

OPEN

Immunological Profiling of Paediatric Inflammatory Bowel Disease Using Unsupervised Machine Learning

AQ2 ^{*†}Tracy Coelho, ^{*Enrico Mossotto}, [‡]Yifang Gao, [§]Rachel Haggarty, ^{*†}James J. Ashton, [†]Akshay Batra, ^{*¶}Imogen S. Stafford, [†]Robert M. Beattie, [‡]Anthony P. Williams, and ^{*Sarah Ennis}

ABSTRACT

Objectives: The current classification of inflammatory bowel disease (IBD) is based on clinical phenotypes, which is blind to the molecular basis of the disease. The aim of this study was to stratify a treatment-naïve paediatric IBD cohort through specific innate immunity pathway profiling and application of unsupervised machine learning (UML).

Methods: In order to test the molecular integrity of biological pathways implicated in IBD, innate immune responses were assessed at diagnosis in 22 paediatric patients and 10 age-matched controls. Peripheral blood mononuclear cells (PBMCs) were selectively stimulated for assessing the functionality of upstream activation receptors including NOD2, toll-like receptor (TLR) 1-2 and TLR4, and the downstream cytokine responses (IL-10, IL-1 β , IL-6, and TNF- α) using multiplex assays. Cytokine data generated were subjected to hierarchical clustering to assess for patient stratification.

Results: Combined immune responses in patients across 12 effector responses were significantly reduced compared with controls ($P=0.003$) and driven primarily by “hypofunctional” TLR responses (P values 0.045, 0.010, and 0.018 for TLR4-mediated IL-10, IL-1 β , and TNF- α , respectively; 0.018 and 0.015 for TLR1-2-mediated IL-10 and IL-1 β). Hierarchical clustering generated 3 distinct clusters of patients and a fourth group of “unclustered” individuals. No relationship was observed between the observed immune clusters and the clinical disease phenotype.

Conclusions: Although a clinically useful outcome was not observed through hierarchical clustering, our study provides a rationale for using an UML approach to stratify patients. The study also highlights the predominance of hypo-inflammatory innate immune responses as a key mechanism in the pathogenesis of IBD.

Key Words: immunological profiling, inflammatory bowel disease, unsupervised machine learning

(*JPGN* 2020;70: 00–00)

Inflammatory bowel disease (IBD) is a cytokine-mediated auto-inflammatory condition, constituting 3 main clinical phenotypes, Crohn disease (CD), ulcerative colitis (UC), and IBD-unclassified

What Is Known

- Dysregulated immune responses are the key drivers of inflammation in inflammatory bowel disease.
- The current clinical classification of inflammatory bowel disease is blind to the underlying immunological phenotype or the molecular basis of the disease.
- Treatment strategies in inflammatory bowel disease are largely guided by the clinical phenotype, disease progression, and course.

What Is New

- Children with inflammatory bowel disease can be stratified into sub-groups or clusters based on their immune response profiles using an unsupervised machine learning approach.
- A predominance of hypo-inflammatory toll-like receptor responses is seen in a significant number of paediatric inflammatory bowel disease patients at diagnosis.

(IBDU) (1). The aetiology of this polygenic immune-mediated disease is complex with a combination of multiple factors converging to produce similar clinical phenotypes. Current treatment protocols are driven by clinical phenotypes, which are blind to the molecular basis of the disease, resulting in sub-optimal outcomes and escalation of treatment strategies, sometimes necessitating invasive surgical procedures with consequential life-long morbidity (2,3). Identifying sub-groups of patients with specific immune defects will facilitate targeted therapeutics based on the molecular characterisation and stratification of the disease.

Received November 28, 2019; accepted March 17, 2020.

AQ3 From the ^{*}Human Genetics and Genomic medicine, University of Southampton, the [†]Department of Paediatric Gastroenterology, the [‡]Cancer Sciences Division, Faculty of Medicine, the [§]NIHR, Southampton Biomedical Research Centre, University Hospital Southampton, and the [¶]Institute for Life Sciences, University of Southampton, Southampton, UK.

Address correspondence and reprint requests to Tracy Coelho, PhD, Human Genetics & Genomic Medicine, Duthie Building (Mp 808), University Hospital Southampton Foundation Trust, Southampton SO16 6YD, UK (e-mail: t.f.coelho@soton.ac.uk).

AQ4

Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal’s Web site (www.jpgn.org).

This work was supported by Crohn’s in Childhood Research Association (CICRA), UK.

Drs Tracy Coelho and Enrico Mossotto are joint first authors. The authors report no conflicts of interest.

Copyright © 2020 The Author(s). Published by Wolters Kluwer Health, Inc. on behalf of the European Society for Pediatric Gastroenterology, Hepatology, and Nutrition and the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition. This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

