

Targeted Analysis of Serum Proteins Encoded at Known Inflammatory Bowel Disease Risk Loci

Kimi Drobin, MSc,* Ghazaleh Assadi, PhD,[†] Mun-Gwan Hong, PhD,* Eni Andersson, MSc,* Claudia Fredolini, PhD,* Björn Forsström, PhD,* Anna Reznichenko, PhD,[†] Tahmina Akhter, MSc,[†] Weronica E. Ek, PhD,^{†,‡} Ferdinando Bonfiglio, PhD,^{†,§} Mark Berner Hansen, MD, DMSci,^{¶,||} Kristian Sandberg, PhD,^{**††} Dario Greco, PhD,^{‡‡} Dirk Repsilber, PhD,^{§§} Jochen M. Schwenk, PhD,^{*a} Mauro D'Amato, PhD,^{†,¶¶,a} and Jonas Halfvarson, MD, PhD^{||,a}

Background: Few studies have investigated the blood proteome of inflammatory bowel disease (IBD). We characterized the serum abundance of proteins encoded at 163 known IBD risk loci and tested these proteins for their biomarker discovery potential.

Methods: Based on the Human Protein Atlas (HPA) antibody availability, 218 proteins from genes mapping at 163 IBD risk loci were selected. Targeted serum protein profiles from 49 Crohn's disease (CD) patients, 51 ulcerative colitis (UC) patients, and 50 sex- and age-matched healthy individuals were obtained using multiplexed antibody suspension bead array assays. Differences in relative serum abundance levels between disease groups and controls were examined. Replication was attempted for CD-UC comparisons (including disease subtypes) by including 64 additional patients (33 CD and 31 UC). Antibodies targeting a potentially novel risk protein were validated by paired antibodies, Western blot, immuno-capture mass spectrometry, and epitope mapping.

Results: By univariate analysis, 13 proteins mostly related to neutrophil, T-cell, and B-cell activation and function were differentially expressed in IBD patients vs healthy controls, 3 in CD patients vs healthy controls and 2 in UC patients vs healthy controls ($q < 0.01$). Multivariate analyses further differentiated disease groups from healthy controls and CD subtypes from UC ($P < 0.05$). Extended characterization of an antibody targeting a novel, discriminative serum marker, the laccase (multicopper oxidoreductase) domain containing 1 (LACC1) protein, provided evidence for antibody on-target specificity.

Conclusions: Using affinity proteomics, we identified a set of IBD-associated serum proteins encoded at IBD risk loci. These candidate proteins hold the potential to be exploited as diagnostic biomarkers of IBD.

Key Words: inflammatory bowel disease, affinity proteomics, LACC1

INTRODUCTION

Inflammatory bowel disease (IBD) is an immune-mediated disease causing chronic inflammation in the gastrointestinal (GI) tract. The disease entity is characterized by

relapsing course of diarrhea, abdominal pain, and weight loss. The prevalence of IBD is approximately 0.5% in the Western world, and the prevalence and incidence of the disease entity

support from Janssen, MSD, and Takeda. K.D., G.A., M.G.H., E.A., C.F., B.F., A.R., T.A., W.E.E., F.B., M.B.H., K.S., D.G., D.R., J.M.S., and M.D. do not have any competing interests to disclose.

^aEqual contribution

Supported by: This study was supported by unrestricted research grants from AstraZeneca (Translational Research Program "Post-genomic applications in IBD: exploitation of genetic information toward improved diagnosis and therapy") and Vetenskapsrådet (VR 2013-3862 to M.D.); and by funds from Örebro University Hospital Research Foundation and Vetenskapsrådet (2011-2764 to J.H.). J.M.S. acknowledges an SRA grant from the Swedish Government (CancerUU) and support from the Science for Life Laboratory Stockholm, the Knut and Alice Wallenberg Foundation, and the KTH Center for Applied Precision Medicine (KCAP), financed by the Erling-Persson Family Foundation. D.R. acknowledges support from the Swedish Knowledge Foundation.

Address correspondence to: Jonas Halfvarson, PhD, Department of Gastroenterology, Faculty of Medicine and Health, Örebro University, SE 70182, Örebro, Sweden (jonas.halfvarson@regionorebrolan.se).

© 2018 Crohn's & Colitis Foundation. Published by Oxford University Press on behalf of Crohn's & Colitis Foundation.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

doi: 10.1093/ibd/izy326

Published online 24 October 2018

Received for publications April 16, 2018; Editorial Decision September 21, 2018.

From the *Affinity Proteomics, SciLifeLab, School of Biotechnology, KTH, Royal Institute of Technology, Stockholm, Sweden; [†]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden; [‡]Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden; [§]Department of Gastrointestinal and Liver Diseases, Biodonostia Health Research Institute, San Sebastián, Spain; [¶]AstraZeneca R&D Mölndal, Innovative and Global Medicines, Mölndal, Sweden; ^{||}Digestive Disease Center, Bispebjerg Hospital, University of Copenhagen, Copenhagen, Denmark; ^{**}Science for Life Laboratory, Drug Discovery & Development Platform & Organic Pharmaceutical Chemistry, Department of Medicinal Chemistry, Uppsala Biomedical Center, Uppsala University, Uppsala, Sweden; ^{††}Department of Physiology and Pharmacology, Karolinska Institutet, Stockholm, Sweden; ^{‡‡}Institute of Biotechnology, University of Helsinki, Helsinki, Finland; ^{§§}School of Medical Sciences, Örebro University, Örebro, Sweden; ^{¶¶}BioDonostia Health Research Institute, San Sebastian and IKERBASQUE, Basque Foundation for Science, Bilbao, Spain; ^aDepartment of Gastroenterology, Faculty of Medicine and Health, Örebro University, Örebro, Sweden

Conflicts of interest: J.H. has received consultant/lecture fees from Abbvie, Celgene, Ferring, Hospira, Janssen, Medivir, MSD, Pfizer, Sandoz, Shire, Takeda, Tillotts Pharma, Janssen, MSD, Takeda, Tillotts, and Vifor Pharma and grant

is increasing globally.¹ Crohn's disease (CD) and ulcerative colitis (UC) represent the 2 main subtypes of the disease. The precise pathophysiology of IBD remains largely unknown, but accumulating evidence suggests that the dysregulated mucosal immune response is caused by a loss of tolerance toward gut microbiota in genetically susceptible individuals.²

Historically, the influence of genetics was first illustrated by high concordance rates in family and twin studies.³ During the last decades, large consortium-based meta-analyses of genome-wide association studies (GWAS) have dramatically increased our knowledge of the genetic architecture of IBD and its subtypes CD and UC. These analyses have also revealed that the genetic information might contribute to the classification of IBD patients into disease-specific subtypes, specifically in relation to the location of inflammation.⁴ Interestingly, genetic risk effects often appear to be mediated by allelic differences (risk vs nonrisk variants) in the modulation of mRNA expression (expression quantitative trait loci [eQTLs]).

