.140	0.200	Yes		
	26	0.157	0.100	Yes		
NR3C1	400	0.077	0.200	Yes	0.453	0.636
	4000	0.111	0.200	Yes		
	26	0.270	0.000	No		
ALPI	400	0.225	0.002	No	0.618	0.639
	4000	0.114	0.200	Yes		
	26	0.141	0.198	Yes		
NPC1L1	400	0.131	0.200	Yes	0.286	0.751
	4000	0.127	0.200	Yes		
	26	0.096	0.200	Yes		
IL15	400	0.088	0.200	Yes	0.968	0.380
	4000	0.107	0.200	Yes		
	26	0.148	0.149	Yes		
DEFA5	400	0.166	0.063	No	0.294	0.745
	4000	0.131	0.200	Yes		
	26	0.136	0.200	Yes		
TJP1	400	0.137	0.200	Yes	1.188	0.305
	4000	0.074	0.200	Yes		
ATOH1	26	0.147	0.154	Yes	0.386	0.680

(*continued on next page*)

Table 7 (*continued*)

	400	0.077	0.200	Yes		
	4000	0.124	0.200	Yes		
	26	0.230	0.001	No		
CD55	400	0.091	0.200	Yes	1.545	0.213
	4000	0.123	0.200	Yes		
	26	0.247	0.000	No		
HSPA1A	400	0.096	0.200	Yes	0.885	0.413
	4000	0.125	0.200	Yes		
	26	0.138	0.200	Yes		
CD326	400	0.122	0.200	Yes	0.298	0.743
	4000	0.076	0.200	Yes		
	26	0.201	0.008	No		
CD86	400	0.167	0.061	No	0.482	0.618
	4000	0.087	0.200	Yes		
	26	0.208	0.005	No		
ACAT2	400	0.100	0.200	Yes	0.202	0.818
	4000	0.086	0.200	Yes		
	26	0.143	0.185	Yes		
CD54	400	0.100	0.200	Yes	0.173	0.841
	4000	0.108	0.200	Yes		
	26	0.395	0.000	No		
TAP1	400	0.131	0.200	Yes	0.093	0.911
	4000	0.221	0.002	No		
	26	0.164	0.071	Yes		
TGM2	400	0.072	0.200	Yes	0.051	0.950
	4000	0.113	0.200	Yes		
IFNG	26	0.272	0.000	No	0.588	0.556

(*continued on next page*)

Table 7 (*continued*)

	400	0.176	0.037	No		
	4000	0.136	0.200	Yes		
	26	0.226	0.001	No		
CD19	400	0.100	0.200	Yes	0.730	0.482
	4000	0.133	0.200	Yes		
	26	0.180	0.030	No		
CD11c	400	0.246	0.000	No	0.198	0.821
	4000	0.241	0.000	No		
	26	0.228	0.001	No		
S-100a8	400	0.155	0.107	Yes	0.287	0.751
	4000	0.128	0.200	Yes		
	26	0.203	0.007	No		
S-100a9	400	0.128	0.200	Yes	0.264	0.768
	4000	0.180	0.030	No		
	26	0.127	0.200	Yes		
FGF7	400	0.098	0.200	Yes	0.069	0.933
	4000	0.129	0.200	Yes		
	26	0.172	0.047	No		
CD117	400	0.117	0.200	Yes	0.288	0.750
	4000	0.132	0.200	Yes		
	26	0.229	0.001	No		
HLAE	400	0.089	0.200	Yes	0.613	0.542
	4000	0.111	0.200	Yes		
	26	0.158	0.096	Yes		
C5A	400	0.099	0.200	Yes	0.937	0.392
	4000	0.114	0.200	Yes		