DOI: 10.1097/MPG.0000000000002719

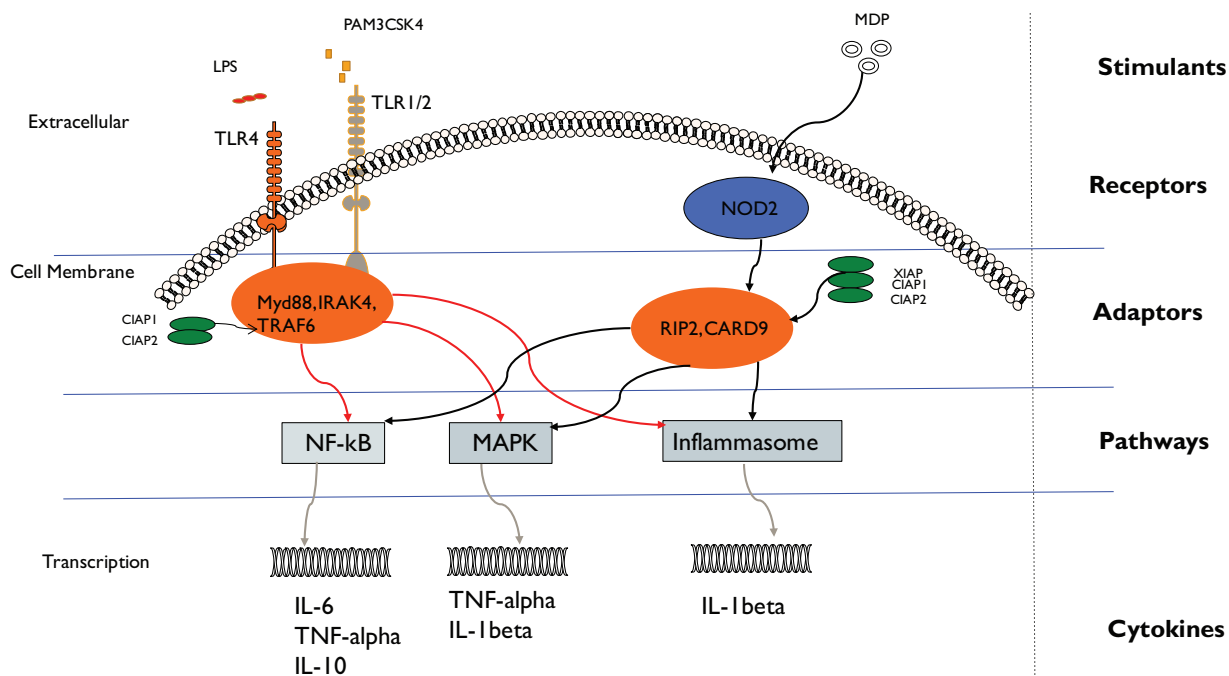


FIGURE 1. Conceptual framework of the immunological assay. PS binds to TLR4, Pam3CSK4 to TLR1/2 and MDP to NOD2. Binding of the ligands to the respective receptors result in the recruitment of adaptor proteins, which in turn activate the signalling pathways including NF- κ B, MAPK and the NLRP3-inflammasome, leading to transcription of effector cytokines. The adaptor proteins including IAPs (XIAP, cIAP1, and cIAP2) play a key role in signal transduction by engaging with other proteins for processes, such as ubiquitination and providing a scaffold for recruiting other effector kinases downstream. cIAP = cellular inhibitors of apoptosis; IRAK4 = interleukin-1 receptor associated kinases 4; MAPK = mitogen-activated protein kinases; Myd88 = myeloid differentiation primary response protein 88; NF- κ B = nuclear factor- κ B; RIP2 = receptor interacting response protein 2; TRAF6 = TNF-receptor associated factor 6; XIAP = X-linked inhibitor of apoptosis.

The primary aim of this study was to profile specific innate immune responses in peripheral blood mononuclear cells (PBMCs) in a treatment-naïve paediatric IBD cohort and assess if patients could be stratified based on their induced cytokine responses. The reason for using treatment-naïve samples was to minimise the masking impact of drug interventions on immune responses.

The cytokine data generated were subjected to hierarchical clustering to assess patient-stratification based on immune response profiles. Hierarchical clustering is a well-researched application of unsupervised machine learning, which seeks to identify natural groupings within a complex data set by reducing variable dimensionality and clustering data points with similar patterns (1).

The study focused on the functional integrity of innate signalling pathways of known biological importance in IBD mediated via the NOD2, toll-like receptor (TLR)-1-2, and TLR4 (4,5), measuring effector cytokine responses using multiplex assays. Figure 1 provides a schematic representation of the molecular basis of the assay.

NOD2 receptor is an intracellular pathogen recognition receptor (PRR), which senses specific conserved fragments, such as muramyl dipeptide (MDP), a peptidoglycan motif found in the cell wall of several Gram-positive and Gram-negative bacteria (6). Upon recognition of MDP, NOD2 receptor sub-units undergo self-oligomerisation, recruiting receptor-interacting serine-threonine-protein kinase 2 (RIPK2) through their caspase recruitment domain (CARD). Further downstream, RIPK2 mediates recruitment of other adaptor proteins and kinases, which are essential for the activation of the nuclear factor- κ B (NF- κ B), mitogen-activated protein kinases (MAPK), and the Nod-like receptor family pyrin domain 3 (NLRP3)-inflammasome cascades (7). Unlike NOD2,

which is an intracellular receptor, TLRs, such as TLR1, TLR2, and TLR4 are cell-surface receptors. TLRs recognize pathogen-associated molecular patterns (PAMPs), such as lipids, lipoproteins, and other microbial products. Mutations in several TLR genes have been implicated in the pathogenesis of IBD. The most well-studied among the TLRs for their association with IBD include TLR2 and TLR4 (8,9). Following interaction with their corresponding stimulants, TLRs interact with adaptors, such as myeloid differentiation primary response protein 88 (MyD88) and TIR-domain-containing adaptor-inducing interferon- β (TRIF), which then activate downstream-signalling cascades. Activation of signalling cascades, such as NF- κ B, MAPK, and NLRP3-mediated inflammasome activation pathways results in the transcription of cytokines including IL-10, IL-6, IL-1 β , and TNF- α , which can be quantified through multiplex assays (10,11). Cytokine analysis following activation of specific receptors indicates the functionality of proteins involved in these pathways, from the stimulated receptor to the specific cytokine produced (10,11).

METHODS

Study Participants

Children under 18 years of age, with suspected IBD were recruited before an established diagnosis over a 12-month period from January 2016 to January 2017. Twenty-two patients diagnosed with IBD following an upper and lower gastrointestinal (GI) endoscopy were included in the study. The diagnosis was made in line with the ESPGHAN revised Porto criteria for IBD (12). Specimens from 10 children with normal endoscopic and histological assessment of the GI biopsies were used as non-IBD controls.

The non-IBD control group underwent investigations in view of nonspecific GI symptoms but had normal investigations, remained well at 6 months follow-up and were subsequently discharged from the service. Blood samples for PBMC extraction from all participants were obtained before commencement of any treatment.

