



OPEN ACCESS

## ORIGINAL ARTICLE

# Bacterial protein signals are associated with Crohn's disease

Catherine Juste,<sup>1</sup> David P Kreil,<sup>2,3</sup> Christian Beauvallet,<sup>4</sup> Alain Guillot,<sup>5</sup> Sebastian Vaca,<sup>6</sup> Christine Carapito,<sup>6</sup> Stanislas Mondot,<sup>1</sup> Peter Sykacek,<sup>2</sup> Harry Sokol,<sup>1,7</sup> Florence Blon,<sup>1</sup> Pascale Lepercq,<sup>1</sup> Florence Levenez,<sup>1</sup> Benoît Valot,<sup>5</sup> Wilfrid Carré,<sup>8</sup> Valentin Loux,<sup>8</sup> Nicolas Pons,<sup>1</sup> Olivier David,<sup>9</sup> Brigitte Schaeffer,<sup>9</sup> Patricia Lepage,<sup>1</sup> Patrice Martin,<sup>4</sup> Véronique Monnet,<sup>1</sup> Philippe Seksik,<sup>7</sup> Laurent Beaugerie,<sup>7</sup> S Dusko Ehrlich,<sup>1</sup> Jean-François Gibrat,<sup>8</sup> Alain Van Dorsselaer,<sup>6</sup> Joël Doré<sup>1</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2012-303786>).

For numbered affiliations see end of article.

## Correspondence to

Dr Catherine Juste, Bâtiment 405, INRA Domaine de Vilvert, Jouy-en-Josas 78350, France; [catherine.juste@jouy.inra.fr](mailto:catherine.juste@jouy.inra.fr)

Received 19 September 2012

Revised 16 December 2013

Accepted 17 December 2013

Published Online First

16 January 2014

## ABSTRACT

**Objective** No Crohn's disease (CD) molecular marker has advanced to clinical use, and independent lines of evidence support a central role of the gut microbial community in CD. Here we explore the feasibility of extracting bacterial protein signals relevant to CD, by interrogating myriads of intestinal bacterial proteomes from a small number of patients and healthy controls.

**Design** We first developed and validated a workflow—including extraction of microbial communities, two-dimensional difference gel electrophoresis (2D-DIGE), and LC-MS/MS—to discover protein signals from CD-associated gut microbial communities. Then we used selected reaction monitoring (SRM) to confirm a set of candidates. In parallel, we used 16S rRNA gene sequencing for an integrated analysis of gut ecosystem structure and functions.

**Results** Our 2D-DIGE-based discovery approach revealed an imbalance of intestinal bacterial functions in CD. Many proteins, largely derived from *Bacteroides* species, were over-represented, while under-represented proteins were mostly from Firmicutes and some *Prevotella* members. Most overabundant proteins could be confirmed using SRM. They correspond to functions allowing opportunistic pathogens to colonise the mucus layers, breach the host barriers and invade the mucosae, which could still be aggravated by decreased host-derived pancreatic zymogen granule membrane protein GP2 in CD patients. Moreover, although the abundance of most protein groups reflected that of related bacterial populations, we found a specific independent regulation of bacteria-derived cell envelope proteins.

**Conclusions** This study provides the first evidence that quantifiable bacterial protein signals are associated with CD, which can have a profound impact on future molecular diagnosis.

## INTRODUCTION

Independent lines of evidence converge to suggest a central role of the gut microbial community in Crohn's disease (CD): microbiota is required for the development of inflammation in genetically predisposed colitis animal models,<sup>1</sup> reinfusion of luminal contents after ileal resection rapidly produces recurrent disease in CD patients,<sup>2</sup> antibiotics delay

## Significance of this study

### What is already known on this subject?

- There are unmet needs for diagnosis, treatment and patient monitoring in Crohn's disease (CD).
- No molecular marker has yet advanced to clinical use in CD.
- The intestinal microbiota is recognised as an essential contributor to disease initiation and perpetuation and, therefore, represents an enormous reservoir for the discovery of novel signatures that could be used as biomarkers and predictors for different disease phenotypes or stages.

### What are the new findings?

- The feasibility of extracting bacterial protein signals relevant to CD by interrogating myriads of intestinal bacteria, even from a small number of subjects.
- Twelve bacterial protein signals and one human protein signal (glycoprotein 2 of zymogen granule membranes, GP2) were robustly quantified by targeted MS-based proteomics, without the need for antibodies and ELISA testing. All of them make sense in the context of our understanding of CD.
- Increased IgA at the surface of microbial cells of CD patients coincides with the over-representation of various bacterial proteins with a high immunogenic potential in CD patients.
- Decreased GP2 at the surface of microbial cells of CD patients may favour adhesion of bacteria to the mucosa and then promote inflammation.

### How might it impact on clinical practice in the foreseeable future?

- Using meta-proteome-wide association studies, we point out new potential biomarkers in CD.



Open Access  
Scan to access more  
free content



CrossMark

**To cite:** Juste C, Kreil DP, Beauvallet C, et al. *Gut* 2014;**63**:1566–1577.

postoperative recurrence of CD,<sup>3</sup> and the hitherto identified susceptibility polymorphisms contribute or relate to bacterial sensing through innate and adaptive immune pathways.<sup>4–6</sup> Finally, a vicious

circle of 'mutualism breakdown' has been postulated, where the host does not tolerate its own microbiota any longer, and inflammation can favour the selection of aggressive symbionts.<sup>7</sup> However, the micro-organisms, or microbial products that signal the disruption of gut homeostasis and may have a critical role in CD, are unknown. Their identification remains a significant challenge due to the high complexity of the intestinal microbiota, forming the most densely populated microbial community in the body. It is composed of  $10^{13}$ – $10^{14}$  micro-organisms belonging to about a thousand different species, most of them anaerobic,<sup>8</sup> and is hitherto largely uncultivable (only 20–30% of enteric bacterial species have been propagated in pure culture). Our group<sup>9–11</sup> and others<sup>12–15</sup> have used diverse culture-independent approaches based on molecular profiling of 16S rRNA genes to compare the microbial diversity of intestinal or faecal samples from CD patients and healthy people. Despite the biases inherent to different methodological approaches, varying sampling sites (faeces or mucosa along the intestine), heterogeneity in clinical phenotypes, and variable statistical power, the overall consensus is that the diversity of dominant bacterial species is reduced in CD, notably among members of the Firmicutes phylum.

Largescale metagenomic sequencing (as in MetaHit and the Human Microbiome Project), which analyses whole genomic DNA directly extracted from human intestinal communities, offers a new dimension for the characterisation of these communities from a functional point of view, and represents an enormous reservoir for the discovery of novel signatures that could be used as biomarkers and predictors for different disease phenotypes or stages.<sup>16–18</sup> Beyond functional metagenomics, metaproteomic studies will reveal the true expression of metabolic and cellular functions that govern physiology, become disrupted in disease, and can have a profound impact on molecular diagnosis. Therefore, while there are unmet needs for diagnosis, treatment and patient monitoring in CD, especially while no molecular marker has advanced to clinical use in CD,<sup>19</sup> and considering that the intestinal microbiota is recognised as an essential contributor to disease initiation and perpetuation, we here demonstrate the feasibility of discovering and validating a range of CD-associated bacterial proteins by using a proteomic approach from discovery to validation. With protein profiling providing assays closer to activated functions, such metaproteome-wide association studies have the potential to

become an important tool in modern medicine, and could answer major yet unmet clinical needs.

## METHODS

### Subjects and samples

We conducted a cross-sectional study including six patients with CD (four women and two men, aged 26 through 41 years) and six healthy controls (HC) matched for age, sex and tobacco use (table 1). Patients were followed and selected in the gastroenterology unit of the Saint-Antoine Hospital (Paris). We made a conscious selection of different phenotypes to avoid an unnaturally uniform patient population for this pilot study. Exclusion criteria, however, were active disease with a Harvey–Bradshaw score >5, and any use of antibiotics within the preceding 2 months. The control group comprised healthy volunteers with neither symptoms nor a family history of gastrointestinal disease, and with no use of medication. All participants gave informed consent to the protocol that was approved by the ethics committee of the hospital.

