

Single-cell microbiota phenotyping reveals distinct disease and therapy-associated signatures in Crohn's disease

Lisa Budzinski^{a,b}, Gi-Ung Kang^a, René Riedel^{a,c}, Toni Sempert^a, Leonie Lietz^{a,b}, René Maier^a, Janine Büttner^d, Bettina Bochow^d, Marcell T. Tordai^e, Aayushi Shah^{a,b}, Amro Abbas^a, Tanisha Momtaz^{a,f}, Jannike L. Krause^a, Robin Kempkens^a, Katrin Lehman^a, Gitta A. Heinz^a, Anne E. Benken^g, Stefanie Bartsch^h, Kathleen Necke^h, Ute Hoffmann^a, Mir-Farzin Mashreghi^a, Robert Biesen^g, Tilmann Kallinich^h, Tobias Alexander^{a,g}, Bosse Jessen^e, Carl Weidinger^e, Britta Siegmund^{e,i}, Andreas Radbruch^a, Anja Schirbel^d, Benjamin Moser^{d,j}, and Hyun-Dong Chang^{a,b}

^aGerman Rheumatology Research Centre Berlin – A Leibniz Institute, Berlin, Germany; ^bDepartment for Cytometry, Institute of Biotechnology, Technische Universität Berlin, Berlin, Germany; ^cBioinformatics and Computational Biology, Department of Cardiology, University Medical Centre Hamburg-Eppendorf, Hamburg, Germany; ^dDepartment of Hepatology and Gastroenterology, Campus Charité Mitte, Charité, Universitätsmedizin Berlin corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany; ^eDepartment of Gastroenterology, Infectious Diseases and Rheumatology, Campus Benjamin Franklin, Charité – Universitätsmedizin Berlin corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany; ^fSchool of Pharmacy, BRAC University, Dhaka, Bangladesh; ^gDepartment of Rheumatology, Campus Charité Mitte, Charité – Universitätsmedizin Berlin corporate member of Freie Universität Berlin and Humboldt-Universität zu Berlin, Berlin, Germany; ^hDepartment of Paediatric Respiratory Medicine, Immunology and Critical Care Medicine, Charité Campus Virchow, Charité Universitätsmedizin Berlin, Berlin, Germany; ⁱBIH Charité Clinician Scientist Program; ^jDRK Kliniken Berlin, Clinic for internal medicine – Gastroenterology, Haematology and Oncology, Nephrology, Centre for chronic gastrointestinal inflammations, Berlin, Germany

ABSTRACT

IgA-coated fractions of the intestinal microbiota of Crohn's disease (CD) patients have been shown to contain taxa that hallmark the compositional dysbiosis in CD microbiomes. However, the correlation between other cellular properties of intestinal bacteria and disease has not been explored further, especially for features that are not directly driven by the host immune-system, e.g. the expression of surface sugars by bacteria. By sorting and sequencing IgA-coated and lectin-stained fractions from CD patients microbiota and healthy controls, we found that lectin-stained bacteria were distinct from IgA-coated bacteria, but still displayed specific differences between CD and healthy controls. To exploit the discriminatory potential of both, immunoglobulin coated bacteria and the altered surface sugar expression of bacteria in CD, we developed a multiplexed single cell-based analysis approach for intestinal microbiota. By multi-parameter microbiota flow cytometry (mMFC) we characterized the intestinal microbiota of 55 CD patients and 44 healthy controls for 11-parameters in total, comprising host-immunoglobulin coating and the presence of distinct surface sugar moieties. The data were analyzed by machine-learning to assess disease-specific marker patterns in the microbiota phenotype. mMFC captured detailed characteristics of CD microbiota and identified patterns to classify CD patients. In addition, we identified phenotypic signatures in the CD microbiota which not only reflected remission after 6 weeks of anti-TNF treatment, but were also able to predict remission before the start of an adalimumab treatment course in a pilot study. We here present the proof-of-concept demonstrating that multi-parameter single-cell bacterial phenotyping by mMFC could be a novel tool with high translational potential to expand current microbiome investigations by phenotyping of bacteria to identify disease- and therapy-associated cellular alterations and to reveal novel target properties of bacteria for functional assays and therapeutic approaches.