Even though the identification of genetic risk variants has largely improved our understanding of disease mechanisms in IBD, many genes and their encoded proteins are still functionally uncharacterized. Thus, the pace of protein characterization has been slower than that at the genetic level, and the repertoire and overall nature of the pool of IBD risk gene products ("IBD risk proteome") still remain largely unknown.

One initiative to increase our understanding of the proteome is the Human Protein Atlas (HPA) project, which aims at generating and applying antibodies to study all of the approximately 20,000 human proteins encoded in the genome.⁵ The current version of the HPA (version 15) includes 25,039 antibodies targeting 17,005 proteins, most of which are annotated to relative expression data and cell- and tissue-specific localization (www.proteinatlas.org).

In the present exploratory study, we took advantage of the HPA repository for the characterization of expression of the IBD risk proteome in human serum. In particular, we screened protein products encoded at IBD risk loci for their potential to distinguish IBD patients from healthy individuals and to further differentiate between different disease subtypes. We thus propose a list of IBD-associated protein targets that may be exploited in follow-up studies for future IBD profiling efforts.

METHODS

Study Population

The IBD patients included in this study were obtained from a cohort previously described.⁶ In short, adult patients with CD and UC were consecutively recruited at the outpatient IBD clinic of Örebro University Hospital, Sweden. After obtaining an informed written consent, blood samples were collected, and the serum was separated after centrifugation at 2400g for 6 minutes at room temperature. All serum samples were stored as aliquots at -80°C . Diagnosis was based on internationally

accepted clinical, endoscopic, radiologic, and histologic criteria.⁷ Medical notes were scrutinized to classify disease characteristics according to the Montreal classification.⁸ A random sample set of 49 CD patients, 51 UC patients, and 50 healthy blood donors (no history of chronic GI disease), matched according to sex and age ± 5 years (sample set, IBD 1), was selected. In addition, 33 CD and 31 UC patients were selected to extend the analyses and explore possible differences between subgroups of CD and UC patients (sample set, IBD 2). Demographics and clinical characteristics of patients with IBD are reported in [Table 1](#). None of the patients were included at disease onset, and only a few patients had early IBD, as illustrated by the information on disease duration in [Table 1](#). The study was approved by the Örebro Regional Ethics Committee (2006/245).

Experimental Strategy

To characterize IBD risk proteome serum expression profiles in patients and controls, we applied an affinity proteomic analysis targeting proteins encoded at known IBD risk loci.⁹ In addition, a small subset of proteins known to be involved in inflammation, including neutrophil regulation, was added as "experimental controls."¹⁰ Quality assessment was followed by data analyses based on univariate and multivariate approaches, and a brief outline is reported in [Figure 1](#).

Antibody Bead Array Assay

Antibody selection and bead coupling

The HPA library (version 15) was screened to identify antibodies targeting any of the 1438 predicted protein products encoded at the 163 IBD risk loci (known at the time the study was initiated).⁹ From the 601 thereby identified, final selection of antibodies suitable for suspension bead arrays (SBAs) was based on availability, binding specificity assessed by protein arrays,¹¹ and concentration (>0.05 mg/mL). This yielded a total of 343 antibodies directed against 205 unique target proteins, corresponding to 104 of the 163 IBD risk loci, which are listed in [Supplementary Table 1](#), together with a small "control" selection of 22 antibodies directed against 13 known neutrophil- and inflammation-associated proteins.

Antibodies were then coupled to magnetic color-coded microspheres (MagPlex, Luminex Corp.) and assessed for coupling efficiency, and SBAs were generated as previously described.¹² Rabbit antihuman albumin (Dako) and donkey antihuman IgG (Jackson ImmunoResearch Laboratories) antibodies were used as controls for sample transfer, whereas rabbit IgG (Jackson ImmunoResearch Laboratories) and bare beads served as negative controls.

Sample randomization and bead array processing

Before the analysis, serum samples were randomized and distributed into their assigned microtiter plate positions

TABLE 1. Demographics and Clinical Characteristics of CD and UC Patients

	IBD 1		IBD 2	
	Crohn's Disease n = 49	Ulcerative Colitis n = 51	Crohn's Disease n = 31	Ulcerative Colitis n = 33
Male sex, No. (%)	33 (67)	35 (69)	21 (68)	21 (64)
Median (range) age at diagnosis, y	28 (10–54)	30 (5–61)	25 (7–46)	26 (12–65)
Median (range) disease duration, y	20 (0–43)	15 (1–39)	17 (0–36)	10 (0–48)
Disease location, No. (%)				
Ileal (L1)	14 (29)		5 (16)	
Colonic (L2)	14 (29)		15 (48)	
Ileocolonic (L3)	21 (43)		11 (35)	
Upper disease (L4)				
Disease behavior, No. (%)				
Nonstricturing, nonpenetrating (B1)	18 (37)		18 (58)	
Stricturing (B2)	21 (43)		9 (29)	
Penetrating (B3)	10 (20)		4 (13)	
Perianal fistulas	6 (12)		7 (23)	
Disease extent, No. (%)				
Proctitis (E1)		6 (12)		7 (21)
Left-sided colitis (E2)		23 (45)		10 (30)
Extensive colitis (E3)		22 (43)		16 (48)
Clinical disease activity, No. (%) ^a				
Remission	34 (69)	39 (76)	23 (74)	26 (79)
Active	14 (29)	12 (24)	8 (26)	7 (21)
Medications, No. (%) ^b				
5ASA/SASP (local or oral)	8 (16)	25 (49)	3 (10)	16 (48)
Corticosteroids (local or oral)	7 (14)	8 (16)	6 (19)	7 (21)
Thiopurines	12 (24)	13 (25)	9 (29)	4 (12)
Methotrexate	2 (4)	2 (4)	3 (10)	1 (3)
Anti-TNF	1 (2)	0 (0)	3 (10)	0 (0)
No drugs	24 (49)	12 (24)	13 (42)	12 (36)
Previous surgical resection, No. (%)	34 (69)	6 (12)	15 (48)	3 (9)

Abbreviations: 5ASA/SASP, 5-aminosalicylates/sulfasalazine; anti-TNF, anti-tumor necrosis factor.

^aData on disease activity were not available in 1 patient with Crohn's disease in IBD 1.

^bSome patients were in a combination of different treatments: IBD 1 CD, n = 3; IBD 1 UC, n = 8; IBD 2 CD, n = 6; IBD 2 UC, n = 6.

using a liquid handling device (Freedom EVO150, Tecan). Serum aliquots were then diluted 1:10 in PBS and labeled using NHS-PEO4-biotin (Pierce). The labeled samples were diluted 1:50 in assay buffer and heat-treated at 56°C for 30 minutes before being combined with the suspension bead array. Details about this protocol have been described previously.¹³ Transfer of liquid volumes <5 µL into and between plates was performed with a liquid handling device (CyBi-SELMA, CyBio). To avoid location effects because of the sequential read-out of plates,¹⁴ diluted samples were randomized across two 384-well plates to obtain a similar distribution of sample sets, age, and sex. In addition, each 384-well plate contained 24 samples

(16 repeated pooled serum samples and 8 samples containing only buffer) to assess technical variance. The 2 assay plates were processed in parallel. The beads were washed in 1x PBS with 1% Tween20 using a plate washer (EL406, Biotek). Lastly, median fluorescence intensity (MFI) levels were obtained for each antibody-coupled bead using Flexmap 3D instruments (Luminex Corp.), accepting a minimum of 35 events for each bead ID.