Peripheral Blood Mononuclear Cell Extractions, Activation, and Luminex Assays

PBMCs were extracted by density gradient centrifugation (Ficoll-Paque Plus, GE Healthcare, Uppsala, Sweden) using up to 10 mL of whole blood per individual, collected in lithium heparin vacutainer tubes. Isolated PBMCs were cryopreserved in liquid nitrogen. For activation, PBMCs were used at a concentration of 0.3×10^5 cells, stimulated with Pam3CysSerLys4 (Pam3CSK4; TLR1-2 receptor stimulant; Invivogen, UK; dose- 10 µg/mL), lipopolysaccharide (LPS; TLR4 stimulant; Sigma-Aldrich, UK; dose- 1 µg/mL), and muramyl-dipeptide (MDP; NOD2 stimulant; Invivogen, UK; dose- 10 µg/mL) for 24 hours. The cellular supernatants were analysed to measure the concentration of IL-10, IL-1β, IL-6, and TNF-α using Luminex assays as per the manufacturer's protocols (Invitrogen, User Manual: Human Cytokine 10-Plex Panel. Catalog no. LHC0001, CA). Luminex assays are flow cytometry-based multiplexing assays, measuring several analytes of interest using spectrally encoded magnetic beads, each with a specific fluorescent signature (13). All samples were assayed in duplicate. Monocyte counts were assessed using whole blood in every individual and the cytokine measurements standardised to a monocyte count of 10%. As there were 3 activating stimulants and 4 cytokines analysed, a total of 12 cytokine responses were assessed per individual.

NOD2 Variants

Patients were assessed for *NOD2* variants using exome sequencing data. Any variation within the *NOD2* gene was extracted from the variant call files generated for each of the pIBD patients and categorised in line with the American College of Medical Genetics (ACGM) guidance to remove “benign” variants and identify “pathogenic” and “likely pathogenic” variants (14). Further details on the sequencing analysis are presented in the online supplementary material (Supplemental Digital Content, <http://links.lww.com/MPG/B813>).

Statistical Analysis and Cytokine Data Normalisation

Statistical analysis was performed using the GraphPad Prism software, version 7 and SPSS (version 25, IBM). Cytokine induction between the patient cohort and controls were compared using unpaired *t*-tests (2-tailed). For the application of machine learning algorithms, raw cytokine data expressed in a conventional format (picograms/mL) were normalised using RobustScaler and StandardScaler within the python Scikit-Learn package (15). Further details on data transformation and scaling statistics of these software tools are presented in the online supplementary material (Supplemental Digital Content, <http://links.lww.com/MPG/B813>).

Hierarchical Clustering

Hierarchical clustering was performed on the normalised data using the R software package (16). An agglomerative hierarchical clustering (bottom-up approach) algorithm was used, which applies pairwise distance matrix between the observations as

clustering criteria. The distances were calculated using Euclidean metrics and average linkage. A tree-based representation of agglomerative clustering starts at the bottom by treating each individual as a singleton cluster and at each step of the algorithm, the 2 individuals with the most similar profiles are merged, with pairs of clusters successively merged into 1 big cluster (root) to include the entire cohort of patients. In this study, clusters within the dendrogram were identified using a “static” tree-cut method, which defines each contiguous branch below a fixed height cut-off as a separate cluster (16).

Ethics Statement

Ethical approval was granted by Southampton & South West Hampshire Research Ethics Committee (09/H0504/125). Informed consent was obtained from the parents/legal guardians of all participants before recruitment to the study. Informed consent was also obtained from older children who were deemed capable of understanding the nature of the study.

RESULTS

Patients and Controls

Twenty-two paediatric patients with IBD were included; 55% male patients ($n = 12$), median age at diagnosis of 13 years (range 5–16 years); Crohn disease ($n = 14$), and ulcerative colitis ($n = 8$). The mean (\pm standard deviation [SD]) paediatric Crohn's disease activity index (PCDAI) and paediatric ulcerative colitis activity index (PUCAI) scores at the time of sample procurement were 37 (± 12.23) and 54.38 (± 12.94), respectively, reflecting a moderate-severe disease activity at diagnosis. The control group included 10 individuals, 60% boys with a median age of 14 years (range 4–16 years). For patient demographics, see Table S1 in the online supplemental digital content (Supplemental Digital Content, <http://links.lww.com/MPG/B813>).

Induced Immune Responses on a Radar Plot

Normalised cytokine values per patient per stimulus were plotted along the equi-angular spokes or radii of a radar plot, each representing a specific cytokine response (Fig. 2). Cytokine data from paediatric controls were used to define the reference ranges in the patient cohort. Values within 2SD from the mean cytokine values in controls were considered within normal response limits, and those above and below 2SD indicated hyper-inflammatory and hypo-inflammatory responses, respectively. None of the patients showed responses in the hyper-inflammatory range (> 2 SD). The majority of cytokine values fell between the mean and -2 SD, suggesting a trend towards hypo-inflammatory responses.

Hierarchical Clustering

In order to identify patterns or trends within the cytokine data across the 22 patients, hierarchical clustering of normalised cytokine profiles was used. With a static height cut-off of 1.25, the dendrogram generated 3 distinct clusters of patients based on the computed similarity of their induced immune profiles and an “unclustered” group of 10 patients (grouped as cluster 4) (Fig. 3). The clustering pattern generated using this height cut-off was also distinctly recognisable through visual inspection of the dendrogram. See Figure S1 in the supplementary (Supplemental Digital Content, <http://links.lww.com/MPG/B813>). All patients within each cluster had similar immune profiles; however, no relationship was observed between the clusters and the clinical