ARTICLE HISTORY

Received 29 August 2024
Revised 20 December 2024
Accepted 7 January 2025

KEYWORDS

Single-cell analysis;
microbiota flow cytometry;
microbiota phenotyping

Introduction

The human intestinal microbiota are intricately linked to human health. They play an essential role in host energy homeostasis and metabolism, but also contribute significantly to the maturation

of the immune system as a steady interaction partner. In consequence, various human diseases, ranging from metabolic to chronic inflammatory diseases and cancer to neurological disorders,¹ have been associated with alterations in the

CONTACT Hyun-Dong Chang  chang@drfz.de  German Rheumatology Research Centre Berlin, A Leibniz Institute, Charitéplatz 1, 10117 Berlin, Germany
 Supplemental data for this article can be accessed online at <https://doi.org/10.1080/19490976.2025.2452250>

© 2025 The Author(s). Published with license by Taylor & Francis Group, LLC.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

composition of the intestinal microbiota. Typically, the microbiota is characterized by sequencing the highly conserved 16S rRNA gene, encoding a bacteria-specific ribosomal RNA of the small ribosomal subunit, which contains variable regions allowing the determination of phylogenetic relationships between bacteria, and thus taxonomic classification. Applied to cohorts of, e.g., patients and healthy controls, 16S rRNA sequencing has revealed alterations in the abundance or presence of certain bacterial taxa, the overall compositional diversity, and the microbial load in multiple diseases. However, the vast inter-individual diversity in the microbiota composition in humans has interfered with the identification of disease-specific taxonomic microbial signatures as biomarkers. Although many studies have contributed to research on the pathogenic potential of certain intestinal microbes, the lack of ubiquitousness and robustness have so far precluded routine clinical application toward the benefit of patients.^{2–4}

Results from the Human Microbiome Project have indicated that despite high taxonomic diversity between individuals, functions of the microbial community are rather conserved, suggesting that the microbial composition is governed by functionality and interaction with the host.⁵ In patients with Crohn's disease (CD), the microbiota have been found to possess distinct metabolomic profiles compared to healthy donors.⁶ Additionally, drastic changes in the host-microbiota interaction and a modified immune response toward components of the microbiota are reflected by an altered immunoglobulin secretion and coating of intestinal bacteria with host immunoglobulins in CD.^{7–10} Consequently, it appears promising to investigate the microbiota as a community of single cells, each a functional unit, shaped by their micro-environment and host-derived factors. Alongside host immunoglobulins, surface sugar moieties appear promising to reflect on microbiota-host interactions as surface glycosylation of bacteria may correlate with metabolic activity, nutritional state and an inflammatory microenvironment resulting in altered interplay between bacteria or with the host.¹¹ Clinical relevance of lectin binding to bacteria in CD has been suggested by linking the binding of the host lectin intelectin-1 to the pathogenesis of intestinal inflammation.¹² However, alterations regarding sugar moieties of the intestinal microbiota in CD have not yet been investigated.

Flow cytometry is a widely used tool to rapidly investigate cellular properties on single-cell level, but its potential has not yet been fully explored for microbiota analyses.¹³ Microbiota flow cytometry (MFC) has been shown to be an effective method to capture compositional dynamics of microbial communities,^{14–17} assessing their complexity and compositional changes by light scatter properties and quantitative DNA staining.^{14,15,18} We and others have previously demonstrated the utility of MFC for monitoring microbiota dynamics longitudinally, both *in vitro* and *ex vivo* in murine colitis,^{16,19} during chemotherapy-treatment of patients with hematological malignancy¹⁷ and to discriminate CD patients from healthy donors.²⁰

Here, we present an advanced multi-parameter microbiota flow cytometry (mMFC) approach to analyze single bacterial cells in complex communities of the human intestinal microbiome phenotypically. Initially, by IgA-Seq, we confirmed that IgA-coating reflects adapted immune responses to shifts in the microbial community as the IgA-coated fractions in CD patients were mainly composed of taxa that also characterized the CD dysbiosis obtained by 16S rRNA sequencing. In contrast, lectin-Seq, i.e. 16S rRNA-based identification of bacteria based on staining and sorting with plant-derived lectins, for wheat germ agglutinin (WGA, target: N-acetyl-glucosamine) and peanut agglutinin (PNA, target: galactose), indicated that the surface sugar expression may mainly be a dynamic property of the microbes and sensitive to the micro-environment.

We have combined the isotype-specific detection of coating with host immunoglobulins (IgA1, IgA2, IgM and IgG) with the staining of sugar residues by lectins on the bacterial surface on a set of stool samples of CD patients to interrogate their microbial phenotypes for disease- and therapy-related biosignatures.²¹ Comparable to what has been described for blood cell phenotyping, we have reduced the complexity of high-dimensional data by clustering microbial cells by phenotypic similarity. Using machine-learning, we could classify CD patients despite cohort heterogeneity from healthy controls. In addition, microbiota phenotyping revealed a biosignature in CD patients stratifying patients achieving remission criteria in response to anti-TNF therapy during but also prior to start of the therapy.

In summary, we demonstrate that mMFC-based single-cell phenotyping of intestinal bacteria is a robust approach to assess altered microbiota properties reflecting the adaptation of the mucosal immune response to taxonomic shifts and phenotypic features of a dysbiotic microbial community in an inflammatory environment so far overlooked by taxonomic profiling.