Antibody validation

A description of methods used for antibody validation can be found in the [Supplementary Data](#).

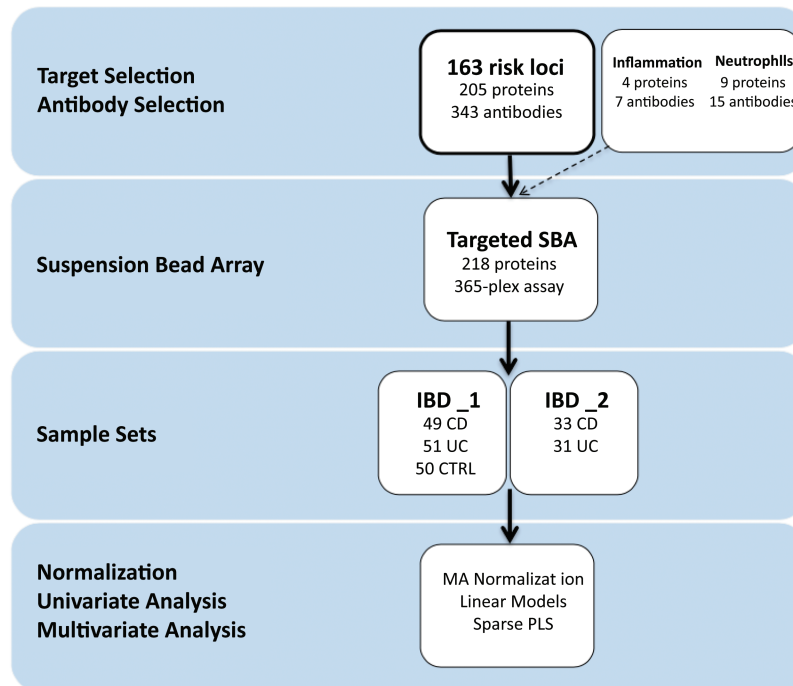


FIGURE 1. Schematic description of study. Target selection was made based on the risk loci GWAS hits ($n = 163$). Thirteen additional proteins involved in inflammation, including neutrophil regulation, were added as “experimental controls.” From these, antibodies available from the Human Protein Atlas were selected to compose a targeted SBA. Two sample sets, denoted IBD 1 and IBD 2, were then screened using this bead array. Upon data processing, both uni- and multivariate analyses were performed to propose targets of interest. Comparisons of IBD patients and subgroups of patients vs healthy controls were performed based on sample set IBD 1, and for comparisons between subgroups of patients, both sample sets (ie, IBD 1 and IBD 2) were used.

Data Processing and Analysis

Data preprocessing

For data evaluation, processing, and analysis, the statistical environment R¹⁵ was used unless otherwise stated. For quality assessment via calculations of coefficient of variations (CV), unprocessed data were employed. Antibodies revealing average MFI levels lower than rabbit IgG (negative control) or higher than anti-albumin (positive control) were removed. Further investigations involved data that were processed with multidimensional MA normalization¹⁶ after exclusion of data generated from sample-free assay buffer, reference sample pools, and the removal of outliers based on robust principal component analysis.¹⁷

Univariate comparisons

For each antibody profile, a linear regression model with adjustment for age and sex was used to find profiles that could separate IBD patients from healthy controls, CD patients from healthy controls, or UC patients from healthy controls in the IBD 1 sample set. Comparisons of CD patients vs UC patients were performed on both sample sets (IBD 1 and IBD 2), and results were merged by a fixed-effects model for meta-analysis using the “meta” package in R.¹⁸ The q -value approach to false

discovery rate (FDR) estimation¹⁹ was used to account for multiple testing. To shortlist candidate proteins for comparison of patients vs healthy controls, we used an FDR cutoff at $q < 0.01$, whereas a less stringent cutoff of $q < 0.05$ was used for comparing UC vs CD.

Multivariate analysis

To identify differentially expressed signatures of proteins, a multivariate approach was applied. Multivariate classification analysis was implemented as sparse partial least squares (sPLS) analysis²⁰ for the same comparisons as in the univariate analysis. Data from both sample sets (IBD 1 and IBD 2) were merged to obtain a data set large enough for a rigorous double cross-validation approach (described below). Data were quantile-normalized using the R package “limma” before analysis.²¹ After model fitting, variable importance was calculated as a variable importance in projection (VIP) value²² for each protein in each of the group comparisons: (1) IBD patients vs healthy controls, (2) CD patients vs healthy controls, (3) UC patients vs healthy controls, (4) CD patients vs UC patients, (5) colonic (L2) CD patients vs UC patients, and (6) ileal (L1/L3) CD patients vs UC patients. VIP was calculated for all variables, and the analyses were optimized for both the number of variables and the number of components to use in the respective

prediction model, within the inner cross-validation loop.^{23, 24} The prediction models were validated, applying the outer leave-one-out cross-validation (LOO) approach, and hit rates were collected. Significance of the observed LOO hit rates was established by resampling analyses, that is, randomly permuting the class labels and re-running the double cross-validation analyses (including inner loops and optimizations) to be able to calculate permutation *P* values for the observed LOO prediction hit rates for the original data.

RESULTS

Data Quality Assessment

At first, we assessed the overall quality of the data and determined the coefficient of variation (CV) of each antibody in replicated and independent samples. As shown in [Supplementary Figure 1](#), the CVs of technical reproducibility (tCVs), calculated from the replicated reference sample pools, were <10% in 279 of 365 antibodies (76%). A denoted biological CV (bCV), describing the variation across all other samples, was also calculated. The median tCVs (9%) were substantially lower when compared with the bCVs (>37%), indicating that the variability in the data set is due to biological differences and not technical artifacts. For the subsequent analyses, antibodies with tCV <15% were included (n = 355). These antibodies were directed against 204 proteins, encoded at 104 genetic risk loci. We did not identify any sample outliers when applying robust PCA analyses (not shown).