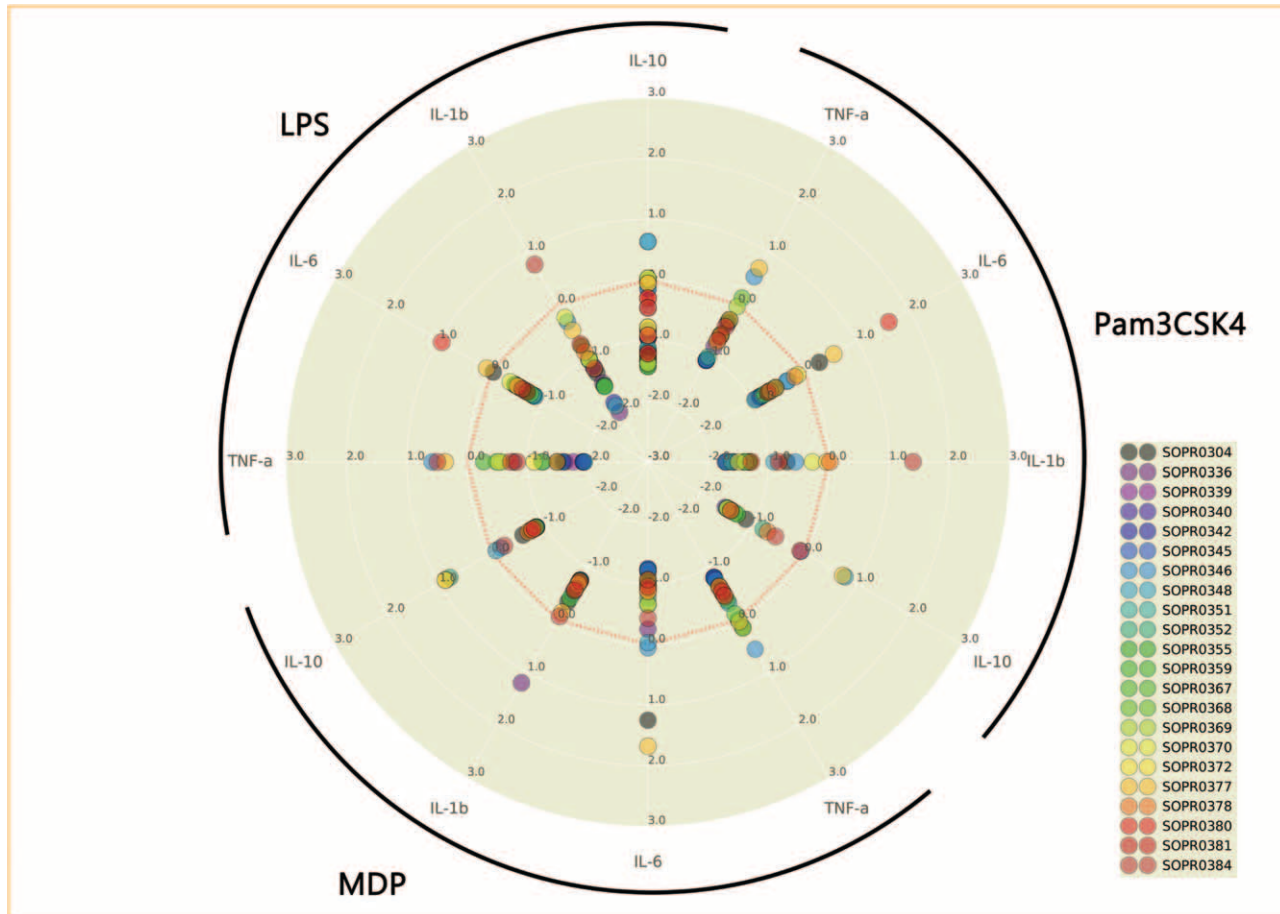


FIGURE 2. Radar plot showing cytokine responses of all patients. The ligands used for stimulation are depicted outside the boundary of each sector. The cytokines analyzed are indicated within the boundary of each sector. The red-dotted line represents the mean obtained from paediatric controls. Responses within 2 SD across the mean are within the range of normality. Every patient is represented by a unique colour. Individual control subjects are not represented in this figure. The numbers 0, 1.0, 2.0, and 3.0 along each vector indicate the mean, 1 SD, 2 SD, and 3 SD, respectively. SD = standard deviation.

disease phenotype or disease severity. Multivariable linear regression did not demonstrate a significant association between any of the 12 cytokine responses and normalised disease severity scores at diagnosis. See online supplementary data (Table S2, Supplemental Digital Content, <http://links.lww.com/MPG/B813>).

Visualisation of Cluster Patterns Using Radar Plots

In order to visualise the cytokine response patterns within individual clusters, radar plots of cytokine responses were generated per patient. Visual inspection of radar plots as groups within each cluster showed similar cytokine profiles, consistent with the stratification observed through an unsupervised machine learning approach, thereby supporting the underlying immunological basis of the clustering pattern. For example, patients in cluster 1 had comparatively reduced induction along the vectors for LPS-induced IL-1 β and TNF- α , MDP-induced IL-6 and TNF- α and Pam3CSK4-induced IL-1 β and TNF- α . The online supplemental digital content includes individual radar plots on patients in clusters 1 to 3 (Figure S2. A-C, Supplemental Digital Content, <http://links.lww.com/MPG/B813>).

Comparing Pooled Cytokine Data Between the Patient Cohort and Controls

Normalised cytokine values for all 22 patients and 10 controls are presented in Table 1. Raw cytokine data are presented in the online SDC Table S3. The final column in Table 1 represents pooled values obtained after summation of all the 12 cytokine responses per individual. There was a significant difference in the pooled responses between patients and controls ($P = 0.003$), indicating a patient cohort signal of overall reduced immune responses. The combined reduced responses were driven predominantly by the toll-like receptor (TLR)-mediated responses (TLR4-induced IL-10 [$P = 0.045$], IL-1 β [$P = 0.010$], and TNF- α [$P = 0.018$] and TLR1-2-induced IL-10 ($P = 0.018$) and IL-1 β ($P = 0.015$)). There was no significant contribution from the NOD2-mediated responses.

MDP-induced Immune Responses in the Context of NOD2 Variants

AQ5

Five NOD2 variants were identified in 8 individuals within the cohort, of which 7 individuals had CD and 1 UC. Two of the variants included R702W and G908R, widely known to be