Every participant was asked to provide a fresh stool sample collected at home in a Stomacher 400 plastic bag (Seward Medical), which was left open in a one-litre hermetic plastic box containing a catalyst (Anaerocult, Merck, Darmstadt, Germany) to generate anaerobic conditions. This faecal material was maintained in a coolbox and transferred within 2 h into an anaerobic chamber (90% N<sub>2</sub>, 5% H<sub>2</sub> and 5% CO<sub>2</sub>) for processing. We had verified in preliminary assays, that measurements at a single time point gave a reliable picture of individual metaproteomes, which showed little variation over time (see online supplementary figure S1).

### Preparation of bacterial fractions and diversity profiling

Given the high complexity of faecal samples that contain bacterial, dietary and host proteins, we first extracted bacterial communities, to focus on the collected bacterial proteomes. Bacterial fractions were extracted in duplicate, at low temperature and in an anaerobic atmosphere (see online supplementary method 1, supplementary figure S2), from freshly collected stool specimens. The final bacterial pellets, as well as 150 mg stool aliquots, were kept at –80°C until further analyses. Diversity profiling was performed by 16S rRNA gene pyrosequencing (see online supplementary method 2).

**Table 1** Gender and age of matched participants and clinical characteristics of Crohn's disease patients at the time of stool collection

Controls	Patients						Common to controls and patients
	Designation	Disease location	Disease activity	Diagnosis year	Medication	Surgery	
Gender/age	Gender/age						Smoking
HC.1 F/39 years	CD.1 F/41 years	L1	4	1994	Azathioprine	Small bowel	No
HC.2 M/36 years	CD.2 M/37 years	L2	2	1987	Azathioprine+Prednisone	Small bowel	Yes
HC.3 F/26 years	CD.3 F/29 years	L1+L4	5	1998	Azathioprine	No	Yes
HC.4 M/27 years	CD.4 M/26 years	L1	5	2006	Budesonide	No	No
HC.5 F/41 years	CD.5 F/41 years	L3	2	1990	Azathioprine	Subtotal colectomy	No
HC.6 F/36 years	CD.6 F/38 years	L1	5	2005	Mesalazine	No	No

Disease location according to the Montreal classification: L1 ileum; L2 colon; L3 ileocolon; L4 upper gastrointestinal tract.

Disease activity according to the Harvey–Bradshaw index.<sup>20</sup>

CD, Crohn's disease; HC, healthy controls.

## Discovery of CD-associated gut microbial proteins using 2D-DIGE/LC-MS/MS

We used two-dimensional differential gel electrophoresis (2D-DIGE), coupled with tandem mass spectrometry (MS/MS) and searches against metagenomic databases (MetaHit) as a non-targeted comprehensive approach to discovering CD-associated proteins. Briefly, microbial fractions were extracted in duplicate for the 12 participants leading to 24 samples, which were analysed in a dye-swap design comprising 12 gels in total (see online supplementary table S1 and supplementary method 3). For differential expression analysis, we applied two complementary methods, both established and commonly used in microarray gene expression analysis: a hierarchical analysis of variance (ANOVA) (false discovery rate (FDR) <10%) and an empirical Bayes moderated single-group *t* test *per* gene (FDR <10%). They represent different approaches to the challenge of comparing small sets of samples for thousands of variables (see online supplementary method 4).<sup>21–22</sup> The complementary candidate lists were combined to yield a set of protein spots identified by at least one of the methods as showing significant differences between CD and HC.

For protein identification, nine gels (see online supplementary table S1) were poststained with SYPRO Ruby (BioRad), and spots of interest were robotically excised under computer-assisted visual control. In-gel trypsin digestion and LC-MS/MS analyses are detailed in online supplementary method 5. Finally, we used the X!TandemPipeline to identify and group the differentially expressed proteins (see online supplementary method 6).

## Validation of CD-associated candidate proteins using selected reaction monitoring (SRM)-based targeted proteomics

A targeted LC-SRM assay was developed to validate a subset of 13 candidate proteins discovered in the original 2D-DIGE non-targeted comprehensive survey. The subset of proteins was defined by choosing the ones containing at least two specific peptides for a protein or a group of proteins with identical

function in phylogenetically close bacterial strains, and that at the same time, had already been identified in previous label-free shotgun experiments run on equivalent samples, preferentially without prefractionation. Details on sample preparation, the SRM-assay development, the list of transitions, chromatographic and acquisition conditions, data processing and statistical analysis of the quantitative datasets using MSstats,<sup>23</sup> are given in online supplementary method 7. The 284 optimised transitions measured for the 13/46 targeted proteins/peptides are detailed in online supplementary table S2.

Other general statistical analyses are detailed in online supplementary method 8.

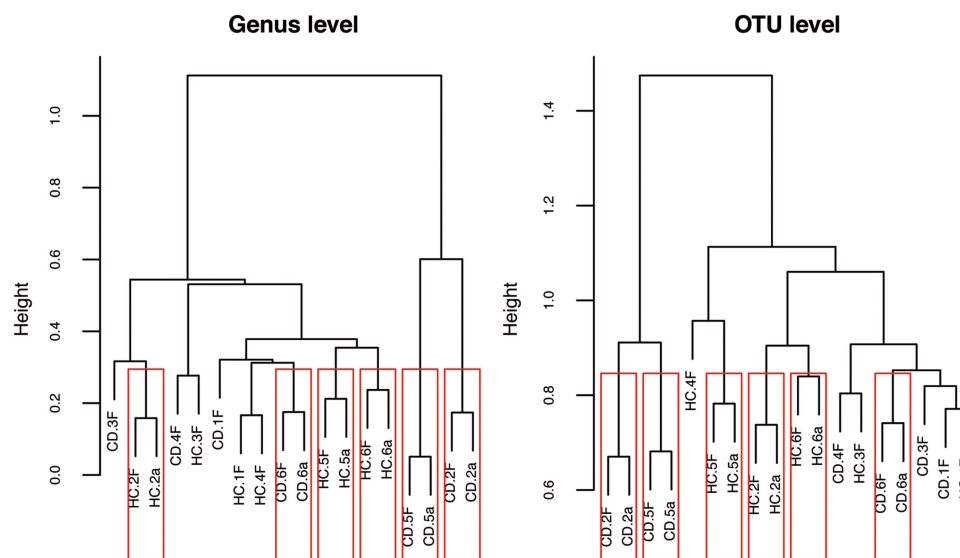
## RESULTS

### Pyrosequencing and quality of the microbial extracts

As illustrated by the dendrogram produced by hierarchical clustering of the 16S rRNA pyrosequencing data at the genus and operational taxonomic unit (OTU) levels (figure 1), microbial extracts were closely related to the corresponding stool total 16S rRNA. This illustrates the ability of our extraction method to preserve the microbial diversity observed in the raw sample material. On the other hand, samples did not cluster by clinical diagnosis, CD, or HC, based on their 16S rRNA gene profile alone. This highlights the need for a complementary proteomics viewpoint.

### Discovery of protein signatures of CD-associated gut microbiota by 2D-DIGE

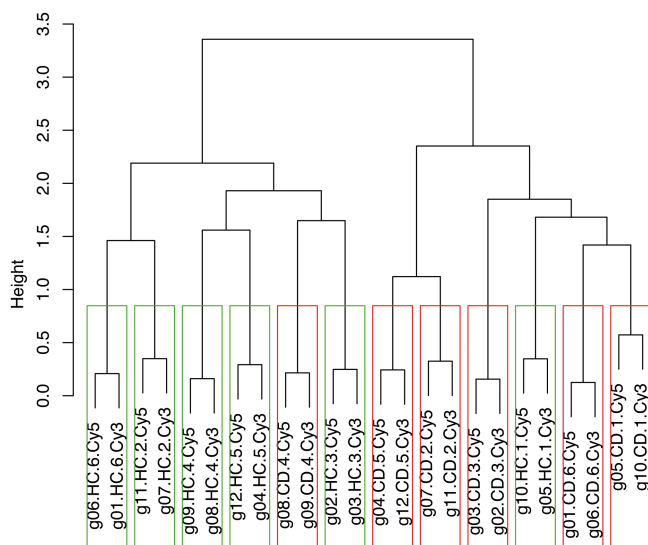
The electrophoretic profile was well conserved across individual samples, and the internal standard (see online supplementary figure S3 and magnified regions thereof in online supplementary figure S4), making it possible to accurately compare spot volumes across the entire experiment. After image alignment and spot co-detection, 2007 protein spots were validated and simultaneously quantified in all 36 images derived from measuring three image channels for each of the 12 gels. There were no pronounced systematic differences between biological replicates



**Figure 1** Structure of all crude samples and half the corresponding extracted microbial pellets profiled by 16S rRNA gene pyrosequencing. Hierarchical clustering of the 16S rRNA gene pyrosequencing dataset (at genus and OTU level) showed a high similarity between population structure of crude samples and those of the corresponding microbial extracts, but did not allow distinguishing between clinical status, CD or HC. HC.1 to HC.6 and CD.1 to CD.6 denote HC and CD patients, respectively; suffix letters, F and a, denote native faeces and bacterial cell extracts, respectively. CD, Crohn's disease; HC, healthy controls.