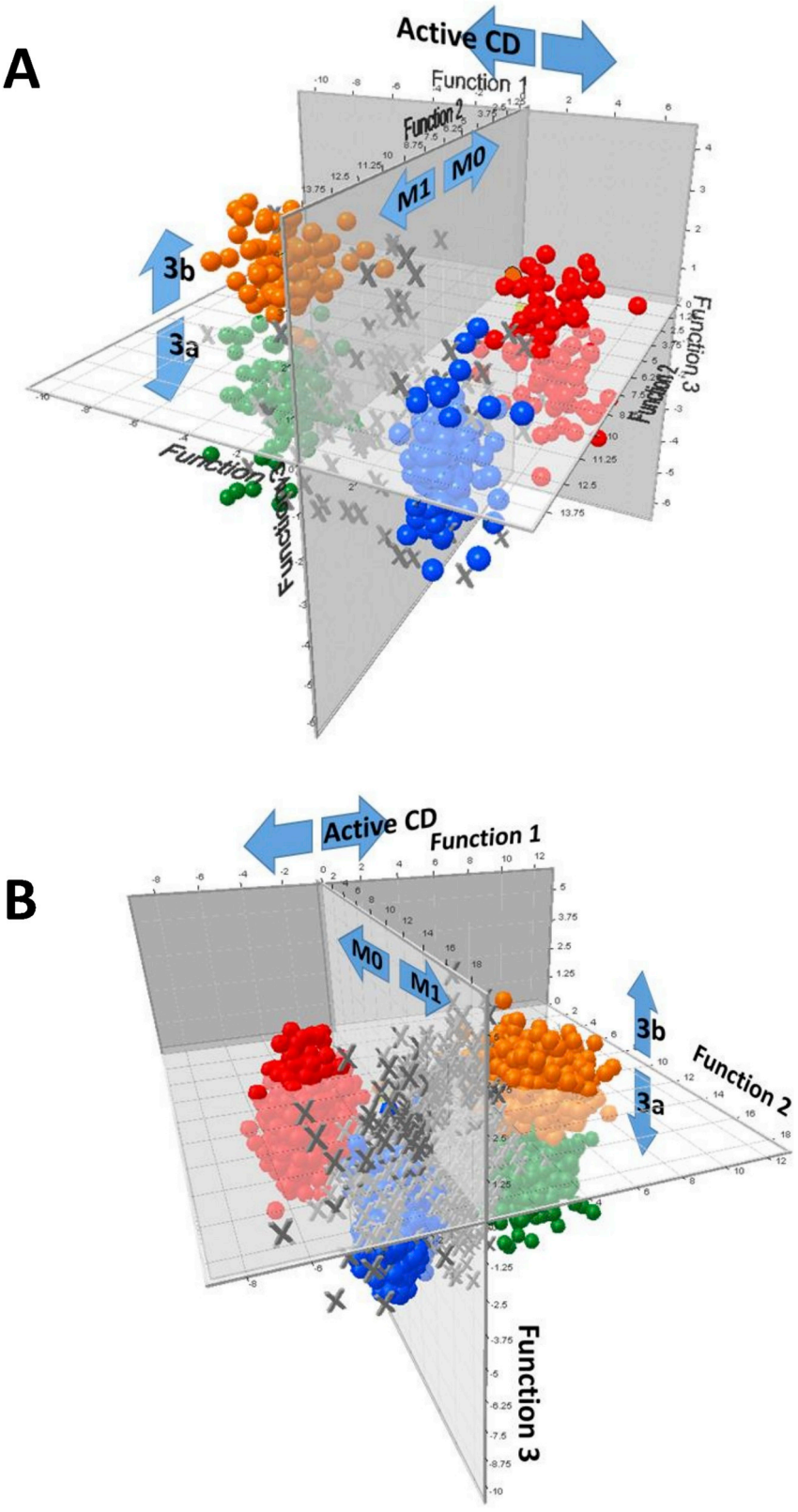

Fig. 5. Marsh score category three function discriminant plots overlayed with simulated treated patient data. Discriminant functions were defined using simulated data from (A) 100 patient/group dataset or (B) 100 patient/group dataset. Symbols: M0 (red); M1 (blue); M3a (green); M3b (orange); simulated treated patient data (grey). Arrows indicate the Marsh score categories resolved by the specific discriminant functions.