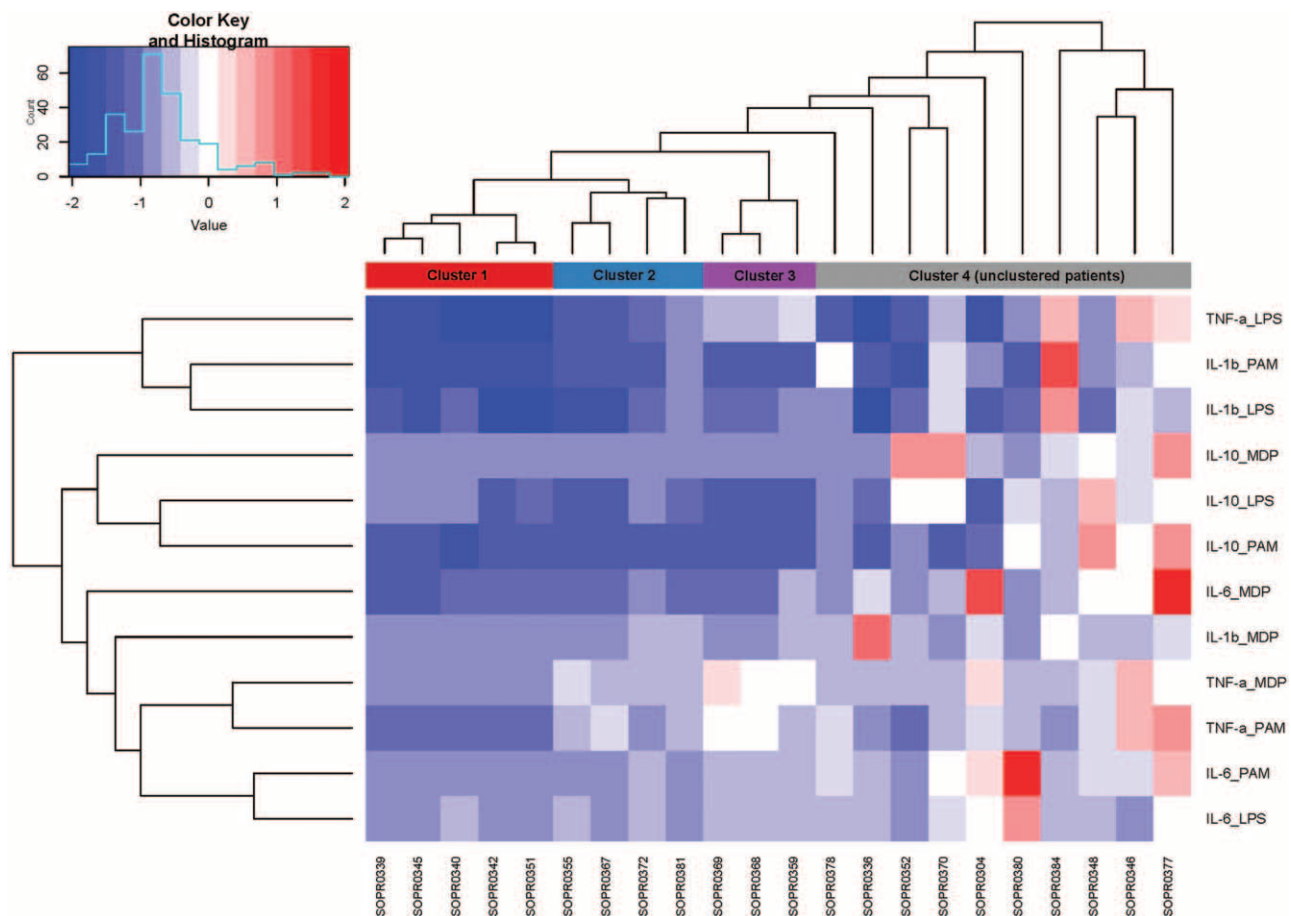


FIGURE 3. Hierarchical clustering to identify immunophenotypes. A cluster heat map based on induced immune profiles in 22 paediatric patients are shown. The maximum and minimum normalized cytokine values are indicated in red and blue, respectively, on the heat map. The probands are indicated on the x-axis and the immune profiles on the right y-axis. Distances were calculated using Euclidean metrics and average linkage. Cluster identification was performed using a static height cut-off method. Hierarchical clustering shows 3 distinct clusters of patients and an “unclustered” group of 10 patients (cluster 4).

associated with CD (17). The other variants included disease-associated mutations R708H, R713H, and V955I, categorised in line with the ACGM guidance (14). One individual with CD (SPR339) had compound heterozygosity for G908R and R708H, as confirmed through segregation analysis. MDP-induced immune responses in this individual fell between the mean and -2 SD. The remaining 7 individuals were heterozygous for the *NOD2* variants. Combined normalised cytokine values for MDP-induced immune responses showed no significant differences between the individuals with *NOD2* variants and those without. Furthermore, no significant differences were observed between the individuals with *NOD2* variants and the controls. See online supplementary data (Table S4, Supplemental Digital Content, <http://links.lww.com/MPG/B813>).

Assessing Inter-cluster Differences

In order to determine if the 5 significantly reduced cytokine responses observed across the 22 patients compared with controls were specific to individual patient clusters rather than the whole group, the within-cluster values were compared against the control group using unpaired *t*-tests. This was conducted in order to qualitatively and statistically analyse how each cluster differed from the control group in terms of their dysfunctional cytokine

responses. Cluster 1 showed significant differences for 4 out of 5, cluster 2 for 3 out of 5 cytokine responses compared with controls and cluster 3 showed only borderline significant differences for TLR1-2-mediated IL-10 induction ($P = 0.047$). See Table S5 in the online supplementary (Supplemental Digital Content, <http://links.lww.com/MPG/B813>). *t*-tests were not conducted for the “unclustered” group of patients (cluster 4).

Cytokine Profiles in Crohn Disease and Ulcerative Colitis

Normalised cytokine data between CD and UC patients were compared using unpaired *t*-tests. No significant differences were observed between the 2 phenotypes. See Table S6 in the online supplementary (Supplemental Digital Content, <http://links.lww.com/MPG/B813>).