Identification of Differentially Abundant Proteins

Univariate analysis

To identify single proteins associated with IBD and subtypes of the disease, univariate analyses were performed. Using the data set IBD 1, the comparison IBD (CD and UC) vs healthy controls yielded significant results for 13 antibodies ([Table 2](#)), and the relative abundance of the 4 top-ranking antibodies is illustrated in [Figure 2A](#). Similarly, an independent comparison of CD patients vs healthy controls resulted in significant differences for 3 antibodies, and the corresponding comparison of UC patients vs healthy controls resulted in significant differences for 2 antibodies ([Table 2](#), [Fig. 2B](#) and [C](#)). When CD patients and UC patients were compared, using the combined data set (IBD 1 and IBD 2), significant results were obtained for 2 proteins, namely serum amyloid protein A (SAA) and cAMP responsive element binding protein 5 (CREB5) ([Table 3](#)). The relative abundance of these 2 proteins is shown in [Figure 3](#). Almost all antibodies that were identified when IBD and subtypes of the disease were compared represented protein products encoded at the 163 IBD risk loci, and only 1 of them (serum amyloid protein A [SAA]) corresponded to the small “control” selection of known neutrophil- and inflammation-associated proteins. No significant results were observed for the comparisons of colonic CD (L2) vs UC, ileal CD (L1/L3) vs UC, and nonstricturing, nonpenetrating CD (B1) vs UC. However, the relative abundance of CREB5

TABLE 2. Antibodies and Corresponding Proteins With Differential Abundance in Patients With IBD and Subtypes of the Disease Compared With Controls, Identified by Univariate Analyses

Comparison	Gene	Antibody	<i>P</i>	q
IBD-CTRL	LACC1	HPA040150	1.3E-05	3.0E-03
	IL2RA	HPA054622	4.7E-05	4.0E-03
	LACC1	HPA061537	6.1E-05	4.0E-03
	LNPEP	HPA043642	6.8E-05	4.0E-03
	CNTF	HPA046534	9.1E-05	4.1E-03
	LPXN	HPA061441	1.0E-04	4.1E-03
	BTNL2	HPA039844	1.9E-04	6.1E-03
	IFNAR2	HPA029229	2.2E-04	6.1E-03
	CARD11	HPA052984	2.3E-04	6.1E-03
	JAK2	HPA058253	3.8E-04	8.6E-03
	PEX13	HPA061468	4.0E-04	8.6E-03
	SLC22A5	HPA063062	5.4E-04	9.8E-03
	IFNG	HPA063125	5.4E-04	9.8E-03
	CD-CTRL	LACC1	HPA040150	1.8E-05
SAA		HPA059733	2.0E-05	2.1E-03
LNPEP		HPA043642	1.2E-04	8.4E-03
UC-CTRL	CNTF	HPA046534	8.3E-05	7.8E-03
	LPXN	HPA061441	8.9E-05	7.8E-03

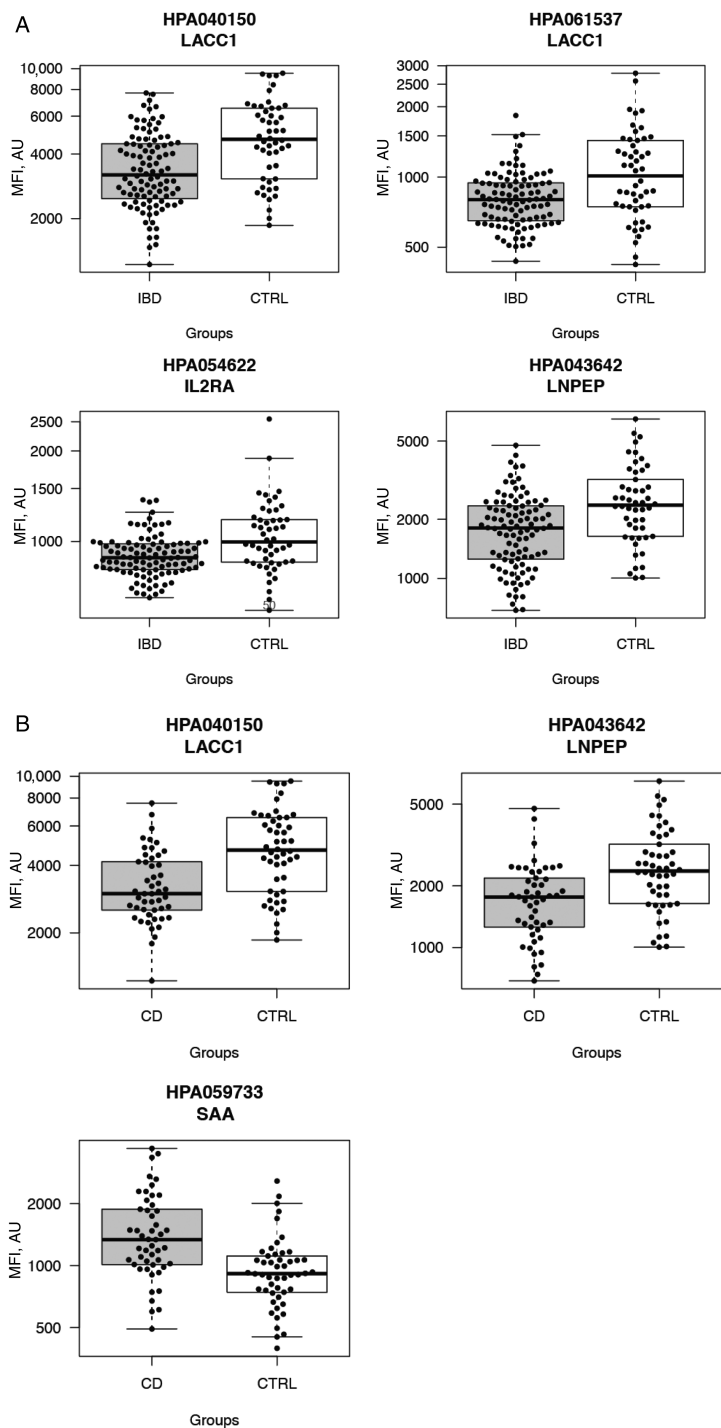


FIGURE 2. Candidate proteins: IBD patients vs healthy controls. The boxplots show the top 4 proteins revealing significant differences in univariate group comparisons ($q < 0.01$). A, Candidate proteins to differ between IBD patients (gray; $n = 100$) and healthy controls (CTRL; white; $n = 50$) are targeted by 2 LACC1 antibodies and shown on the upper panel: HPA040150 ($q = 0.003$) and HPA061537 ($q = 0.004$). On the lower panel are IL2RA (HPA054622; $q = 0.003$) and LNPEP (HPA043642; $q = 0.003$); hence, all revealed higher levels in controls compared with the combined UC and CD cases. B, When comparing CD ($n = 49$) and controls ($n = 50$), antibodies against LACC1 (HPA040150; $q = 0.002$) and LNPEP (HPA043642; $q = 0.008$) revealed higher protein levels in controls, whereas levels of SAA (HPA059733; $q = 0.002$) were higher in CD patients. C, When comparing UC ($n = 51$) with controls ($n = 50$), the levels for CNTF (HPA046534; $q = 0.008$) and LPXN (HPA061441; $q = 0.008$) were all lower in UC cases. The presented data are for normalized MFI values.