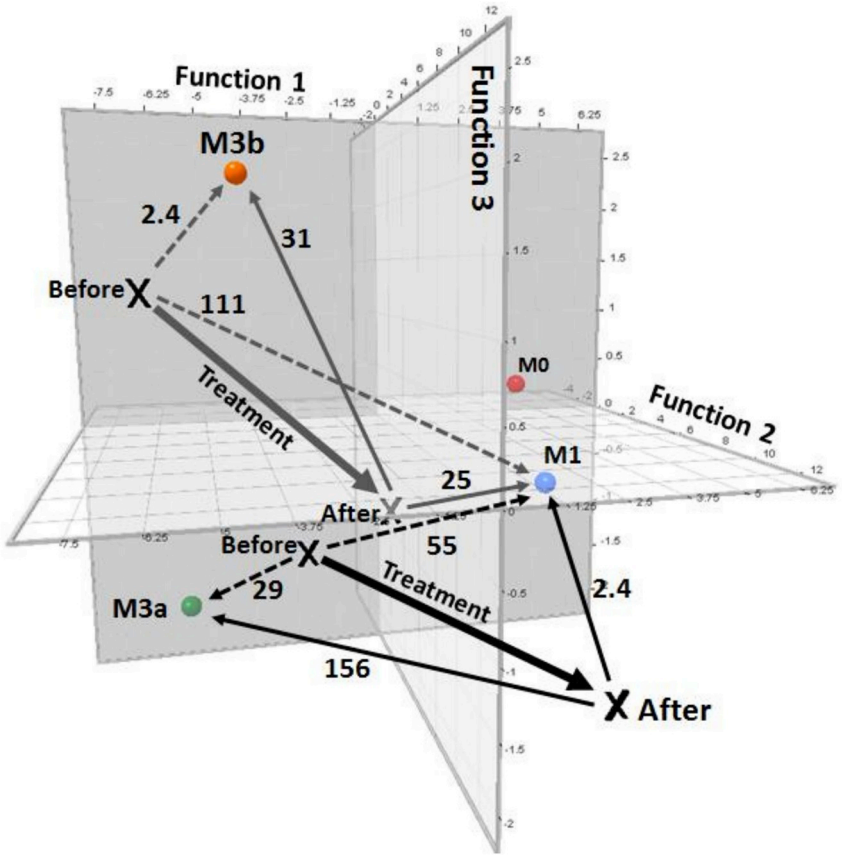

Fig. 6. Mahalanobis distances of two patients with before and after treatment biopsies overlayed on the Marsh score category three function discriminant centroid plot of the 100 patient/group dataset. Symbols: M0 (red); M1 (blue); M3a (green); M3b (orange); simulated treated patient data (black). Broken arrows indicate the Mahalanobis distances of the before treatment biopsy to Marsh score standards. Unbroken arrows indicate the Mahalanobis distances of the after treatment biopsy to Marsh score standards. The thicker unbroken arrow defined the direction of biopsies movement due to treatment.

define useful high resolution discriminant equations. Second, the experimental histological [16] and gene expression data [17] were derived from a patient cohort that did not contain all Marsh score categories (i.e. Marsh score 2 and 3c were notably absent or under-represented, respectively). Therefore, before a quantitative Marsh score continuum can be realised enough biopsies representative of all Marsh score categories must be used to define the discriminant equations. Finally, the specific gene expression signature of CD must be determine to facilitate the accurate classification of CD unambiguously from other intestinal inflammation pathologies.

Nevertheless, this study has used histological and gene expression data from empirical pilot studies to define simulated datasets which were used to determine high resolution discriminant equations able to quantitatively define the discrete Marsh score categories (M0, M1 and M3a and M3b). In addition, statistical analysis of the discrete Marsh score scale defined using gene expression and histological data generated equations may potentially convert the semi-quantitative, discrete Marsh score scale into a quantitative high definition continuum capable of accurately monitoring patient treatment progression.

Disclosures

The authors have no financial, professional, or personal conflicts of interest relevant to the manuscript to declare.