DISCUSSION

This study adopted a unique approach of first assessing the immunological phenotype of a treatment-naïve paediatric cohort through selected innate immunity pathway profiling, followed by patient stratification through unsupervised machine learning. The

TABLE 1. Tabular matrix with normalised cytokine data

Subjects	Hierarchical clusters	TLR4 induction (with LPS)				NOD2 induction (with MDP)				TLR1-2 induction (with Pam3CSK4)				Sum	
		IL-10	IL-1 β	IL-6	TNF- α	IL-10	IL-1 β	IL-6	TNF- α	IL-10	IL-1 β	IL-6	TNF- α		
Patients	SOPR0339	1	-0.952	-1.503	-0.690	-1.768	-0.856	-0.753	-1.242	-0.807	-1.464	-1.604	-0.846	-1.062	-13.548
	SOPR0340	1	-0.945	-1.180	-0.551	-1.924	-0.845	-0.728	-1.221	-0.808	-1.519	-1.717	-0.935	-1.067	-13.440
	SOPR0342	1	-1.260	-1.872	-0.825	-1.930	-0.857	-0.749	-1.197	-0.804	-1.481	-1.695	-0.863	-1.061	-14.593
	SOPR0345	1	-0.829	-1.558	-0.805	-1.585	-0.868	-0.761	-1.248	-0.807	-1.405	-1.572	-0.842	-1.066	-13.345
	SOPR0351	1	-1.124	-1.934	-0.840	-1.949	-0.829	-0.756	-1.221	-0.759	-1.426	-1.702	-0.955	-1.055	-14.549
	SOPR0355	2	-1.427	-1.543	-0.679	-1.499	-0.845	-0.698	-0.963	-0.371	-1.459	-1.312	-0.692	-0.607	-12.095
	SOPR0367	2	-1.289	-1.568	-0.752	-1.241	-0.856	-0.744	-1.037	-0.627	-1.339	-1.319	-0.768	-0.296	-11.836
	SOPR0372	2	-0.780	-1.043	-0.489	-1.112	-0.784	-0.681	-0.885	-0.525	-1.412	-1.502	-0.661	-0.715	-10.589
	SOPR0381	2	-1.213	-0.760	-0.693	-0.830	-0.828	-0.559	-1.066	-0.549	-1.416	-0.832	-0.710	-0.607	-10.062
	SOPR0378	3	-0.913	-0.909	-0.508	-1.491	-0.708	-0.594	-0.872	-0.445	-0.710	0.018	-0.192	-0.294	-7.618
	SOPR0359	4	-1.286	-0.794	-0.563	-0.274	-0.743	-0.415	-0.656	-0.004	-1.287	-1.394	-0.661	-0.569	-8.646
	SOPR0368	4	-1.347	-1.057	-0.458	-0.489	-0.858	-0.713	-0.979	-0.115	-1.457	-1.319	-0.566	0.120	-9.238
	SOPR0369	4	-1.377	-1.057	-0.457	-0.577	-0.851	-0.717	-1.062	0.150	-1.483	-1.386	-0.555	-0.043	-9.415
	SOPR0336	5	-1.172	-2.054	-0.627	-1.929	-0.870	1.195	-0.251	-0.533	-1.480	-1.396	-0.662	-0.809	-10.587
	SOPR0348	6	0.632	-1.044	-0.448	-0.776	-0.091	-0.665	-0.029	-0.320	0.775	-0.911	-0.321	-0.382	-3.580
	SOPR0352	6	-0.002	-1.226	-0.803	-1.257	0.792	-0.564	-0.765	-0.519	-0.797	-1.615	-0.811	-1.002	-8.570
	SOPR0370	6	0.026	-0.248	-0.362	-0.523	0.872	-0.690	-0.659	-0.643	-1.482	-0.273	-0.127	-0.585	-4.696
	SOPR0346	6	-0.147	-0.331	-0.803	0.582	-0.184	-0.503	0.058	0.557	-0.071	-0.554	-0.344	0.524	-1.215
	SOPR0384	6	-0.458	0.758	-0.595	0.492	-0.251	-0.068	-0.431	-0.467	-0.558	1.395	-0.555	-0.689	-1.427
	SOPR0380	7	-0.300	-1.222	0.948	-0.726	-0.782	-0.740	-0.950	-0.635	-0.083	-1.288	1.614	-0.427	-4.590
SOPR0304	8	-1.353	-1.312	-0.038	-1.622	-0.606	-0.346	1.251	0.144	-1.127	-0.692	0.282	-0.343	-5.762	
SOPR0377	8	-0.057	-0.497	0.087	0.361	0.881	-0.151	1.680	0.031	0.730	-0.035	0.557	0.685	4.272	
Controls	SOPR0330	.	-0.691	-0.033	0.431	-0.183	-0.122	0.034	2.458	0.296	0.959	0.866	0.823	0.417	5.257
	SOPR0364	.	0.264	0.945	-0.529	0.871	2.578	2.684	-0.938	0.921	0.342	1.960	-1.015	-0.423	7.660
	SOPR0365	.	1.201	-0.409	-0.354	-0.799	1.052	-0.528	-0.067	-0.641	0.894	-0.826	-0.164	-0.861	-1.501
	SOPR0371	.	0.792	-0.639	-0.589	-0.374	-0.333	-0.615	-0.518	-0.756	1.054	0.142	-0.530	-0.861	-3.226
	SOPR0374	.	-1.405	-1.755	-0.459	-1.863	-0.833	-0.746	-1.149	-0.759	-1.501	-1.743	-0.881	-0.953	-14.044
	SOPR0375	.	-0.472	-0.816	-0.337	-0.079	-0.394	-0.516	0.142	0.034	-0.338	0.400	-0.145	2.061	-0.461
	SOPR0376	.	0.933	-0.398	0.235	0.197	-0.010	-0.493	1.061	-0.550	0.997	-0.501	0.629	-0.509	1.592
	SOPR0379	.	1.453	1.498	2.818	2.147	-0.752	-0.751	-0.240	-0.617	0.277	-0.942	2.468	0.556	7.914
	SOPR0382	.	-1.200	0.065	-0.634	0.427	-0.646	0.660	-0.367	2.565	-1.278	0.036	-0.603	1.353	0.378
	SOPR0383	.	-0.875	1.542	-0.582	-0.346	-0.540	0.271	-0.384	-0.493	-1.407	0.608	-0.582	-0.780	-3.568
<i>P</i> values		0.045	0.010	0.178	0.018	0.188	0.164	0.116	0.267	0.018	0.015	0.245	0.169	0.003	

Normalised cytokine data following stimulation with the receptor-specific ligands including LPS (TLR4 stimulant), MDP (NOD2 stimulant), and Pam3CSK4 (TLR1-2 stimulant). Induction of 4 cytokines including IL-10, IL-1 β , IL-6, and TNF- α were assessed per stimulant, thereby generating 12 assay conditions in total. The column at the extreme right represents summation of the values across all the cells in each row. The bottom row includes *P* values obtained by applying a 2-tailed student *t*-test across the values between the patient cohort and controls for the 12 cytokine responses. The *t*-tests were significantly different for 5 TLR-mediated responses (highlighted in red).

study showed a significantly reduced cytokine induction in the patient cohort compared with controls, which was driven predominantly by hypo-inflammatory TLR-mediated responses with no significant contribution from NOD2-mediated responses. No significant differences were observed in MDP-induced immune responses between individuals harbouring potentially pathogenic *NOD2* variants and those without the *NOD2* variants. Application of unsupervised machine learning approaches, such as hierarchical clustering to the cytokine data enabled stratification of patients into distinct clusters based on their induced immune profiles. Generation of radar plots per individual for visual representation of the cytokine data demonstrated similar patterns within clusters, thereby providing a supportive conceptual evidence of the immunological basis of the hierarchical clustering.