Results

Host IgA-coating but not lectin-staining mark known colitogenic bacteria

We isolated bacteria based on IgA-coating, WGA-binding and PNA-binding, respectively, from 14 CD patients (CD cohort 1) and 8 healthy controls

with fluorescence-activated cell sorting and subjected them to full-length 16S rRNA gene sequencing to characterize their composition in more detail. We used beta-dissimilarity by Bray-Curtis to compare the composition between samples or sorted fractions, which considers presence and distribution of taxa and ranges from 0 for samples of equal composition to 1 for maximally different samples. To visualize differences between the samples of the two groups, we computed the distance in beta-dissimilarity between donors and projected the distribution to a principal coordinate analysis plot.

The taxonomic composition of the IgA-coated fractions of both cohorts was significantly different between CD and healthy samples (Figure 1a, $R^2 = 0.1$, $p = 0.007$). Bacterial taxa found to be

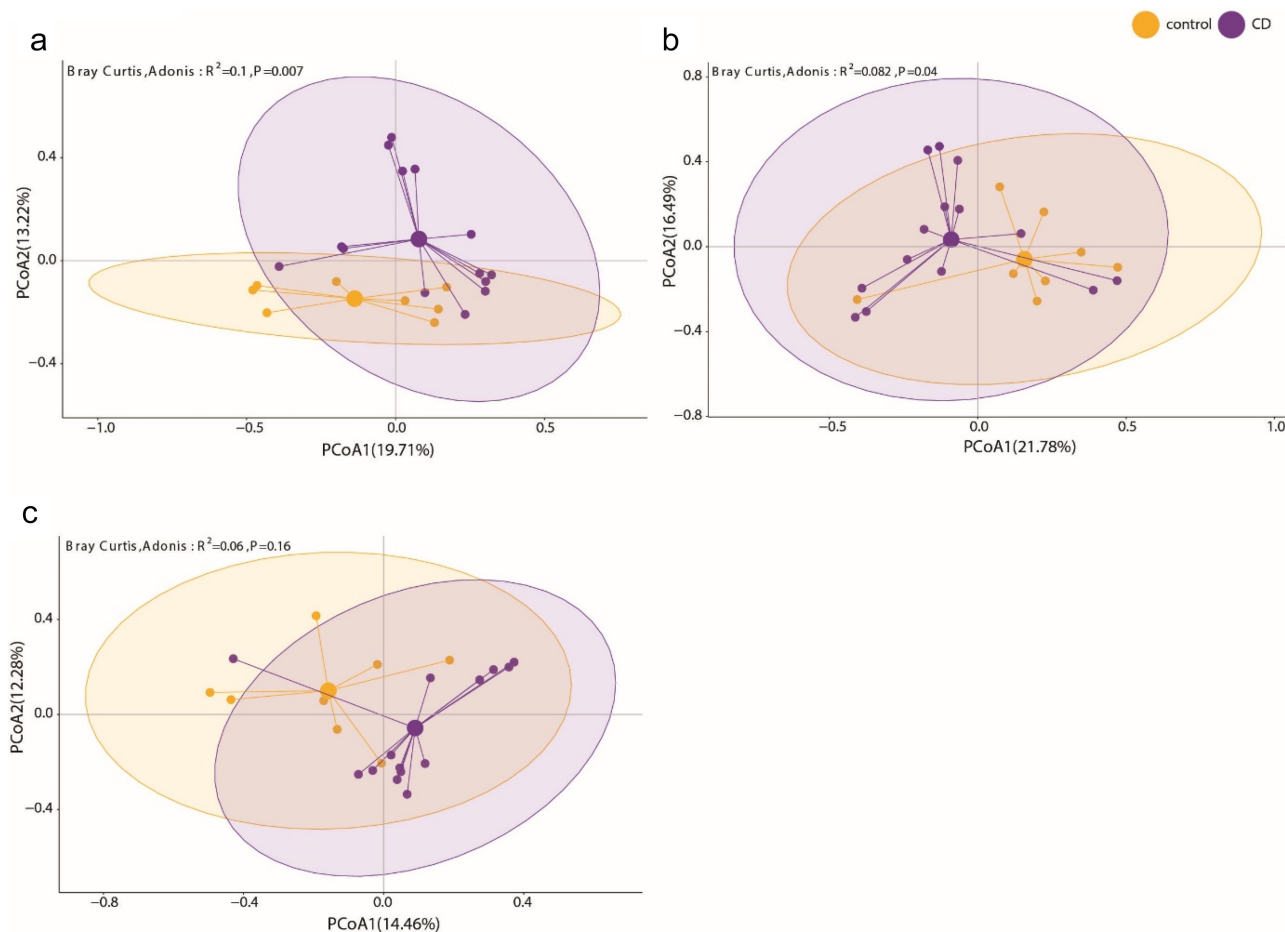


Figure 1. IgA-seq and lectin-seq potentially reveal functionally distinct bacterial taxa. Bacteria from CD patients ($n = 14$, purple) and healthy donors ($n = 8$, orange) were isolated by fluorescence activated cell sorting (FACS) according to (a) IgA-coating, (b) staining with wheat germ agglutinin (WGA) and (c) staining with peanut agglutinin (PNA) and submitted to full-length 16S rRNA gene sequencing. (a, b, c) the principal coordinate projection represent the Bray-Curtis dissimilarity of the taxonomic composition of the respective sorted fractions of CD and healthy donor samples.