References

[1] A. Alaedini, P.H. Green, Narrative review: celiac disease: understanding a complex autoimmune disorder, Ann. Intern. Med. 142 (2005) 289–298.
[2] L.M. Sollid, Molecular basis of celiac disease, Annu. Rev. Immunol. 18 (2000) 53–81.
[3] C. Catassi, S. Gatti, A. Fasano, The new epidemiology of celiac disease, J. Pediatr. Gastroenterol. Nutr. 59 (2014) S7–S9.
[4] C. Catassi, I. Bearzi, G.K.T. Holmes, Association of celiac disease and intestinal lymphomas and other cancers, Gastroenterology 128 (2005) S79–S86.
[5] P.J. Ciclitira, A.L. King, J.S. Fraser, AGA technical review on coeliac sprue: American gastroenterological association, Gastroenterology 120 (2001) 1526–1540.
[6] A. Ensari, Gluten-sensitive enteropathy (celiac disease) controversies in diagnosis and classification, Arch. Pathol. Lab Med. 134 (2010) 826–836.
[7] M. Walker, J. Murray, An update in the diagnosis of coeliac disease, Histopathology 58 (2010) 1–14.
[8] G. Oberhuber, G. Granditsch, H. Vogelsong, The histopathology of coeliac disease: time for a standardised report scheme for pathologists, Eur. J. Gastroenterol. Hepatol. 11 (1999) 1185–1194.
[9] G.R. Corazza, V. Villanacci, Coeliac disease: some considerations on the histological diagnosis, J. Clin. Pathol. 58 (2005) 573–574.
[10] M.N. Marsh, Gluten, major histocompatibility complex and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('Celiac sprue'), Gastroenterology 102 (1992) 330–354.
[11] L. Greco, R. Troncone, B. De Vizia, et al., Discriminant analysis for the diagnosis of childhood celiac disease, J. Paed. Gastro. Nutr. 6 (1987) 538–542.
[12] M. Mayer, L. Greco, R. Troncone, et al., Early prediction of relapse during gluten challenge in childhood celiac disease, J. Paed. Gastro. Nutr. 8 (1989) 474–479.
[13] C. Catassi, M. Rossini, I.M. Rätsch, et al., Dose dependent effects of protracted ingestion of small amounts of gliadin in coeliac disease: a clinical and jejunal morphometric study, Gut 34 (1993) 1515–1519.
[14] E.J. Ciaccio, C.A. Tennyson, S.K. Lewis, et al., Distinguishing patients with celiac disease by quantitative analysis of video capsule endoscopy images, Comput. Methods Progr. Biomed. 100 (2010) 39–48.
[15] R. Troncone, A. Franzese, G. Mazzarella, et al., Gluten sensitivity in a subset of children with insulin dependent diabetes mellitus, Am. J. Gastroenterol. 98 (2003) 590–595.
[16] R.P.G. Charlesworth, N.M. Andronicos, D.R. Scott, et al., Can the sensitivity of the histopathological diagnosis of coeliac disease be increased and can treatment progression be monitored using mathematical modelling of histological sections?–A

pilot study, Adv. Med. Sci. 62 (2017) 136–142.
[17] Charlesworth RPG, Agnew LL, Scott DR, et al. Coeliac disease gene expression data can be used to classify biopsies along the Marsh score severity scale. J Gasterenterol Hepatol. doi: https://doi.org/10.1111/jgh.14369.
[18] L. Martino, D. Luengo, J. Miguez, Efficient sampling from truncated bivariate Gaussians via Box- müller transformation, Electron. Lett. 48 (2012) 1533.
[19] W.J. Thistleton, J.A. Marsh, K. Nelson, C. Tsallis, Generalised box-müller method for generating Gaussian random derivatives, information theory, IEEE Transactions on 53 (2007) 4805–4810.
[20] P. Legendre, L. Legendre, Numerical Ecology, Elsevier, 2012.
[21] A.S. Shirkhorshidi, S. Aghabozorgi, T.Y. Wah, A comparison study on similarity and dissimilarity measures in clustering continuous data, PLoS One 10 (2015) e0144059 https://doi.org/10.1371/journal.pone.0144059.
[22] P.C. Mahalanobis, On the generalised distance in statistics, Proc. Natl. Inst. Sci. India 2 (1936) 49–55.