(samples from a clinical group), and most protein spots (93%) were unchanged between patients and controls allowing reliable normalisation of the data (see online supplementary figure S5). A cluster tree based on the pairwise distances between 2D-DIGE profiles is shown in figure 2. Microbial fractions prepared in duplicate from the same stool specimen always clustered together, reflecting good technical reproducibility, and pairs of duplicates tended to cluster by clinical status, indicating a clinically relevant strong signal. A list of 141 candidate spots (7% of total, 53 increased and 88 decreased in CD patients) was obtained by hierarchical ANOVA or empirical Bayes moderated single-group t test, and all visible selected spots were repeatedly excised from nine SYPRO Ruby poststained gels for LC-MS/MS-based identification (see online supplementary table S1). Eighty-nine spots were found to contain bacterial proteins from a single functional category, and which could be attributed to a defined bacterial subpopulation, with most proteins being from *Bacteroides/Parabacteroides* species, or *Prevotella* species, or members of the order *Clostridiales* (see online supplementary table S3). For robust reporting, however, only a subset of 59 spots were retained (30 increased and 29 decreased in CD patients), which could be identified independently in several gels containing different patient-control pairs (see online supplementary figure S6 for the sequential spot selection process, and see online supplementary table S3 for lists of proteins and peptides). Human proteins were identified in five additional spots with differential signal. Results are summarised in the heat map of figure 3, showing the normalised volumes of those 59 bacterial and 5 human protein spots.

Of the 30 bacterial protein spots which were increased in CD patients, 25 were from the phylum Bacteroidetes, essentially *Bacteroides* species, three were from the phylum Firmicutes,



**Figure 2** Cluster tree based on the pairwise distances between 2D-DIGE profiles. Similarities between patterns (normalised volumes of 2007 spots) were assessed by unsupervised hierarchical clustering. HC.1 to HC.6 and CD.1 to CD.6 denote HC and CD patients, respectively; g01–12 denote gel numbers. Microbial fractions prepared in duplicate from the same stool specimen always clustered together, reflecting good technical reproducibility of our method, and pairs of duplicates tended to cluster by clinical status, CD or HC, indicating a clear clinically relevant signal in the proteomics data. CD, Crohn's disease; HC, healthy controls; 2D-DIGE, two-dimensional difference gel electrophoresis.

order *Clostridiales*, and two were from the phylum Proteobacteria (see the lower half of figure 3). Human proteins IgA immunoglobulins and carboxypeptidase A1 were identified in two additional spots that we found to be over-represented in CD patients (lower half of figure 3). Proteins that were identified in these spots are reported in table 2, where they are organised according to their lineage and function. Of the 29 bacterial protein spots which were decreased in CD patients, 18 were from the phylum Firmicutes, invariably *Clostridiales* whenever order or lower phylogenetic affiliation could be determined, three were from *Prevotella* species, and three others from undefined *Bacteroidales* members, one was from *Escherichia coli*, and four from unknown bacteria (see the upper half of figure 3). Another interesting result was the presence of fragments of human GP2 (pancreatic glycoprotein 2 of zymogen granule membranes) in three under-represented protein spots (upper half of figure 3). Proteins that were identified in these spots are listed in table 3 with their lineage and function.

### Validation of protein signatures of CD-associated gut microbiota by SRM

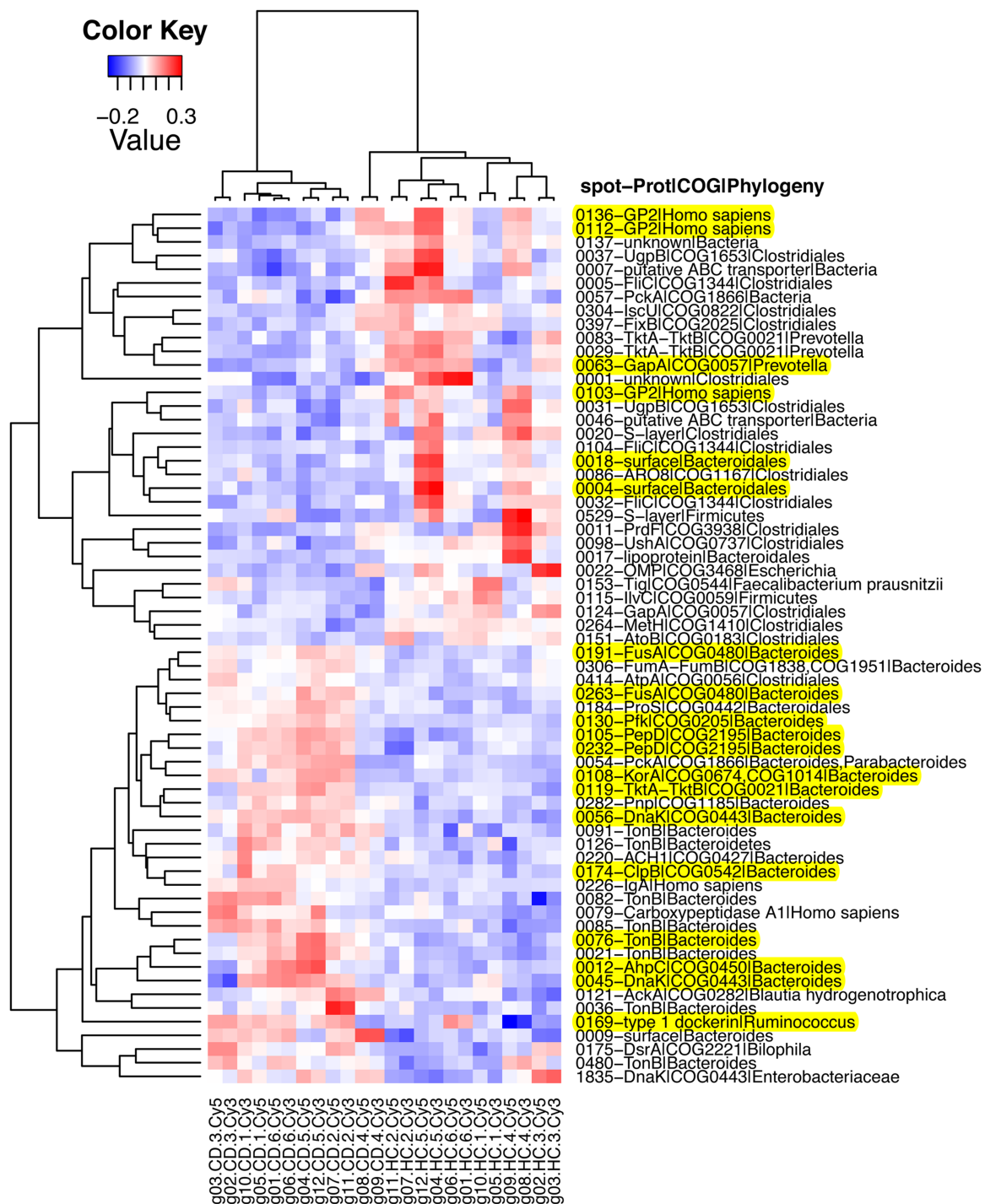
A subset of 13 proteins (highlighted in yellow on figure 3) found to be differentially abundant between CD and HC on the basis of 2D-DIGE were selected to be validated using a targeted LC-SRM assay. Totally, 46 peptides were chosen and 284 transitions were finely optimised using heavy isotope-labelled synthetic peptides spiked into a sample pool in order to allow the precise relative quantification of the 13 candidate proteins in the sample cohort without further sample fractionation other than a stacking gel (see online supplementary table S2). Thus, all 13 proteins could be unambiguously detected with 2–6 specific peptides in single injections of the total bacterial protein extracts. Results are summarised in figure 4 representing the fold change value (differential expression) and the adjusted p value for each targeted protein from the triplicate analyses of the six CD versus six HC individual samples. Details on individual peptide quantification are given in online supplementary table S4. The differential expression of all candidates detected in the discovery experiment was validated and fold changes spanning 14–28 were detected with a very high confidence.<sup>23</sup>

Clearly, we could validate in CD patients a significant elevation of *Bacteroides* proteins that participate in the protection against oxidative stress (AhpC), in protein synthesis, folding and repair (FusA, DnaK and ClpB), in energy saving, and the maintenance of a high carbon flux within both glycolysis and pentose phosphate pathways (PPi-dependent Pfk and TktA-TktB), in the biosynthesis of precursors through the reductive branch of the tricarboxylic acid cycle (KorA), in nutrient acquisition and sensing of the environment (TonB), and in adhesion and colonisation (PepD), while some of these proteins (DnaK, AhpC and TonB-dependent receptors) are recognised for their strong immunogenic properties. We also confirmed elevation of type 1 dockerin from members of the family *Ruminococcaceae*, in CD patients. Other confirmed proteins included the glycolytic enzyme GapA of *Prevotella*, a cell surface protein of undefined *Bacteroidales* members, and the human protein GP2, which were all depleted in CD patients (figure 4).