enriched in the IgA-coated fraction in CD patients included *Ruminococcus gnavus* sp. and *Escherichia-Shigella* sp. which were also dominant in CD samples on bulk level (Supplementary Figure S1, Supplementary Figure S2). While in healthy donors, *Faecalibacterium* sp., *Subdoligranulum* sp. and *Lachnospiraceae* NK4A136 sp. were enriched in the IgA-coated bacteria and were also dominant in the bulk sample. A less apparent but still significant separation was observed between CD and healthy controls when considering WGA-stained bacteria (Figure 1b, $R^2 = 0.082$, $p = 0.04$). We here found *Ruminococcus gnavus* group sp. and *Escherichia-Shigella* sp. enriched exclusively in the CD sorts, while again *Faecalibacterium prausnitzii* and *Lachnospiraceae* NK4A136 sp. were representatives for the sorted fractions from controls (Supplementary Figure S3A). Similar results were found when sorting the PNA-positive bacteria, albeit they did not show significant differences in composition between CD and healthy (Figure 1c, $R^2 = 0.06$, $p = 0.16$). Here, *Ruminococcus gnavus* group sp. was not enriched in CD. *Escherichia-Shigella* sp. and *Lachnoclostridium* sp. were still found enriched among the PNA-stained bacteria in CD (Supplementary Figure S3B). Overall, *Veillonella* sp. which was prominently found in CD samples when looking at the total microbiome, was not found in any of the sorts. Instead *Blautia* sp. and *Stenotrophomonas* sp. which were not conspicuous in the total microbiome were found significantly enriched in the respective enriched populations. This raised the question, how well the taxonomic composition and microbial phenotypes represent the CD and healthy state.

Single-cell analysis reveals Crohn's disease specific phenotypic alterations of the microbiota

To address this in more detail, we enlarged our cohort and characterized 55 CD samples (CD cohort 1) and compared their microbiota phenotype to healthy controls (HC, $n = 44$). Bacteria were stained with the DNA dye Hoechst 33,422 and isotype-specific anti-human IgA1, IgA2, IgM and IgG antibodies to assess surface coating in one staining panel and with the lectins wheat germ

agglutinin (WGA, target: N-acetyl-glucosamine), peanut agglutinin (PNA, target: galactose), *Solanum tuberosum* agglutinin (STL, target: N-acetyl-glucosamine) and Concanavalin A (ConA, target: mannose) in a second panel to assess surface sugar expression (Figure 2a).²¹ The selection of lectins was based on commercial availability, diversity of their staining patterns to represent the variety of intestinal microbiota phenotypes, and recognition of sugars which have the potential to interact with lectins of the immune system, e.g. galactose and mannose.^{22–24} To analyze the multi-dimensional data of each donor, we applied a dimensionality-reducing clustering algorithm to group cells with similar phenotypic properties. For clustering, we applied a map (SOM) with 2025 clusters, which was trained on a larger data set (see methods), to the data of each staining panel. We then combined the clusters of both panels to a total of 4050 clusters to describe the microbiota phenotype of each sample. The frequency of cells per cluster was used to quantify the overall difference (Bray-Curtis dissimilarity) between samples and used for the determination of specific biosignatures (Figure 2). Taking all clusters into account, we observed a significant difference between CD and healthy controls, however, the high interindividual variability between all the samples resulted in a notable overlap between the cohorts (Supplementary Figure S4A, $R^2 = 0.05$, $p = 0.001$).

By statistical testing (Wilcoxon) and recursive feature elimination (RFE) (Figure 2b,c), 187 out of the 4050 clusters were selected that would best represent a CD-specific phenotypic microbiota biosignature in comparison to healthy controls. This selection resulted in an improved separation of the CD from the healthy samples (PCoA1: 34.46%, PCoA2: 11.86%, $R^2 = 0.18$, $p = 0.001$, Figure 3a). Many of the 187 clusters discriminating CD vs. healthy showed a decreased abundance in the CD samples (Supplementary Figure S5, S6) and were characterized by a higher DNA and SSC signal (Figure 3c). Clusters enriched in CD samples contained primarily microbial cells with a low DNA signal and increased coating with hIgA1/2 (Figure 3c, Supplementary Figure S5). The lectin staining was not obviously increased in CD patients (Figure 3c, Supplementary Figure S6).