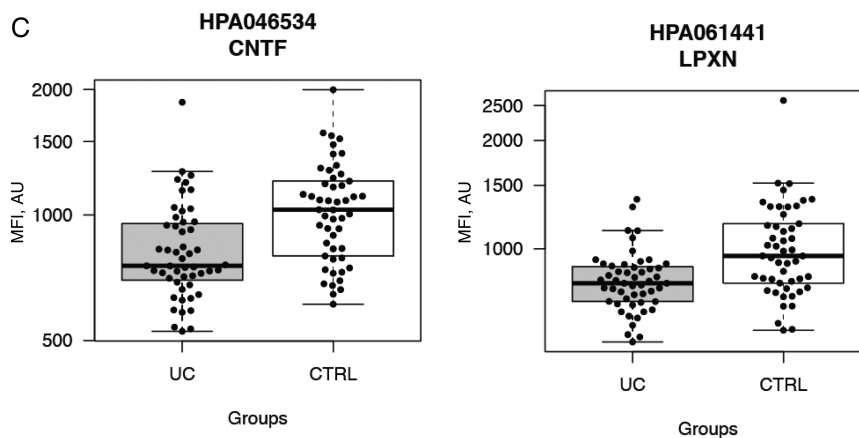


FIGURE 2. (Continued)

TABLE 3. Antibodies and Corresponding Proteins With Differential Abundance in Patients With Crohn's Disease Compared With Patients With Ulcerative Colitis in Sample Sets IBD 1 and IBD 2, Identified by Univariate Analyses

Gene	Antibody	IBD 1 <i>P</i>	IBD 2 <i>P</i>	Meta <i>P</i>	Meta <i>q</i>
SAA	HPA059733	8.75E-02	1.21E-01	3.45E-05	1.24E-02
CREB5	HPA057734	2.83E-02	3.63E-02	2.04E-04	3.68E-02

differed between CD patients with complicated disease behavior, that is, stricturing (B2) or penetrating (B3) disease, and patients with ulcerative colitis. The performance characteristics (tCV, bCV, and ρ) of the significantly altered antibodies can be found in [Supplementary Table 2](#).

Multivariate analysis

We used a multivariate sPLS model to identify signatures of proteins differentially expressed between different groups of patients and healthy controls. Our sPLS modeling resulted in fitted sPLS models, together with optimal numbers of components (latent variables for the PLS discriminant analysis) and optimal numbers of proteins used for each model. Most proteins identified in univariate analysis as being significantly altered in IBD, CD, and UC vs controls, and in CD vs UC, were also identified in the discriminant sPLS analyses ([Supplementary Table 3](#)). In contrast to the univariate analyses, differentiating protein expression signatures between CD subtypes (colonic = L2, ileal = L1/L3, or complicated disease behavior = B2/B3) and UC were identified in the multivariate analyses ([Supplementary Table 3](#)). Differently expressed protein signatures were also observed when UC patients with active disease were compared with those in remission, but no differences were observed when CD patients were stratified by disease activity ([Supplementary Table 3](#)). [Figure 4](#) shows sPLS

scores plots to visualize how the found biosignatures could classify the different disease classes in our data.

To assess the predictive accuracy of our model, hit rates were collected, and their significance assured that a leave-one-out cross-validation approach could be applied. The hit rates and their respective resampling significances, obtained in our sPLS analyses, are shown in [Supplementary Table 4](#). All comparisons resulted in significant resampling *P* values, but the observed hit rates for CD vs UC and colonic CD vs UC were comparatively low.

LACC1-Specific Analyses

A primary finding from the univariate and multivariate analyses is that the protein laccase (multicopper oxidoreductase) domain containing 1 (LACC1; previously known as C13orf31) was differentially abundant in the serum samples from IBD patients compared with healthy controls. LACC1 has been recently characterized as a novel player in IBD and inflammation, involved in the control of macrophage immunometabolic function in general, and we sought to further investigate our results. In the assays, LACC1 was detected by 2 different antibodies (HPA040150 and HPA061537), which are raised toward 2 independent epitopes, located at the N-terminal and C-terminal portions of the target protein, respectively. The obtained serum profiles

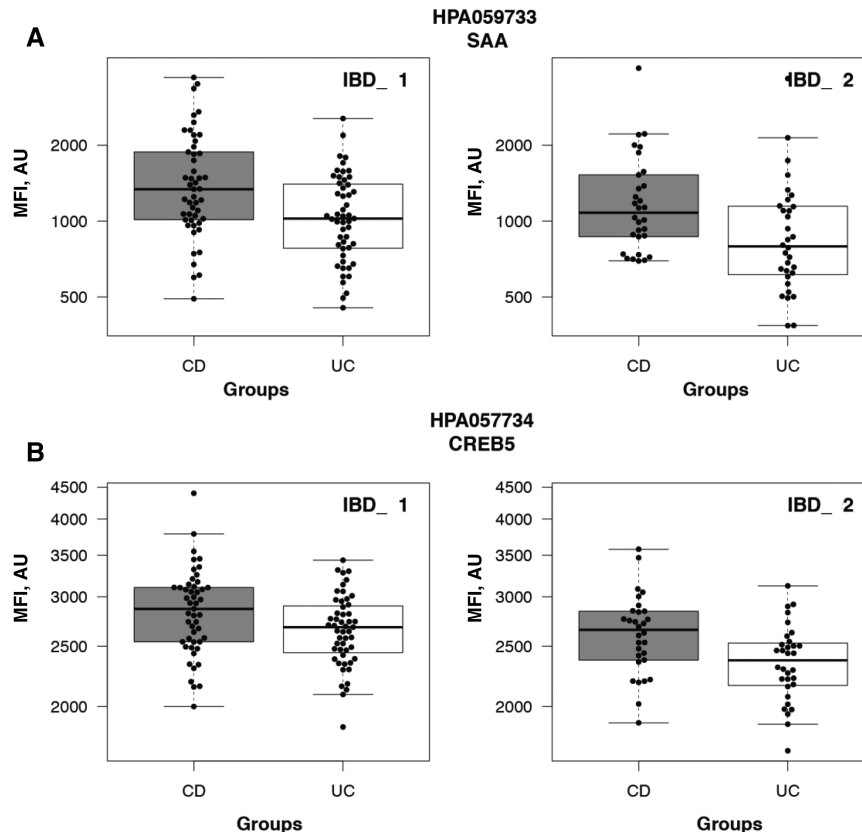


FIGURE 3. Candidate proteins: CD vs UC. The boxplots show the proteins that differed between CD and UC patients in the 2 study sets, IBD 1 and IBD 2, with $q < 0.05$. Levels of (A) SAA (HPA059733; $q = 0.012$) and (B) CREB5 (HPA057734; $q = 0.037$) were both higher in CD patients. The presented data are for normalized MFI values.

from these 2 antibodies correlated with a Spearman ρ of 0.6 (Supplementary Fig. 2), indicating concordant selectivity for a common target in the matrix of human serum. The antibody HPA040150 was further validated as it shows a stronger association when comparing IBD and CD patients with controls. Western blot analyses of protein extracts from LACC1-overexpressing cells established that HPA040150 binds to denatured LACC1 (Supplementary Fig. 3A). In addition, epitope mapping analyses of the antibody were performed. Three epitopes were identified (Supplementary Fig. 3B), and a Uniprot BLAST search (100% identity and E value threshold < 10) showed that only LACC1 matches the peptides. Finally, immuno-capture mass spectrometry analysis was conducted and confirmed that HPA040150 enriched its intended target LACC1 from serum (Supplementary Fig. 3C).