### Correlating functional shifts with the structure of the bacterial community

We then investigated the question whether the imbalance in bacterial protein abundance that we observed in CD corresponded to changes in gene expression in a stable bacterial community,





**Figure 3** Cluster heat map constructed from the normalised volumes of spots with significant different intensities between HC and CD patients, and that could be robustly identified. Spot numbers and meaningful names for the associated functions are in the right margin. Similarities between patterns are visualised by unsupervised hierarchical clustering. HC.1 to HC.6 and CD.1 to CD.6 denote HC and CD patients, respectively; g01–12 denote gel numbers. Since all spot variables were centred at the mean (the mean has been subtracted to each value), the new mean for each spot variable is now at 0, making half the values negative as indicated in the colour key. Blue and red tones therefore signify under-represented and over-represented, respectively. Proteins highlighted in yellow in the right margin are those that were chosen for SRM validation. As different forms of the same protein (typically TonB-dependent receptors of *Bacteroides*) may occur in different spots, the number of highlighted spots exceeds 13. Proteins annotated 'surface', 'TonB', 'OMP' and 'lipoprotein' in the right margin, may be grouped into 'cell envelope proteins' in the text when several categories are concerned, including proteins of unknown function that have specific features known to be characteristic of cell envelope localisation. CD, Crohn's disease; HC, healthy controls; SRM, selected reaction monitoring.

or whether they reflected a remodelling of the population structure, or whether both events could have occurred. OTU richness estimated by the bias-corrected Chao 1 richness estimator, was significantly lower ( $p=0.015$ ) in the CD group (840 OTUs, SD

259) compared with the HC group (1371 OTUs, SD 286), but the  $\alpha$  diversity Simpson index did not significantly differ between the two groups. This means that a lower number of species was present, which was more evenly distributed in the

**Table 2** List of proteins that were discovered to be over-represented in CD, using the without a priori 2D-DIGE approach

Lineage	Protein name	Cellular function/pathway	2D-DIGE	SRM (fold change)
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	AhpC	Protection against oxidative stress	+	+ (4.4)
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	Pnp	Protection against oxidative stress	+	
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; NA; NA	ProS	Protein synthesis, folding, and repair	+	
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	FusA	Protein synthesis, folding, and repair	+	+ (5.2)
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	DnaK	Protein synthesis, folding, and repair	+	+ (4.1)
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	ClpB	Protein synthesis, folding, and repair	+	+ (4.8)
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	PPI-dependent Pfk	Energy saving and maintenance of a high flux of carbon within both glycolysis and pentose phosphate pathways	+	+ (3.9)
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	TktA-TktB	Energy saving and maintenance of a high flux of carbon within both glycolysis and pentose phosphate pathways	+	+ (5.2)
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	FumA-FumB	Biosynthesis of precursors through the reductive branch of the tricarboxylic acid cycle	+	
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	KorA	Biosynthesis of precursors through the reductive branch of the tricarboxylic acid cycle	+	+ (4.3)
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	TonB-dependent receptors	Nutrient acquisition and sensing of the environment	+	+ (11.3)
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	Other cell envelope proteins	Nutrient acquisition and sensing of the environment	+	
Bacteria; Bacteroidetes; NA; NA; NA; NA	TonB-dependent receptors	Nutrient acquisition and sensing of the environment	+	
Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	PepD	Adhesion and colonisation	+	+ (3.9)
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides	ACH1	Pyruvate metabolism	+	
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Bacteroidaceae; Bacteroides/Parabacteroides	PckA	Energy saving and maintenance of a high flux of carbon within both glycolysis and pentose phosphate pathways	+	
Bacteria; Firmicutes; Clostridia; Clostridiales; Ruminococcaceae; Ruminococcus	type 1 dockerins	Cellulosome assembly	+	+ (27.8)
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	AtpA	ATP production	+	
Bacteria; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Blautia	AckA	ATP production	+	
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales; Enterobacteriaceae; NA	Dnak	Protein synthesis, folding, and repair	+	
Bacteria; Proteobacteria; delta/epsilon subdivisions; Deltaproteobacteria; Desulfobivirionales; Desulfobivirionaceae; Bilophila	DsrA	Energy conservation by reducing sulfite	+	
Human	IgA immunoglobulins	Coating bacterial cells	+	
Human	Carboxypeptidase A1	Coating bacterial cells	+	

Those that were confirmed by SRM are indicated in the last column on the right, with their fold change.

CD, Crohn's disease; SRM, selected reaction monitoring; 2D-DIGE, two-dimensional difference gel electrophoresis; NA, not assigned.

CD group. Specific traits of CD microbiota are illustrated by figure 5. Thirty-four OTUs varied or tended to vary in abundance (see online supplementary table S5). Those that were increased in CD were related to *Bacteroides vulgatus* and *Ruminococcus obeum* (genus *Blautia*) and interestingly included one OTU similar to the potentially anti-inflammatory butyrate-producing bacterium SR1/1, while those that were decreased in CD were related to the butyrate-producing bacterium L2-21, to *Roseburia faecis* (T) M72/1, *Faecalibacterium prausnitzii* A2-165, or other clostridial members, and also included one OTU most similar to *Prevotella oralis*. The heat map of figure 6 shows a positive correlation between the abundance of most of the varying protein groups and the abundance of the related

varying OTUs. There were, however, a number of interesting exceptions suggesting additional effects at work, for instance a subset of TonB proteins attributed to *Bacteroides* members that were increased in CD patients independently of a modulation of the corresponding bacterial populations (see the green-yellow bands on the right middle part of figure 6).

## DISCUSSION

The present work is a clear demonstration that environmental proteomics of gut microbes can provide molecular signatures of IBDs, becoming a powerful complementary tool for their study and, ultimately, their diagnosis and treatment. So far, long lists of candidate biomarker proteins, notably in oncology,

**Table 3** List of proteins that were discovered to be under-represented in CD, using the without a priori 2D-DIGE approach

Lineage	Protein name	Cellular function/pathway	2D-DIGE	SRM (fold change)
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	flagellins FlhC	Nutrient acquisition and sensing of the environment	–	
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	UgpB	Nutrient acquisition and sensing of the environment	–	
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	S-layer proteins	Nutrient acquisition and sensing of the environment	–	
Bacteria; Firmicutes; NA; NA; NA; NA	S-layer proteins	Nutrient acquisition and sensing of the environment	–	
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	UshA	Nucleotide transport and metabolism	–	
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	ARO8	Transcription, amino acid transport and metabolism	–	
Bacteria; Firmicutes; Clostridia; Clostridiales; Ruminococcaceae; Faecalibacterium; Faecalibacterium prausnitzii	Tig	Folding of newly synthesised proteins	–	
Bacteria; Firmicutes; NA; NA; NA; NA	IlvC	Amino acid transport and metabolism	–	
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	Meth	Amino acid transport and metabolism	–	
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	PrdF	Amino acid transport and metabolism	–	
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	IscU	General metabolism	–	
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	FixB	General metabolism	–	
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	AtoB	General metabolism	–	
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	GapA	Glycolysis	–	
Bacteria; Firmicutes; Clostridia; Clostridiales; NA; NA	Unknown function	Unknown	–	
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; NA; NA	Surface proteins	Nutrient acquisition and sensing of the environment	–	–(3.2)
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; NA; NA	Lipoprotein	Nutrient acquisition and sensing of the environment	–	
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella	TktA-TktB	Energy saving and maintenance of a high flux of carbon within both glycolysis and pentose phosphate pathways	–	
Bacteria; Bacteroidetes; Bacteroidia; Bacteroidales; Prevotellaceae; Prevotella	GapA	Glycolysis	–	–(14)
<i>Escherichia</i>	Outer membrane proteins	Nutrient acquisition and sensing of the environment	–	
Bacteria; NA; NA; NA; NA; NA	ABC transporters	Nutrient acquisition and sensing of the environment	–	
Bacteria; NA; NA; NA; NA; NA	PckA	Energy saving and maintenance of a high flux of carbon within both glycolysis and pentose phosphate pathways	–	
Bacteria; NA; NA; NA; NA; NA	Unknown function	Unknown	–	
HUMAN	Fragments of GP2	Coating bacterial cells	–	–(2.4)