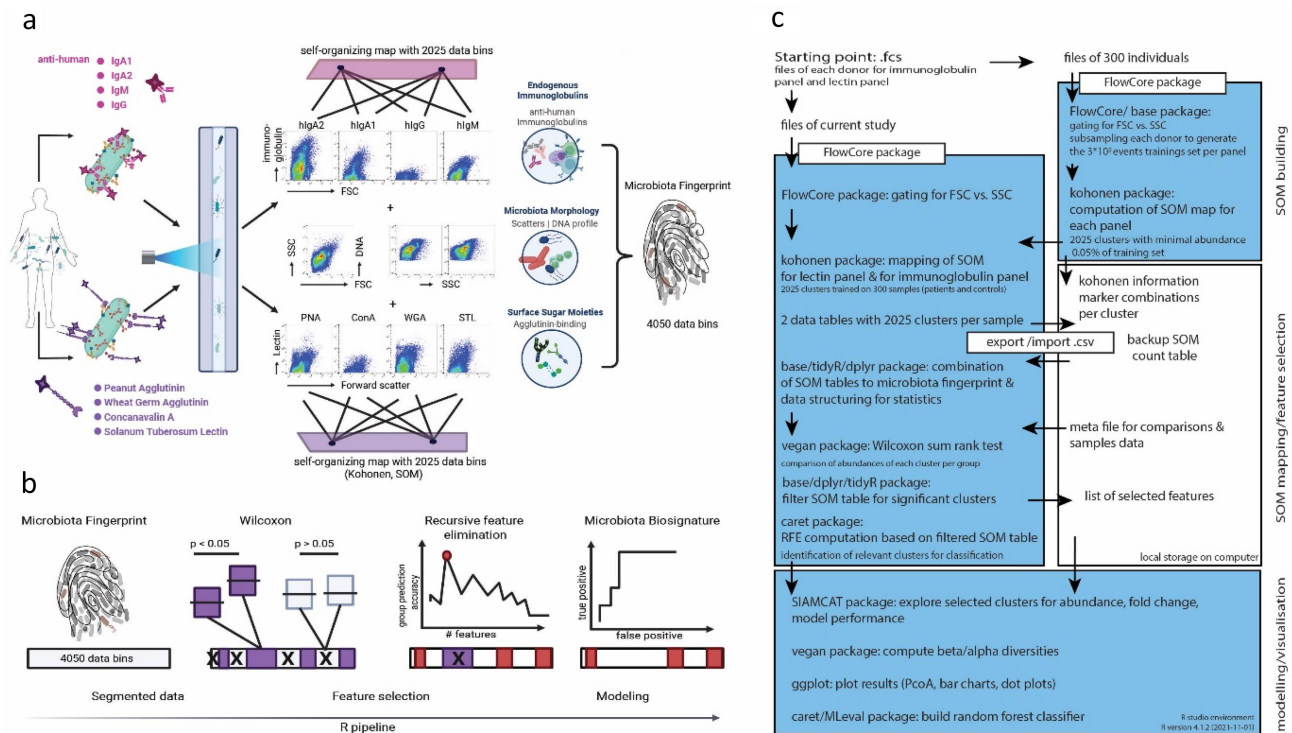


Figure 2. Microbiota phenotyping by flow cytometry and machine learning to identify specific biosignatures in human stool samples. (a) Human intestinal bacteria from stool samples were stained with monoclonal antibodies specific for the human immunoglobulins IgA1, IgA2, IgM and IgG and with the lectins peanut agglutinin (PNA), wheat germ agglutinin (WGA), *Solanum tuberosum* lectin (STL) and concanavalin a (ConA). Each staining panel also included the cell wall/membrane-permeable DNA dye hoechst 33,342. After data acquisition in a flow cytometer, the cells of each staining panel were clustered according to a previously defined self-organizing map (SOM) into 2025 clusters. The abundance of cells per cluster in the total of 4050 clusters represented the overall microbiota phenotype of a sample. (b) The clusters were filtered by Wilcoxon statistical evaluation and recursive feature elimination to select the significant and most relevant clusters defining the specific microbiota biosignature for random forest model-based classification of disease and comparison of samples by Bray-Curtis dissimilarity (β -diversity) projection. (c) Outline of all relevant data processing steps and used packages (R) for computing a microbiota phenotype.

We trained a random-forest model-based classifier based on these 187 clusters on CD cohort 1, which discriminated CD samples from healthy controls with an AUROC = 0.91 (specificity = 0.85 and sensitivity = 0.79) (Figure 3b). To validate the classification and thus the biosignature, we analyzed a second independent cohort of 19 CD patients, selected directly the 187 previously identified clusters and tested the performance of our classification model on this sample set. The 19 patients were each sampled twice, first prior (CD cohort 2 baseline, active CD) and second 6 weeks in anti-TNF therapy (CD cohort 2 therapy). Our classification model was also able to classify CD cohort 2 despite heterogeneity in disease activity with an AUROC = 0.92 for the baseline and 0.93 for 6w therapy samples, respectively (Figure 3b).