DISCUSSION

The identification of susceptibility risk variants has afforded limited success in translating genetic discoveries to enhanced knowledge of the serum proteome in IBD. Exploiting gene content from known IBD risk loci, we implemented a screen of the IBD risk proteome for the identification

of serum targets eventually suitable for inclusion in follow-up biomarker discovery efforts. For this purpose, we took advantage of the HPA large antibody repository and an affinity proteomic approach, which allowed the identification of proteins showing differential abundance in healthy vs IBD serum, and in serum from different subgroups of patients. Although these proteins mostly relate to neutrophil, T-cell, and B-cell activation and function, our strategy shows that multiplexed bead-based antibody array approaches may be successfully adopted to exploit existing genetic information for the delineation of targeted proteomic studies. However, inherent limitations in this study should be noted and eventually avoided in follow-up investigations: (1) the inclusion of patients with a long disease duration may have introduced bias due to disease history; (2) there is a lack of data on endoscopic activity that may have been correlated with specific protein profiles; (3) acknowledging the exploratory nature of our study, assays for known and expected protein signatures may not be reported; and (4) instead of using fully validated assays, the multiplexed approach may offer only insufficient analytical sensitivity and specificity, hence reducing the number of risk candidates identified in serum. Another important limitation is that we

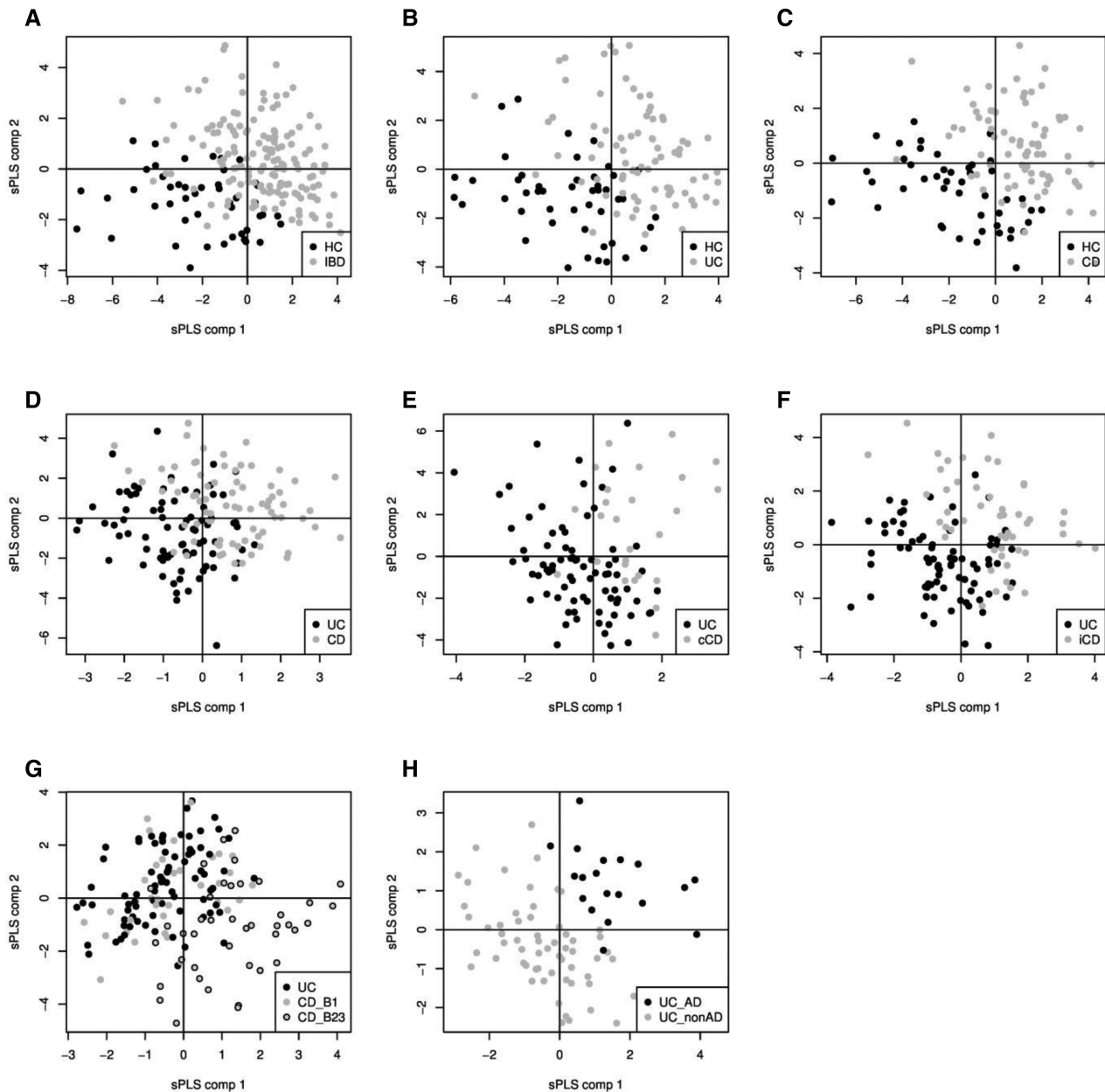


FIGURE 4. sPLS scores plots of comparisons between CD patients and subgroups of CD patients, ulcerative colitis patients, and healthy controls. The sPLS scores plots visualize how the found biosignatures could classify CD patients and subgroups of CD patients, ulcerative colitis patients, and healthy controls, respectively. Abbreviations: cCD, colonic CD; iCD, ileal CD; CD_B1, nonstricturing, nonpenetrating CD; CD_B23, stricturing or penetrating CD; UC_AD, UC with active disease; UC_nonAD, UC with nonactive disease.

were only able to measure 204 proteins out of 1438 proteins encoded in 104 out of the 163 risk loci. Although our analyses were based on an unbiased approach, in targeting all proteins encoded at IBD risk loci, we were not able to define the exact mechanisms underlying specific predictive protein profiles or to establish a relationship between individual single nucleotide polymorphisms (SNPs) and protein expression levels (protein quantitative trait loci [pQTLs]), as this requires large-scale

samples and analytical techniques that ensure exact relative measurement of protein levels.