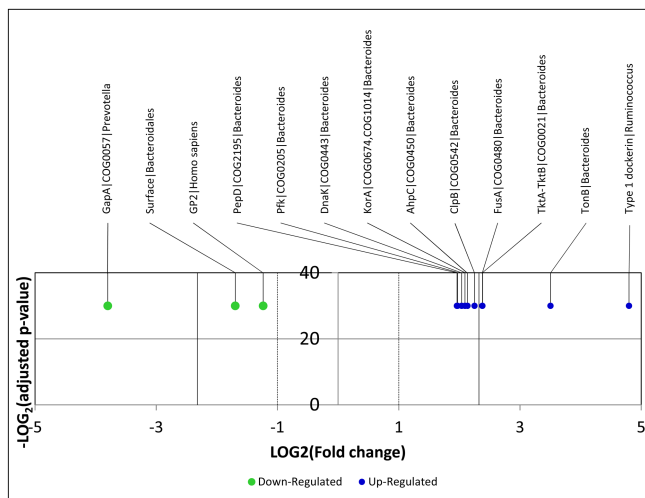
Those that were confirmed by SRM are indicated in the last column on the right, with their fold change.

CD, Crohn's disease; SRM, selected reaction monitoring; 2D-DIGE, two-dimensional difference gel electrophoresis.

remarkably never progressed from research discovery to clinical application because de novo development of multiple ELISA would have been prohibitive in time and money. SRM, by contrast, offers a valuable alternative to antibody-based validations, as it allows the accurate and robust multiplexed quantification of a range of proteins in complex samples without the need for antibodies and ELISA testing. This technology has, for instance, recently been used successfully for the verification of diagnostic and prognostic cancer biomarkers.<sup>24 25</sup> In the present study, we first developed and validated a workflow for the discovery of protein signals specific to CD-associated gut microbial communities without any a priori assumption of the metabolic and/or cellular functions that can accompany CD. Then we developed a SRM assay to confirm a set of candidates, as de novo development of ELISA assays would not have been feasible in a reasonable timeframe and poorly adapted to multiple verifications.

Simple unsupervised hierarchical clustering of samples based on their 2D-DIGE profiles showed that five pairs of replicates out of six within each group, CD and HC, already clustered together, indicating a clear signal, whereas, hierarchical clustering based on 16S rRNA gene pyrosequencing (at the genus and OTU levels) did not allow a distinction between clinical status. This illustrates well that the metaproteomic approach is a powerful tool for highlighting functional imbalances even in the absence of clear major shifts in the dominant bacterial groups or

species. The 2D-DIGE strategy further solved the specific question of detecting differences between groups, a problem that faces even greater challenges in label-free shotgun metaproteomics.<sup>26–31</sup> A set of proteins from members of the Bacteroidetes phylum, largely *Bacteroides* species, were over-represented in CD microbiota. By contrast, under-represented proteins were mostly from *Clostridiales* (Firmicutes phylum), and more rarely from *Prevotella* and undefined *Bacteroidales* members. Functions that we found to be increased in the *Bacteroides/Bacteroidales* population from CD microbiota included proteins corresponding, in general, to functions related to strategic adaptation for survival in challenging environments. For instance, DnaKs and other chaperones, such as ClpB homologues, have a defensive role against oxidative, nitrosative, nutritional, osmotic and pH stresses that are likely to occur in the gut of CD patients.<sup>32–35</sup> AhpCs represent another important defence mechanism to cope with reactive nitrogen intermediates and reactive oxygen species.<sup>36 37</sup> An over-representation of proteins involved in the binding and import of nutrients (TonB-dependent receptors and other cell envelope proteins) could be related to an increased need for carbon substrates and/or for micronutrients to fuel increased central metabolism in *Bacteroidales*.<sup>38 39</sup> Consistent with this, a set of key enzymes that maximise the energy yield from monosaccharide catabolism in *Bacteroides* (PPi-dependent Pfk, PEP-carboxykinase PckA and



**Figure 4** Volcano plot representing results of the LC-SRM assays on the 13 targeted proteins. The logarithmic fold changes (CD vs HC) are plotted against negative logarithmic adjusted p values calculated with the R package MSstats and performed from triplicate injections.<sup>23</sup> All targeted proteins were found to be either upregulated or downregulated in CD patients compared with controls, and the results validated all candidates identified in the discovery experiments. CD, Crohn's disease; HC, healthy controls; SRM, selected reaction monitoring.

fumarate hydratase FumA-FumB) were also over-represented in CD-associated microbial proteomes.<sup>40</sup> Finally, a higher abundance of PepD could contribute to increased surface colonisation by *Bacteroides* members observed in CD patients.<sup>41–44</sup> Remarkably, the overexpression of 10 of these proteins was reliably confirmed by quantitative SRM measurements in unfractionated bacterial protein extracts. Therefore, we reach an attractive hypothesis, that a number of *Bacteroidales* members, essentially *Bacteroides* species, might be adapted or promoted under environmental conditions specific to the gut of CD patients, and that a set of overabundant bacterial proteins that can be quantified by SRM, could be regarded as bacterial signatures for CD. Moreover, all of them make sense in the context of our understanding of CD. Indeed, a number of these proteins have been proposed as major traits for allowing opportunistic pathogens, including *Bacteroides fragilis*, to colonise mucus layers, breach host barriers, and invade the mucosae. For instance, DnaK and AhpC, which are usually intracellular proteins, may also be found in the outer membrane where they bind human plasminogen and enhance its conversion into plasmin by host activators, a scenario which might promote colonisation and host invasion.<sup>45</sup> DnaK proteins exhibit strong immunostimulatory properties both at the level of the innate and adaptive immune system, and can, moreover, promote the processing and presentation of other bacterial or food antigens by chaperoning them.<sup>46</sup> AhpC of several bacteria also demonstrate immunogenic properties as assessed by high titres of seric antibodies against purified/recombinant targets or whole proteome maps,<sup>47</sup> and the  $\beta$ -barrel domains of TonB-dependent receptors are suspected to play an important role in the virulence of Gram-negative bacteria, exposing epitopes on the bacterial surface.<sup>48</sup> Consistent with this and with a previous report,<sup>49</sup> we found an over-representation of microbiota-associated secretory IgA in CD patients. These results encourage inspection of the immunome of the intestinal microbiota to capture those strongly IgA-coated bacteria as well as a set of