Multiple studies assessing induced immune function have been described previously, however, the vast majority have been conducted in adult IBD patients following an established diagnosis

and treatment (18). In our study, immunological assays were performed in treatment-naïve paediatric IBD patients where PBMCs were obtained at or before diagnosis, thereby removing the potential impact of drug interventions on immune responses. Previously published studies have clearly demonstrated the effect of drug treatments on the functioning of immunological assays (19,20).

The reduced cytokine induction observed in this study may reflect signalling defects in the key inflammatory pathways resulting in impaired cytokine production. Reduced or absent innate immune induction has been reported previously, particularly in the context of known IBD-associated mutations in genes, such as *NOD2* (21,22). The findings from our study, combined with the weight of evidence from previously published studies (21,22) suggest that poor or maladaptive immune response to pathogenic insults with a consequential poor bacterial clearance and persistence of secondary inflammation may be a key driving mechanism in the

pathogenesis of IBD. Another possible mechanism to explain the reduced cytokine induction observed in our patient cohort is down regulation of cytokine production in PBMCs because of the inhibitory effect of stromal factors secreted by the inflamed bowel tissue including TGF- β and IL-10 (23–25).

In this study, hypo-inflammatory immune responses were driven predominantly by reduction of TLR-mediated responses. TLRs are cell-surface or endosomal pathogen-recognition receptors, widely expressed in the GI epithelium. Alterations in TLR expression have been associated with the pathogenesis of IBD (26). TLR2 forms a hetero-dimer with TLR1 or TLR6, which is necessary for signal transduction. Current evidence indicates a protective role for TLR1-TLR2-mediated responses (27). Results have shown that poor expression of TLR1 during acute infection can lead to chronic bowel inflammation and dysbiosis. Experimental evidence suggests that a TLR2 focussed agonistic approach could hold promise for future therapeutic applications in IBD. On the other hand, although TLR4 signalling is protective against invading bacteria, it may trigger or aggravate mucosal inflammation through release of nitric oxide with consequent pro-inflammatory cytokine induction. Although the precise mechanistic influence of TLR4 on IBD pathogenesis is unknown, current evidence is indicative of a bidirectional role (27).

Although CD and UC are distinct clinical entities with possibly different immunological mechanisms, all patients in this study were grouped together with the primary objective of exploring novel groupings within the cohort based on immunological profiles regardless of the clinical diagnosis. The binary classification of IBD into CD and UC although clinically very useful, is blind to the immunological phenotype of the disease (28). It may be argued that given the complex polygenic nature of IBD and the resultant biological heterogeneity, the 2 conditions represent a phenotypic spectrum of an underlying immune dysregulation (29). Whilst there are defined diagnostic criteria for CD and UC, there is frequently a histological overlap and subsequent re-classification of the disease sub-type particularly in children (30). There is also a significant overlap in the genetic susceptibility to IBD, with patients frequently having both disease sub-types in their family history (31). Furthermore, diverse genetic mutations can present with similar clinical phenotypes adding further challenges to the understanding of the genetic nature of the disease (32).

Our study was limited by modest cohort size and profiling of a limited set of innate immune cytokine responses. To confirm wider applicability of the study findings, implementation of larger cohorts and more extensive immune profiling is required. Furthermore, as all patients had active inflammation at diagnosis when the samples were collected, it is unclear if the observed hypo-functionality is an epiphenomenon of inflammation. To follow this prospectively, it is important to demonstrate a maintained “cluster signature” by using PBMCs collected from patients during remission, whilst not on any treatment that would impact on the assay function. It would be practically challenging, however, to identify a sizable cohort of patients with controlled disease but not on immunosuppressive treatment. Although specific immune pathway defects were not identified in this study, the observed clustering pattern suggests that there is a potential for stratifying patients based on immunological profiling with a futuristic goal of implementing targeted treatments.

Despite the emergence of robust multi-omic technologies over the last few years, there is an unmet need for novel groupings in complex immune-mediated conditions like IBD. The concept of machine learning to classify IBD into novel categories based on the disease biology is exciting and a clear steer towards personalised medicine (1,31). The scope and clinical relevance of this approach in IBD is ever expanding with the advancements in the field of

immune therapies and treatment with biologics. As a futuristic model of precision medicine, patient stratification on the basis of immunological and multi-omic profiling will lead to the development of tailored treatment strategies with optimal outcomes and minimal side-effects.

Acknowledgments: The authors are grateful to all patients and their families. We thank Nikki J. Graham for technical assistance in the DNA laboratory. We also thank Dr Kathy Potter, Jenna Watt, and the technical team from tissue bank for providing consistent laboratory support.