Previously, only a small number of studies have aimed to identify serum proteins associated with IBD,²⁵⁻²⁷ but these have been in general limited to a small number of selected targets. In recent years, some initiatives to increase our understanding have been initiated. Di Narzo et al. reported novel aging-associated proteomic traits and proteomic traits in Crohn's disease patients,

corresponding to 41 distinct genes, that were significantly influenced by SNP genotypes in cis.²⁸ The HPA, aiming at generating and applying antibodies directed toward every human protein, represents another recent initiative. The project has been built on the systematic generation of antibodies through a predefined pipeline.^{29–32} Here we took advantage of this large antibody repository to assay serum expression levels of proteins translated from IBD-related risk loci⁹ and proteins related to inflammation including neutrophil activation. Based on this approach, we identified a number of proteins that appear to be differentially expressed in sera from IBD patients compared with healthy controls. Among the top-ranking proteins, we found targets related to cytokine signaling (IL2RA³³), immune-metabolic regulation and reactive oxygen species production (LACC1³⁴), antigen presentation (LNPEP³⁵), T-cell regulation (CARD11,³⁶ BTNL2,³⁷ and IFNG³⁸), B-cell signaling (LPXN), and neurotransmitter synthesis (CNTF³⁹). To our knowledge, only a few of these proteins, such as CNTF³⁹ and LNPEP,³⁵ have been known to be present in sera. LACC1, whose serum expression has not been previously studied, was identified to be differentially expressed in IBD patients compared with healthy controls. Genetic variation in the *LACC1* gene has been associated with IBD, leprosy, and systemic and nonsystemic juvenile idiopathic arthritis (JIA),^{9,40–42} whereas the corresponding protein product has been recently characterized as a master regulator of macrophage immune-metabolic function.⁴⁰ In the present study, we observed reduced serum LACC1 levels in IBD patients (primarily CD patients) compared with controls. Notably, reduced LACC1 expression has been reported to parallel the finding of hypomorphic (reduced function) *LACC1* gene variants, which are associated with increased risk of CD.⁴⁰

Lower abundance of LNPEP, an aminopeptidase involved in antigen presentation,⁴³ was also observed in CD patients compared with healthy controls. Interestingly, LNPEP comes from the IBD risk locus containing also the endoplasmic reticulum aminopeptidases 1 and 2 (ERAP1 and ERAP2), which are both involved in the regulation of innate immune response⁴⁴ and have well-established roles in IBD and other auto-inflammatory diseases.^{9,45} Lower levels of serum expression were observed for CNTF and LPXN in UC cases compared with controls. CNTF is believed to be involved in the recruitment of macrophages,⁴⁶ whereas LPXN has been reported to have an inhibitory role in B-cell function⁴⁷ and to be involved in the development of colorectal cancer in UC patients.⁴⁸

Interestingly, the results from specific antibodies for an epitope shared by 2 isoforms of the SAA, namely A1 and A2, point to increased abundance of the SAA protein in serum from CD vs UC patients. This finding is supported by a previously reported correlation between serum SAA levels and endoscopic activity in patients with CD.⁴⁹ Similarly, increased abundance of antibodies toward CREB5, a transcription factor regulating a variety of cellular responses,⁵⁰ was observed in CD patients when compared with UC patients. This observation

is consistent with the previously observed genetic association between *CREB5* and CD, but not UC,⁹ and the earlier reported upregulation of CREB expression in peripheral blood mononuclear cells of CD patients.^{51,52}

By implementing an sPLS model, we were able to identify signatures (ie, combinations) of proteins associated with IBD and its subtypes CD and UC. Based on these signatures, a predictive accuracy of 0.76–0.83 was observed when compared with healthy controls. The accuracy of our model dropped when different subtypes of IBD were compared against each other, even though our protein signatures remained significant, indicating that the found protein signatures were truly discriminative for all tested comparisons. Interestingly, a better accuracy was observed for serum profiles from ileal rather than colonic CD, when these were compared with UC profiles. This observation may add further support to recent genetic data, indicating that the genetic distance between ileal CD and UC is actually larger than that between colonic CD and UC.⁴ Differentially expressed signatures of proteins were also observed between CD patients with complicated disease behavior (stricturing [B2] or penetrating disease [B3]) and CD patients with nonstricturing, nonpenetrating disease (B1), and with UC patients. This finding may seem contrary to that of Cleynen et al., who reported little or no genetic association with disease behavior.⁴ However, our findings might have been biased by differences in other phenotypic characteristics such as disease location and age at onset.

In conclusion, we took advantage of a large repository of antibodies to exploit genetic information from IBD risk loci to identify disease-associated serum profiles. Among others, we report interesting results for LACC1, a novel immune player in IBD that shows downregulated expression in CD patients. A number of additional candidate proteins have the potential to be prioritized as selected targets in future biomarker discovery efforts in IBD.

SUPPLEMENTARY DATA

Supplementary data are available at *Inflammatory Bowel Diseases* online.

ACKNOWLEDGMENTS

The authors wish to acknowledge the entire staff of the Plasma Profiling group and the Human Protein Atlas team for their tremendous efforts.

REFERENCES

- Molodecky NA, Soon IS, Rabi DM, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*. 2012;142:46–54.e42; quiz e30.
- Kaser A, Zeissig S, Blumberg RS. Inflammatory bowel disease. *Annu Rev Immunol*. 2010;28:573–621.
- Ek WE, D'Amato M, Halfvarson J. The history of genetics in inflammatory bowel disease. *Ann Gastroenterol*. 2014;27:294–303.
- Cleynen I, Boucher G, Jostins L, et al; International Inflammatory Bowel Disease Genetics Consortium. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. *Lancet*. 2016;387:156–167.