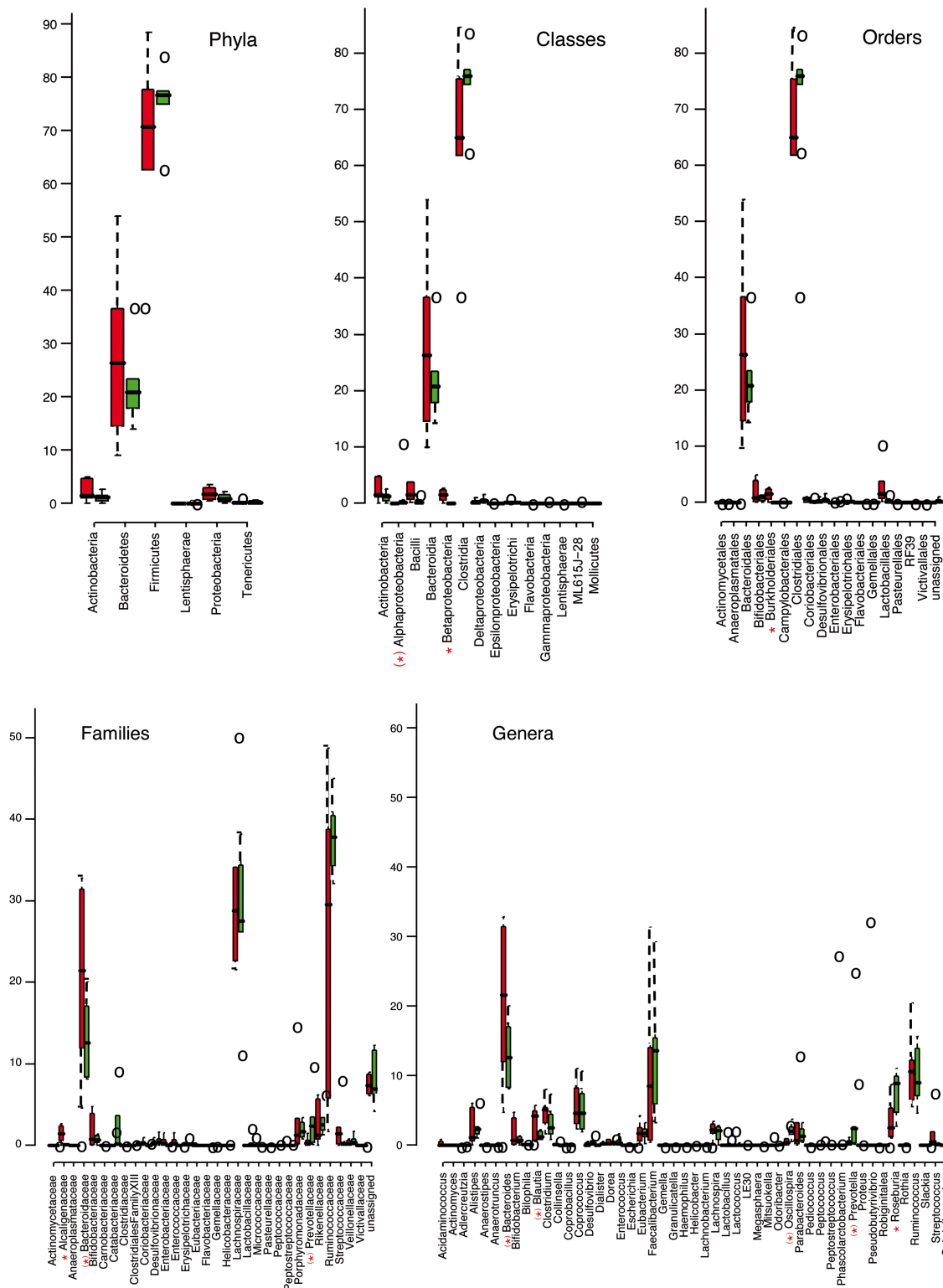
derived antigenic peptides that could be used to map and distinguish specific antibody profiles in subgroups of IBD patients, and thus facilitate appropriate therapeutic options, evaluation of treatment efficacy, and long-term follow-up.

Conversely, many members of the order *Clostridiales* and some of the genus *Prevotella* appeared unable to meet the ecological challenge imposed in CD patients, as judged from the decrease in abundance of key proteins involved in diverse cellular and biochemical functions within this population. For instance, 25 different flagellin FliC proteins attributed to *Clostridiales* members were identified in three under-represented spots, while common enteric flagellins are proposed as major targets of the CD-associated aberrant immune response.<sup>50</sup> The under-representation of trigger factor Tgf attributed to *Faecalibacterium prausnitzii* further suggests that this numerically dominant and potentially anti-inflammatory subgroup,<sup>51</sup> might fail to sustain efficient protein synthesis in CD patients. Finally, under-representation of key enzymes, notably GapA and TktA-TktB, attributed to *Prevotella* members could reflect the inability of this subpopulation in maintaining the flux of carbon within both glycolysis and pentose phosphate pathways. Interestingly, the CD-associated under-representation of GapA from *Prevotella* was confirmed by SRM. Consistent with this, we found that the relative abundances of members of the lineage *Prevotellaceae-Prevotella* tended to decrease in CD patients, and that one OTU assigned to *Prevotella* was under-represented in CD.

While sequencing of the 16S rRNA-encoding genes revealed many positive correlations between the abundances of the varying protein groups and the abundances of the related varying OTUs, it also highlighted a number of unexpected interesting deviations. In particular, some of the TonB-dependent receptors and other uncharacterised proteins localised in bacterial cell envelopes of *Bacteroides* species were increased in CD patients independently of a modulation of the corresponding bacterial populations. Therefore, findings in the present study clearly point to functional changes beyond what can be explained by a mere shift in populations, and extend earlier observations of possible dissociations between structural and functional changes in obese individuals<sup>29–30</sup> or CD patients.<sup>31</sup> The fact that completely different approaches independently identified *Bacteroides* membrane proteins as over-represented in CD,<sup>31</sup> and moreover, that this was confirmed here by SRM, also gives strong support for regarding these proteins as possible relevant bacterial signatures for CD, and highlights the need for future work focusing on microbial cell envelopes in health and disease, inasmuch as this subcellular fraction constitutes the first line of interaction with the host.

Our study also provides the first and unexpected clue towards under-representation of intestinal microbiota-associated GP2 in CD patients. This was confirmed by SRM, and may favour adhesion of bacteria to the mucosa, and then promote inflammation. Indeed, it has been demonstrated recently that recombinant human GP2 binds *Escherichia coli* Type I fimbriae, a bacterial adhesin commonly expressed by members of the intestinal microbiota. A role in host defence has been proposed in which GP2 may serve as a physical barrier that prevents bacteria from binding to host cell receptors.<sup>52</sup> Consistent with this, a higher bacterial biomass on the mucosa, and an adherent mucosal biofilm enriched with *Bacteroides fragilis* were shown to be prominent features in IBD patients.<sup>43</sup> Finally, the question arises as to whether decreased GP2 binding to bacteria might be related to increased anti-GP2 titres reported in CD patients.<sup>53</sup>