REFERENCES

1. Mossotto E, Ashton JJ, Coelho T, et al. Classification of paediatric inflammatory bowel disease using machine learning. *Sci Rep* 2017;7:2427. **AQ6**
2. Ruemmele FM, Veres G, Kolho KL, et al., European Crohn's and Colitis Organisation; European Society of Pediatric Gastroenterology, Hepatology and Nutrition. Consensus guidelines of ECCO/ESPGHAN on the medical management of pediatric Crohn's disease. *J Crohns Colitis* 2014;8:1179–207.
3. Turner D, Levine A, Escher JC, et al., European Crohn's and Colitis Organization; European Society for Paediatric Gastroenterology, Hepatology, and Nutrition. Management of pediatric ulcerative colitis: joint ECCO and ESPGHAN evidence-based consensus guidelines. *J Pediatr Gastroenterol Nutr* 2012;55:340–61.
4. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature* 2011;474:307–17.
5. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature* 2012;491:119–24.
6. Inohara N, Ogura Y, Fontalba A, et al. Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J Biol Chem* 2003;278:5509–12.
7. Caruso R, Warner N, Inohara N, et al. NOD1 and NOD2: signaling, host defense, and inflammatory disease. *Immunity* 2014;41:898–908.
8. Wang F, Tahara T, Arisawa T, et al. Genetic polymorphisms of CD14 and Toll-like receptor-2 (TLR2) in patients with ulcerative colitis. *J Gastroenterol Hepatol* 2007;22:925–9.
9. Hume GE, Fowler EV, Doecke J, et al. Novel NOD2 haplotype strengthens the association between TLR4 Asp299gly and Crohn's disease in an Australian population. *Inflamm Bowel Dis* 2008;14:585–90.
10. Oeckinghaus A, Ghosh S. The NF-kappaB family of transcription factors and its regulation. *Cold Spring Harb Perspect Biol* 2009;1:a000034.
11. Oeckinghaus A, Hayden MS, Ghosh S. Crosstalk in NF-kappaB signaling pathways. *Nat Immunol* 2011;12:695–708.
12. Levine A, Koletzko S, Turner D, et al., European Society of Pediatric Gastroenterology, Hepatology, and Nutrition. ESPGHAN revised porto criteria for the diagnosis of inflammatory bowel disease in children and adolescents. *J Pediatr Gastroenterol Nutr* 2014;58:795–806.
13. Prabhakar U, Eirikis E, Miller BE, et al. Multiplexed cytokine sandwich immunoassays: clinical applications. *Methods Mol Med* 2005;114:223–32.
14. Richards S, Aziz N, Bale S, et al., ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17:405–24.
15. Oliphant TE. SciPy: open source scientific tools for Python. *Comput Sci Eng* 2007;9:10.
16. Langfelder P, Zhang B, Horvath S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* 2008;24:719–20.
17. Hugot JP, Chamaillard M, Zouali H, et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;411:599–603.
18. Coelho T, Andreoletti G, Ashton JJ, et al. Immuno-genomic profiling of patients with inflammatory bowel disease: a systematic review of genetic and functional in vivo studies of implicated genes. *Inflamm Bowel Dis* 2014;20:1813–9.

19. Wong SH, Gao Q, Tsoi KK, et al. Effect of immunosuppressive therapy on interferon gamma release assay for latent tuberculosis screening in patients with autoimmune diseases: a systematic review and meta-analysis. *Thorax* 2016;71:64–72.
20. Edwards A, Gao Y, Allan RN, et al. Corticosteroids and infliximab impair the performance of interferon-gamma release assays used for diagnosis of latent tuberculosis. *Thorax* 2017;72:946–9.
21. van Heel DA, Ghosh S, Butler M, et al. Muramyl dipeptide and toll-like receptor sensitivity in NOD2-associated Crohn's disease. *Lancet* 2005;365:1794–6.
22. Netea MG, Ferwerda G, de Jong DJ, et al. The frameshift mutation in Nod2 results in unresponsiveness not only to Nod2- but also Nod1-activating peptidoglycan agonists. *J Biol Chem* 2005;280:35859–67.
23. Smythies LE, Sellers M, Clements RH, et al. Human intestinal macrophages display profound inflammatory anergy despite avid phagocytic and bacteriocidal activity. *J Clin Invest* 2005;115:66–75.
24. Babyatsky MW, Rossiter G, Podolsky DK. Expression of transforming growth factors alpha and beta in colonic mucosa in inflammatory bowel disease. *Gastroenterology* 1996;110:975–84.
25. de Waal Malefyt R, Abrams J, Bennett B, et al. Interleukin 10(IL-10) inhibits cytokine synthesis by human monocytes: an autoregulatory role of IL-10 produced by monocytes. *J Exp Med* 1991;174:1209–20.
26. Cario E, Podolsky DK. Differential alteration in intestinal epithelial cell expression of toll-like receptor 3 (TLR3) and TLR4 in inflammatory bowel disease. *Infect Immun* 2000;68:7010–7.
27. Lu Y, Li X, Liu S, et al. Toll-like receptors and inflammatory bowel disease. *Front Immunol* 2018;9:72.
28. Torres J, Colombel JF. Genetics and phenotypes in inflammatory bowel disease. *Lancet* 2016;387:98–100.
29. Olivera P, Danese S, Jay N, et al. Big data in IBD: a look into the future. *Nat Rev Gastroenterol Hepatol* 2019;16:312–21.
30. Rialon KL, Crowley E, Seemann NM, et al. Long-term outcomes for children with very early-onset colitis: Implications for surgical management. *J Pediatr Surg* 2018;53:964–7.
31. Ashton JJ, Mossotto E, Ennis S, et al. Personalising medicine in inflammatory bowel disease-current and future perspectives. *Transl Pediatr* 2019;8:56–69.
32. Cleynen I, Boucher G, Jostins L, et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet* 2016;387:156–67.

MPG**Journal of Pediatric Gastroenterology and Nutrition**

Manuscript No. JPGN-19-951

Dear Author,

During the preparation of your manuscript for typesetting, some queries have arisen. These are listed below. Please check your typeset proof carefully and mark any corrections in the margin as neatly as possible or compile them as a separate list. This form should then be returned with your marked proof/list of corrections to the Production Editor.

QUERIES: to be answered by AUTHOR

QUERY NO.	QUERY DETAILS	RESPONSE
<AQ1>	Please check the suggested running title for correctness.	
<AQ2>	Please confirm whether surnames/family names (red) have been identified correctly in the author byline.	
<AQ3>	Affiliations have been set as per style. Please check for accuracy of information.	
<AQ4>	Please check and confirm the corresponding author's name and relevant details for correctness.	
<AQ5>	Please expand MDP in the section heading as well as define it at the first instance of occurrence in the text, if possible.	
<AQ6>	References have been updated using PubMed. Please check.	