5. Uhlén M, Fagerberg L, Hallström BM, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015;347:1260419.
6. Fransén K, Franzén P, Magnusson A, et al. Polymorphism in the retinoic acid metabolizing enzyme CYP26B1 and the development of Crohn's disease. *PLoS One*. 2013;8:e72739.
7. Lennard-Jones JE. Classification of inflammatory bowel disease. *Scand J Gastroenterol Suppl*. 1989;170:2–6; discussion 16.
8. Silverberg MS, Satsangi J, Ahmad T, et al. Toward an integrated clinical, molecular and serological classification of inflammatory bowel disease: report of a working party of the 2005 Montreal World Congress of Gastroenterology. *Can J Gastroenterol*. 2005;19(Suppl A):5A–36A.
9. Jostins L, Ripke S, Weersma RK, et al; International IBD Genetics Consortium (IIBDGC). Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–124.
10. Muthas D, Reznichenko A, Balendran CA, et al. Neutrophils in ulcerative colitis: a review of selected biomarkers and their potential therapeutic implications. *Scand J Gastroenterol*. 2017;52:125–135.
11. Sjöberg R, Mattsson C, Andersson E, et al. Exploration of high-density protein microarrays for antibody validation and autoimmunity profiling. *N Biotechnol*. 2016;33:582–592.
12. Drobin K, Nilsson P, Schwenk JM. Highly multiplexed antibody suspension bead arrays for plasma protein profiling. *Methods Mol Biol*. 2013;1023:137–145.
13. Byström S, Ayoglu B, Häggmark A, et al. Affinity proteomic profiling of plasma, cerebrospinal fluid, and brain tissue within multiple sclerosis. *J Proteome Res*. 2014;13:4607–4619.
14. Kato BS, Nicholson G, Neiman M, et al. Variance decomposition of protein profiles from antibody arrays using a longitudinal twin model. *Proteome Sci*. 2011;9:73.
15. R Development Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: The R Foundation for Statistical Computing; 2011.
16. Hong MG, Lee W, Nilsson P, et al. Multidimensional normalization to minimize plate effects of suspension bead array data. *J Proteome Res*. 2016;15(10):3473–3480.
17. Hubert M, Rousseeuw PJ, Vanden Branden K. ROBPCA: a new approach to robust principal component analysis. *Technometrics*. 2005;47:64–79.
18. Schwarzer G. Meta: An R package for meta-analysis. *R News*. 2007;7:40–45.
19. Storey JD. A direct approach to false discovery rates. *J R Stat Soc Series B Stat Methodol*. 2002;64:479–498.
20. Lê Cao KA, Boitard S, Besse P. Sparse PLS discriminant analysis: biologically relevant feature selection and graphical displays for multiclass problems. *BMC Bioinformatics*. 2011;12:253.
21. Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol*. 2004;3:Article3.
22. Mehmood T, Liland KH, Snipen L, et al. A review of variable selection methods in partial least squares regression. *Chemometr Intell Lab Syst*. 2012;118:62–69.
23. Eriksson L, Andersson PL, Johansson E, Tysklind M. Megavariable analysis of environmental QSAR data. Part I—a basic framework founded on principal component analysis (PCA), partial least squares (PLS), and statistical molecular design (SMD). *Mol Divers*. 2006;10:169–186.
24. Lee LC, Liang CY, Jemain AA. Partial least squares-discriminant analysis (PLS-DA) for classification of high-dimensional (HD) data: a review of contemporary practice strategies and knowledge gaps. *Analyst*. 2018;143:3526–3539.
25. Meuwis MA, Fillet M, Geurts P, et al. Biomarker discovery for inflammatory bowel disease, using proteomic serum profiling. *Biochem Pharmacol*. 2007;73:1422–1433.
26. Li N, Wang X, Zhang Y, et al. Comparative proteomics analysis of serum proteins in ulcerative colitis patients. *Mol Biol Rep*. 2012;39:5659–5667.
27. Wasinger VC, Yau Y, Duo X, et al. Low mass blood peptides discriminative of inflammatory bowel disease (IBD) severity: a quantitative proteomic perspective. *Mol Cell Proteomics*. 2016;15:256–265.
28. Di Narzo AF, Telesco SE, Brodmerkel C, et al. High-throughput characterization of blood serum proteomics of IBD patients with respect to aging and genetic factors. *PLoS Genet*. 2017;13:e1006565.
29. Agaton C, Galli J, Höiden Guthenberg I, et al. Affinity proteomics for systematic protein profiling of chromosome 21 gene products in human tissues. *Mol Cell Proteomics*. 2003;2:405–414.
30. Nilsson P, Paavilainen L, Larsson K, et al. Towards a human proteome atlas: high-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics*. 2005;5:4327–4337.
31. Uhlen M, Ponten F. Antibody-based proteomics for human tissue profiling. *Mol Cell Proteomics*. 2005;4:384–393.
32. Uhlen M, Oksvold P, Fagerberg L, et al. Towards a knowledge-based human protein atlas. *Nat Biotechnol*. 2010;28:1248–1250.
33. Rubin LA, Kurman CC, Fritz ME, et al. Soluble interleukin 2 receptors are released from activated human lymphoid cells in vitro. *J Immunol*. 1985;135:3172–3177.
34. Cader MZ, Boroviak K, Zhang Q, et al. C13orf31 (FAMIN) is a central regulator of immunometabolic function. *Nat Immunol*. 2016;17:1046–1056.
35. Mizutani S, Akiyama H, Kurauchi O, et al. In vitro degradation of angiotensin II (A-II) by human placental subcellular fractions, pregnancy sera and purified placental aminopeptidases. *Acta Endocrinol (Copenh)*. 1985;110:135–139.
36. Wang D, You Y, Case SM, et al. A requirement for CARMA1 in TCR-induced NF-kappa B activation. *Nat Immunol*. 2002;3:830–835.
37. Nguyen T, Liu XK, Zhang Y, Dong C. BTNL2, a butyrophilin-like molecule that functions to inhibit T cell activation. *J Immunol*. 2006;176:7354–7360.
38. Khor B, Gardet A, Xavier RJ. Genetics and pathogenesis of inflammatory bowel disease. *Nature*. 2011;474:307–317.
39. Kazim SF, Cardenas-Aguayo Mdel C, Arif M, et al. Sera from children with autism induce autistic features which can be rescued with a CNTF small peptide mimetic in rats. *PLoS One*. 2015;10:e0118627.
40. Assadi G, Saleh R, Hadizadeh F, et al. LACC1 polymorphisms in inflammatory bowel disease and juvenile idiopathic arthritis. *Genes Immun*. 2016;17:261–264.
41. Zhang FR, Huang W, Chen SM, et al. Genomewide association study of leprosy. *N Engl J Med*. 2009;361:2609–2618.
42. Wakil SM, Monies DM, Abouelhoda M, et al. Association of a mutation in LACC1 with a monogenic form of systemic juvenile idiopathic arthritis. *Arthritis Rheumatol*. 2015;67:288–295.
43. Saveanu L, van Ender P. The role of insulin-regulated aminopeptidase in MHC class I antigen presentation. *Front Immunol*. 2012;3:57.
44. Cifaldi L, Romania P, Lorenzi S, et al. Role of endoplasmic reticulum aminopeptidases in health and disease: from infection to cancer. *Int J Mol Sci*. 2012;13:8338–8352.
45. Agrawal N, Brown MA. Genetic associations and functional characterization of M1 aminopeptidases and immune-mediated diseases. *Genes Immun*. 2014;15:521–527.
46. Kobayashi H, Mizisin AP. CNTFR alpha alone or in combination with CNTF promotes macrophage chemotaxis in vitro. *Neuropeptides*. 2000;34:338–347.
47. Chew V, Lam KP. Leupaxin negatively regulates B cell receptor signaling. *J Biol Chem*. 2007;282:27181–27191.
48. Fan S, Pan Z, Geng Q, et al. Layered signaling regulatory networks analysis of gene expression involved in malignant tumorigenesis of non-resolving ulcerative colitis via integration of cross-study microarray profiles. *PLoS One*. 2013;8:e67142.
49. Yarus AJ, Quintero MA, Jain A, et al. Serum amyloid A as a surrogate marker for mucosal and histologic inflammation in patients with Crohn's disease. *Inflamm Bowel Dis*. 2017;23:158–164.
50. Wen AY, Sakamoto KM, Miller LS. The role of the transcription factor CREB in immune function. *J Immunol*. 2010;185:6413–6419.
51. Ochaion A, Bar-Yehuda S, Cohen S, et al. The anti-inflammatory target A(3) adenosine receptor is over-expressed in rheumatoid arthritis, psoriasis and Crohn's disease. *Cell Immunol*. 2009;258:115–122.
52. Qiao YQ, Huang ML, Xu AT, et al. Lncrna DQ786243 affects treg related CREB and foxp3 expression in Crohn's disease. *J Biomed Sci*. 2013;20:87.