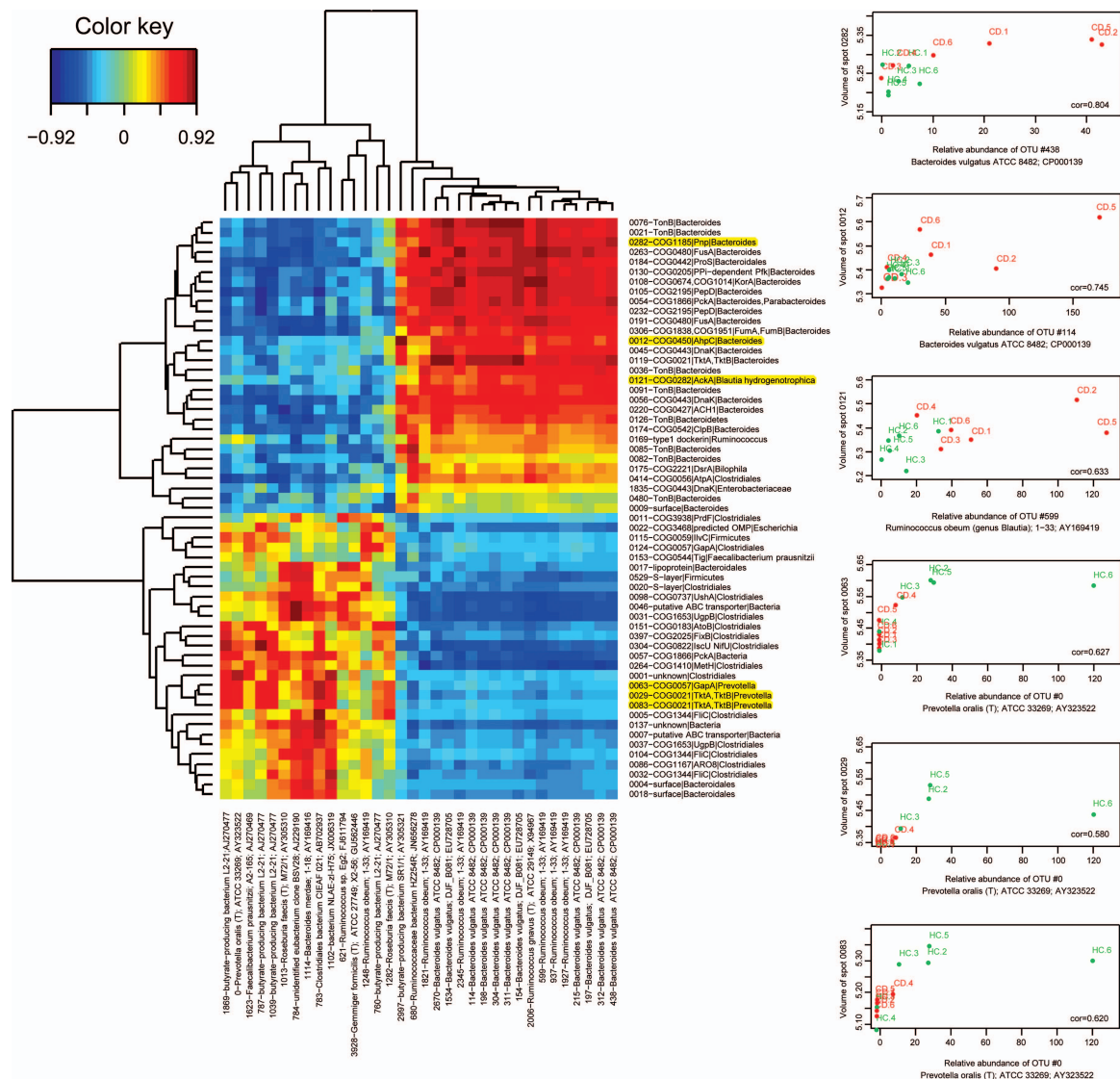




**Figure 5** Box plots of the relative abundances of faecal bacterial populations found by 454 pyrosequencing. Differences between Crohn's patients (□) and HCs (□) at the different phylogenetic levels were considered \*significant for  $p \leq 0.05$ , and (\*) tendencies were reported up to  $p \leq 0.10$  ('glm' with the 'quasibinomial family'). Specific traits of CD microbiota were significantly increased abundances of members in the lineage *Betaproteobacteria-Burkholderiales-Alcaligenaceae*, a tendency towards increased abundances of *Bacteroidaceae-Bacteroides* and *Blautia*, significantly lower numbers of *Roseburia*, and a tendency towards lower numbers of *Alphaproteobacteria*, *Prevotellaceae-Prevotella* and *Oscillospira*. On the other hand, inter-individual variability was higher in CD patients, which is in agreement with heterogeneity of CD. CD, Crohn's disease; HCs, healthy controls.

In conclusion, our metaproteomics approach spanning discovery to confirmation demonstrates, for the first time, the feasibility of extracting bacterial protein signals relevant to CD, by interrogating myriads of intestinal bacteria, even from a small

number of patients and HCs. We provide an initial list of CD-associated microbial proteins extracted from a typical group of patients, which could represent major common features in CD patients. The next step should be to validate the specificity



**Figure 6** Heat map of the correlation matrix between abundance of the varying protein groups and the abundance of the related varying OTUs (left panel). Red and hotter orange tones indicate a positive correlation between abundance of a protein group and abundance of the corresponding OTUs; green and yellow tones indicate absence of correlation as for example, TonB proteins and other uncharacterised surface proteins of *Bacteroides* in spots 0082, 0480 and 0009, and DnaK of *Enterobacteriaceae* in spot 1835 (right middle part of the image). Activities that were found to be increased and decreased in CD patients fell in the upper and lower halves of the image, respectively. OTUs that were found to be increased and decreased in abundance in CD patients fell into the right and left halves of the image, respectively. Examples of positive correlations are detailed on right panel and highlighted in yellow on panel A. CD, Crohn's disease.

and sensitivity of bacterial protein signals either in individual clinical trials with well-defined and homogenous CD populations, or in a comprehensive study with a larger heterogeneous patient cohort. Given that effects are harder to detect in smaller samples, one can expect that even more subtle differences could be detected in larger cohorts, and that inclusion and accurate quantification of additional predefined sets of proteins could be used to refine recognition of IBD entities in the very near future. The extraction and robust quantification of bacterial protein signals is also a way to identify disrupted protein networks that drive onset and perpetuation of CD and, therefore, candidate targets for IBD treatment based on gut-ecological intervention strategies.

#### Author affiliations

<sup>1</sup>UMR1319 Micalis, INRA, Jouy-en-Josas, France

<sup>2</sup>Chair of Bioinformatics, Boku University Vienna, Vienna, Austria

<sup>3</sup>Department of Life Sciences, University of Warwick, Warwickshire, UK

<sup>4</sup>UMR1313 GABI, Iso Cell Express (ICE), INRA, Jouy-en-Josas, France

<sup>5</sup>Plate-forme d'Analyse Protéomique de Paris Sud-Ouest (PAPPSO), INRA, Gif-sur-Yvette, France

<sup>6</sup>Laboratoire de Spectrométrie de Masse BioOrganique (LSMBO), IPHC, Université de Strasbourg, Strasbourg, France

<sup>7</sup>Gastroenterology and Nutrition Unit, Hôpital Saint-Antoine, AP-HP, Paris, France

<sup>8</sup>UR1077, Mathématique Informatique et Génome (MIG), INRA, Jouy-en-Josas, France

<sup>9</sup>UR341, Mathématiques et Informatique Appliquées (MIA), INRA, Jouy-en-Josas, France

**Acknowledgements** We are grateful to Bertrand Nicolas for his help in the preparation of figures.

**Contributors** CJ, DPK, CC, HS, PS, LB, JD: conception and design of the study; CB, AG, SV, CC, SM, PL, FB, FL, WC, VL, NP: acquisition of data; CJ, DPK, CB, AG, SV, CC, SM, FB, WC, VL, NP: analysis and interpretation of data; CJ, DPK, CB, SV, CC: drafting of the manuscript; DPK, AG, CC, HS, PS, J-FG, SDE, AVD, JD, BV, PM, VM: critical revision of the manuscript for important intellectual content; CJ, DPK,

SV, CC, OD, BS: statistical analysis; CB, AG: technical or material support; CJ, HS, PS, LB: study supervision.

**Funding** The Boku Chair of Bioinformatics acknowledges funding by the Vienna Science and Technology Fund (WWTF), Baxter AG, Austrian Research Centres Seibersdorf, and the Austrian Centre of Biopharmaceutical Technology. We acknowledge the 'Fondation pour la Recherche Médicale' for funding the triple quadrupole instrument for targeted proteomics experiments.

**Competing interests** None.

**Patient consent** Obtained.

**Ethics approval** This study was conducted with the approval of the Ethics Committee of the St Antoine hospital, Paris.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Open Access** This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 3.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/3.0/>