In parallel, all samples were analyzed by 16S rRNA gene amplicon sequencing to determine the

composition of the samples on taxonomic level. The identified taxa were also further filtered by Wilcoxon statistical test and RFE. In total, 82 taxa were selected for the best discrimination between CD patients and healthy controls (Supplementary Figure S2). The dissimilarity between CD and healthy controls based on these 82 taxa increased to $R^2 = 0.067$ ($p = 0.001$) (Figure 4a) compared to $R^2 = 0.047$ ($p = 0.001$) when all taxa were considered (Supplementary Figure S4B). The discrimination between healthy controls and CD samples based on these 82 taxa with a random-forest model performed similarly to the cytometric profiling in training (AUROC = 0.92, specificity = 0.86, sensitivity = 0.82) and in validation with the second CD cohort (AUROC = 0.96 for baseline and 0.91 6w therapy samples) (Figure 4b). The main discriminator was an increased abundance of *Ruminococcus gnavus* sp. in CD patients.

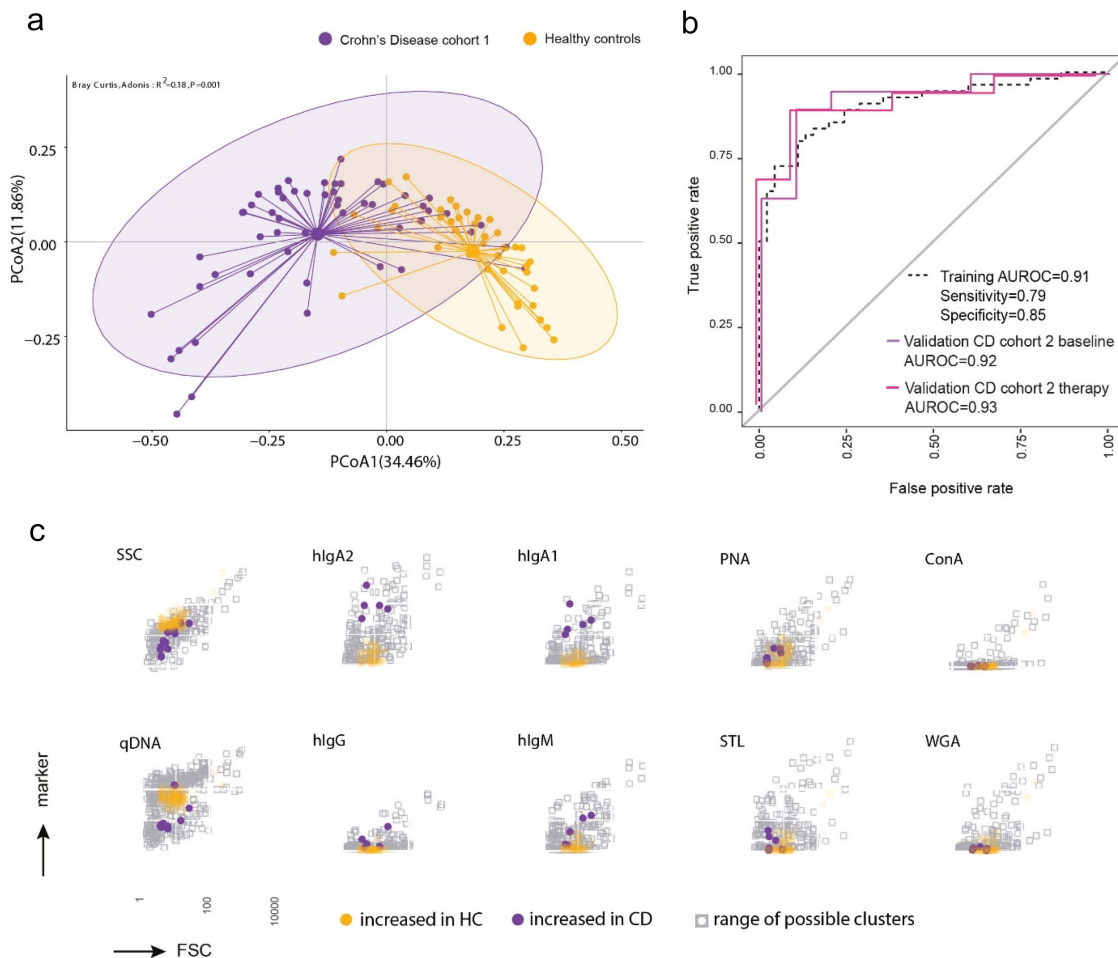


Figure 3. Microbiota phenotyping reveals a specific cytometric biosignatures of Crohn's disease. Samples of CD patients from cohort 1 ($n = 55$) and healthy controls ($n = 44$) were stained for host immunoglobulins and surface sugars. Following SOM clustering and cluster selection, 187 clusters were identified. (a) Principal coordinate projection representing the Bray-Curtis dissimilarity between samples of CD cohort 1 and healthy controls according to the 187 selected clusters. (b) The 187 clusters were used to train a random forest model classifying between CD patients and healthy controls. The model was validated with 19 patients from an independent CD cohort, sampled at two time points: before therapy (baseline) and after 6 weeks of anti-tnf therapy (therapy) and 10 new healthy donors. The performance of the binary classifier model is illustrated by the AUROC curves. (c) Projection of the selected clusters containing increased abundance of bacterial cells in either CD patients or healthy controls visualizing the mean fluorescence intensity in each stained parameter in relation to all clusters.