## REFERENCES

- Rath HC, Herfarth HH, Ikeda JS, *et al.* Normal luminal bacteria, especially *Bacteroides* species, mediate chronic colitis, gastritis, and arthritis in HLA-B27/ human  $\beta$ 2 microglobulin transgenic rats. *J Clin Invest* 1996;98:945–53.
- D'Haens GR, Geboes K, Peeters M, *et al.* Early lesions of recurrent Crohn's disease caused by infusion of intestinal contents in excluded ileum. *Gastroenterology* 1998;114:262–7.
- Rutgeerts P, Hiele M, Geboes K, *et al.* Controlled trial of metronidazole treatment for prevention of Crohn's recurrence after ileal resection. *Gastroenterology* 1995;108:1617–21.
- Mathew CG. New links to the pathogenesis of Crohn disease provided by genome-wide association scans. *Nat Rev Genet* 2008;9:9–14.
- Cho JH. The genetics and immunopathogenesis of inflammatory bowel disease. *Nat Rev Immunol* 2008;8:458–66.
- Cleynen I, González JR, Figueroa C, *et al.* Genetic factors conferring an increased susceptibility to develop Crohn's disease also influence disease phenotype: results from the IBDchip European Project. *Gut* 2013;62:1556–65.
- Cerf-Bensussan N, Gaboriau-Routhiau V. The immune system and the gut microbiota: friends or foes? *Nat Rev Immunol* 2010;10:735–44.
- Tap J, Mondot S, Levenez F, *et al.* Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol* 2009;11:2574–84.
- Sekik P, Rigottier-Gois L, Gramet G, *et al.* Alterations of the dominant faecal bacterial groups in patients with Crohn's disease of the colon. *Gut* 2003;52:237–42.
- Mangin I, Bonnet R, Sekik P, *et al.* Molecular inventory of faecal microflora in patients with Crohn's disease. *FEMS Microbiol Ecol* 2004;50:25–36.
- Manichanh C, Rigottier-Gois L, Bonnaud E, *et al.* Reduced diversity of faecal microbiota in Crohn's disease revealed by a metagenomic approach. *Gut* 2006;55:205–11.
- Kleessen B, Kroesen AJ, Buhr HJ, *et al.* Mucosal and invading bacteria in patients with inflammatory bowel disease compared with controls. *Scand J Gastroenterol* 2002;37:1034–41.
- Scanlan PD, Shanahan F, O'Mahony C, *et al.* Culture-independent analyses of temporal variation of the dominant fecal microbiota and targeted bacterial subgroups in Crohn's disease. *J Clin Microbiol* 2006;44:3980–8.
- Gophna U, Sommerfeld K, Gophna S, *et al.* Differences between tissue-associated intestinal microfloras of patients with Crohn's disease and ulcerative colitis. *J Clin Microbiol* 2006;44:4136–41.
- Frank DN, St Amand AL, Feldman RA, *et al.* Molecular-phylogenetic characterization of microbial community imbalances in human inflammatory bowel diseases. *Proc Natl Acad Sci USA* 2007;104:13780–5.
- Qin J, Ruiqiang L, Raes J, *et al.* A human gut microbial gene catalog established by deep metagenomic sequencing. *Nature* 2010;464:59–65.
- Human Microbiome Jumpstart Reference Strains Consortium Nelson KE, Weinstock GM, Highlander SK, *et al.* A catalog of reference genomes from the human microbiome. *Science* 2010;328:994–9.
- Le Chatelier E, Nielsen T, Qin J, *et al.* Richness of human gut microbiome correlates with metabolic markers. *Nature* 2013;500:541–6.
- Vermeire S, Ferrante M, Rutgeerts P. Recent advances: personalised use of current Crohn's disease therapeutic options. *Gut* 2013;62:1511–15.
- Harvey RF, Bradshaw JM. A simple index of Crohn's disease activity. *The Lancet* 1980;1:514.
- Kreil DP, Kar NA, Lilley KS. DNA microarray normalization methods can remove bias from differential protein expression analysis of 2D difference gel electrophoresis results. *Bioinformatics* 2004;20:2026–34.
- Karp NA, Kreil DP, Lilley KS. Determining a significant change in protein expression with DeCyder during a pair-wise comparison using two-dimensional difference gel electrophoresis. *Proteomics* 2004;4:1421–32.
- Chang CY, Picotti P, Hüttenhain R, *et al.* Protein significance analysis in selected reaction monitoring (SRM) measurements. *Mol Cell Proteomics* 2012;11:M111. 014662.
- Pan S, Chen R, Brand RE, *et al.* Multiplex targeted proteomic assay for biomarker detection in plasma: a pancreatic cancer biomarker case study. *J Proteome Res* 2012;11:1937–48.
- Hüttenhain R, Soste M, Selevsek N, *et al.* Reproducible quantification of cancer-associated proteins in body fluids using targeted proteomics. *Sci Transl Med* 2012;4:142ra94.
- Verberkmoes NC, Russell AL, Shah M, *et al.* Shotgun metaproteomics of the human distal gut microbiota. *ISME J* 2009;3:179–89.
- Cantarel BL, Erickson AR, Verberkmoes NC, *et al.* Strategies of metagenomic-guided whole-community proteomics of complex microbial environments. *PLoS ONE* 2011;6:e21713.
- Rooijers K, Kolmeder C, Juste C, *et al.* An iterative workflow for mining the human intestinal metaproteome. *BMC Genomics* 2011;12:6. <http://www.biomedcentral.com/1471-2164/12/6>
- Kolmeder CA, de Been M, Nikkilä J, *et al.* Comparative metaproteomics and diversity analysis of human intestinal microbiota testifies for its temporal stability and expression of core functions. *PLoS ONE* 2012;7:e29913.
- Ferrer M, Ruiz A, Lanza F, *et al.* Microbiota from the distal guts of lean and obese adolescents exhibit partial functional redundancy besides clear differences in community structure. *Environ Microbiol* 2013;15:211–26.
- Erickson AR, Cantarel BL, Lamendella R, *et al.* Integrated metagenomics/ metaproteomics reveals human host-microbiota signatures of Crohn's disease. *PLoS ONE* 2012;7:e49138.
- Keshavarzian A, Banan A, Farhadi A, *et al.* Increases in free radicals and cytoskeletal protein oxidation and nitration in the colon of patients with inflammatory bowel disease. *Gut* 2003;52:720–8.
- Schill R, Breuer RI, Klein F, *et al.* Comparison of the composition of faecal fluid in Crohn's disease and ulcerative colitis. *Gut* 1982;23:326–32.
- Nugent SG, Kumar D, Rampton DS, *et al.* Intestinal luminal pH in inflammatory bowel disease: possible determinants and implications for therapy with aminosaliclates and other drugs. *Gut* 2001;48:571–7.
- Baumgart DC, Sandborn WJ. Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet* 2007;369:1641–57.
- Chen L, Xie QW, Nathan C. Alkyl hydroperoxide reductase subunit C (AhpC) protects bacterial and human cells against reactive nitrogen intermediates. *Mol Cell* 1998;1:795–805.
- Rocha ER, Smith CJ. Role of the alkyl hydroperoxide reductase (ahpCF) gene in oxidative stress defense of the obligate anaerobe *Bacteroides fragilis*. *J Bacteriol* 1999;181:5701–10.
- Spence C, Wells WG, Smith CJ. Characterization of the primary starch utilization operon in the obligate anaerobe *Bacteroides fragilis*: regulation by carbon source and oxygen. *J Bacteriol* 2006;188:4663–72.
- Martens EC, Koropatkin NM, Smith TJ, *et al.* Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* 2009;284:24673–7.
- Macy JM, Ljungdahl LG, Gottschalk G. Pathway of succinate and propionate formation in *Bacteroides fragilis*. *J Bacteriol* 1978;134:84–91.
- Van Houdt R, Michiels CW. Role of bacteria cell surface structures in *Escherichia coli* biofilm formation. *Res Microbiol* 2005;156:626–33.
- Stoodley P, Sauer K, Davies DG, *et al.* Biofilms as complex differentiated communities. *Annu Rev Microbiol* 2002;56:187–209.
- Swidsinski A, Weber J, Loening-Baucke V, *et al.* Spatial organization and composition of the mucosal flora in patients with inflammatory bowel disease. *J Clin Microbiol* 2005;43:3380–9.
- Kleessen B, Kroesen AJ, Buhr HJ, *et al.* Mucosal and invading bacteria in patients with inflammatory bowel disease compared with controls. *Scand J Gastroenterol* 2002;37:1034–41.
- Knaust A, Martin VR, Weber MVR, *et al.* Cytosolic proteins contribute to surface plasminogen recruitment of *Neisseria meningitidis*. *J Bacteriol* 2007;189:3246–55.
- Routsias JG, Tzioufas AG. The role of chaperone proteins in autoimmunity. *Ann NY Acad Sci* 2006;1088:52–64.
- Chitlaru T, Gat O, Grosfeld H, *et al.* Identification of in vivo-expressed immunogenic proteins by serological proteome analysis of the *Bacillus anthracis* secretome. *Infect Immun* 2007;75:2841–52.

- 48 Gribun A, Katcoff DJ, Hershkovits G, *et al.* Cloning and characterization of the gene encoding for OMP-PD porin: the major *Photobacterium damsela* outer membrane protein. *Curr Microbiol* 2004;48:167–74.
- 49 van der Waaij LA, Kroese FGM, Visser A, *et al.* Immunoglobulin coating of faecal bacteria in inflammatory bowel disease. *Eur J Gastroenterol Hepatol* 2004;16:669–74.
- 50 Sitaraman SV, Klapproth JM, Moore DA III, *et al.* Elevated flagellin-specific immunoglobulins in Crohn's disease. *Am J Physiol Gastrointest Liver Physiol* 2005;288:G403–6.
- 51 Sokol H, Pigneur B, Watterlot L, *et al.* *Faecalibacterium prausnitzii* is an anti-inflammatory commensal bacterium identified by gut microbiota analysis of Crohn disease patient. *Proc Natl Acad Sci U S A* 2008;105:16731–6.
- 52 Yu S, Lowe AW. The pancreatic zymogen granule membrane protein, GP2, binds *Escherichia coli* type 1 Fimbriae. *BMC Gastroenterol* 2009;9: 58–64.
- 53 Roggenbuck D, Hausdorf G, Martinez-Gamboa L, *et al.* Identification of GP2, the major zymogen granule membrane glycoprotein, as the autoantigen of pancreatic antibodies in Crohn's disease. *Gut* 2009;58:1620–8.