Thus, our results demonstrate that microbiota phenotyping was able to robustly classify CD patients similarly to 16S rRNA sequencing.

Identification of a microbiota signature reflecting therapy response

We next set out to test, whether the response of CD patients to the treatment with TNF blockers is reflected in the microbial signature. For this we analyzed the samples of CD cohort 2. The 19 CD patients of cohort 2 had received anti-TNF therapy (adalimumab) either for the first time ($n = 12$) or after an anti-

TNF therapy break of at least 4 weeks ($n = 7$, details in Table 1) and were profiled with mMFC and 16S rRNA gene sequencing at baseline prior to anti-TNF therapy and 6 weeks after therapy initiation. At 6 weeks, the CD patients were grouped into those having achieved remission or not according to the Harvey-Bradshaw index (HBI), which assesses, e.g. the general well-being of the patient, their stool frequency and abdominal mass.²⁵ 11 out of the 19 patients fulfilled remission criteria (HBI <5) while 8 patients did not (HBI ≥5). At baseline, the remission group exhibited lower, but not significantly different, disease score values for HBI (remission: mean HBI =

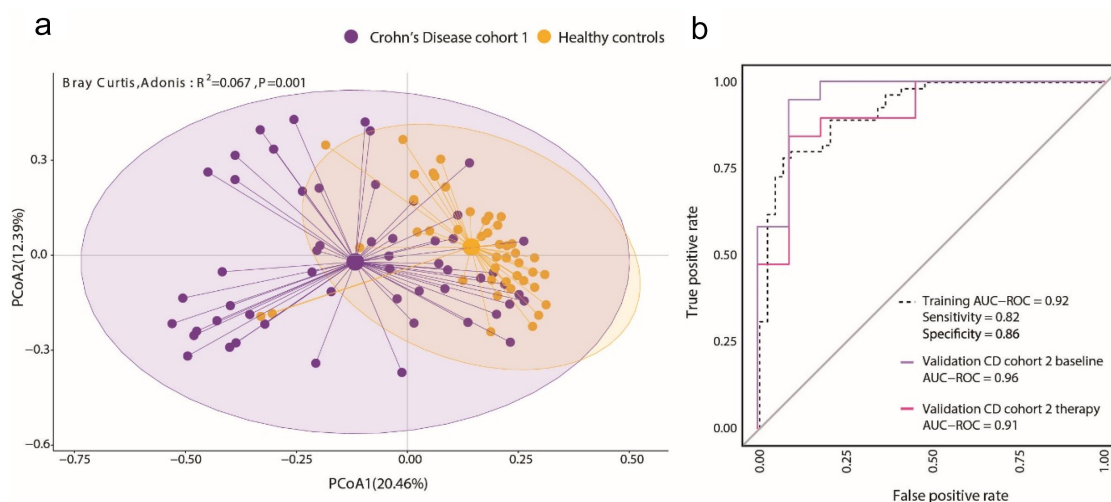


Figure 4. The microbiome composition is altered in Crohn's disease patients. Samples of CD patients from cohort 1 ($n = 55$) and healthy controls ($n = 44$) were analyzed by 16S rRNA V3/V4 amplicon sequencing. Identified genus level taxa also underwent selection as described resulting in 82 selected taxa. (a) Principal coordinate projection representing the Bray-Curtis dissimilarity between samples of CD cohort 1 and healthy controls according to the 82 selected taxa. (b) A random forest model classifier to distinguish between CD patients and healthy controls was trained with the 82 taxa. The model was validated with the 19 patients from CD cohort 2, sampled at two time points: before therapy (baseline) and after 6 weeks of anti-tnf therapy (therapy) and with the 10 new healthy donors. The performance of the taxonomy-based binary classifier model is illustrated by the AUROC curves.

Table 1. Donor information.

	CD cohort 1	CD cohort 2 (remission group)		CD cohort 2 (no-remission group)		controls
female/male	29/25	3/8		4/4		35/20
age	43 ± 13	36 ± 10		43 ± 10		48 ± 17
	in study	pre study	in study	pre study	in study	
Biologicals tota	34	3	11	4	8	0
anti-TNF	21	2 (1 × infliximab, 1 × adalimumab)*	11 (adalimumab)	3 (2 × infliximab, 1 × adalimumab)*	8 (adalimumab)	0
anti-IL12/23	10	4	0	2	0	0
anti-Integrin	3	0	0	0	0	0
small molecule	0	0	0	0	0	0
immune suppression	10	5	6	6	0	0
no current therapy	10	0	0	1	0	0

The table outlines the characteristics of patients included in the study.

*These patients had no anti-TNF therapy for at least 4 weeks before the start of the adalimumab study. The reasons for the anti-TNF therapy termination of these patients were therapy failure, severe side effects or personal reasons.

11.4 ± 4.4, no-remission: mean HBI = 12.4 ± 3.8) and Crohn's disease activity index (CDAI) (remission: mean CDAI = 305, no-remission: mean CDAI = 308). At the initiation of therapy both groups showed elevated CRP levels (remission: 34, no-remission: 26). The ultrasonography-based Limberg score²⁶ was similar in both groups (Table 2). In patients of the remission group the HBI score was reduced to 3.2 ± 3.2 ($p = 0.04$), while in patients attributed to the no-remission group the HBI score was reduced to 5 ± 2.9 ($p = 0.014$) indicating that the majority of patients responded to the therapy even when not achieving the remission criteria of the study.²⁷ When sampling

after 6 weeks of therapy, we detected significant differences both in the phenotypic biosignature and in the taxonomic composition of the intestinal microbiota between remission and no-remission patients (Figure 5a,e). 24 phenotypic clusters were identified to discriminate microbiota of remission patients and no-remission patients resulting in a clear separation of the two groups by Bray-Curtis dissimilarity ($R^2 = 0.396$, $p = 0.001$) (Figure 5a). Accordingly, the random-forest model classified the samples into remission and no-remission with an AUROC = 0.99 (specificity = 0.92, sensitivity = 0.99) (Figure 5b, solid line). The clusters dominant in no-remission patients

Table 2. Clinical parameters of CD cohort 2 at baseline and after 6 weeks TNF therapy.

	remission patients			no-remission patients		
	average value	std. deviation	Wilcoxon, paired p	average value	std. deviation	Wilcoxon, paired p
HBI baseline	11.4	4.4	0.04	12.4	3.8	0.014
HBI 6 weeks in therapy	3.2	3.2		5	2.9	
CDAI baseline	305	118.2	0.001	308	124	0.016
CDAI 6 weeks in therapy	105	95.3		136	88.2	
Limberg score* baseline	2	1	0.35	2	1	0.9
Limberg score ^{1*} 6 weeks in therapy	1.5	1		2	1	
CRP baseline	34	42.4	0.65	26	33.1	0.078
CRP 6 weeks in therapy	7	12.2		7	11.8	

*0: normal bowel wall thickness, preservation of wall layer stratification, and no signal on color Doppler; 1: wall thickening and absent color Doppler signal; 2: wall thickening with spot-like focal increases in vascularity; 3: wall thickening and diffuse stretches of increased mural vascularity; 4: wall thickening with increased color Doppler signal in the bowel wall with extension into the mesentery.

comprised bacteria displaying increased coating with the host immunoglobulins hIgA1/2 and hIgM compared to remission patients (Figure 5c, Supplementary Figure S7A). This signature associated with remission was not present at baseline, as the 24 clusters did not significantly differentiate between the groups ($p = 0.472$, Supplementary Figure S7B) at this sampling time point. On the taxonomic level, only 3 taxa were identified. *Coprococcus comes* was strongly associated with remission, with the limitation that it was only detectable in six out of 11 remission patients. *Bacteroides fragilis* was associated with no remission, which was detectable in only 5 out of the 8 no-remission patients (Figure 5d). Nevertheless, the Bray-Curtis dissimilarity projection based on the three selected taxa showed a clear separation between remission and no-remission ($R^2 = 0.303$, $p = 0.001$, Figure 5e). The random-forest model for the 16S rRNA sequencing data performed with an AUROC = 0.91 (specificity = 0.82, sensitivity = 0.75) (Figure 5b, dotted line). Combining phenotypic signature with 16S sequencing did not improve differentiation of the remission and no-remission group at week 6 (Supplementary Figure S8A), as in the combined RFE, only phenotypic clusters and no taxa were selected (data not shown). Finally, we tested whether the therapy-induced changes in the phenotypic signature of the microbiota of CD patients achieving remission resulted in a convergence toward the microbial biosignature of healthy controls. For this,

we applied the selected 187 clusters of the CD vs. healthy controls biosignature (Figure 3) to the samples of CD cohort 2. We then compared the change in dissimilarity of each patient from baseline to week 6 to the mean value of all healthy controls (Supplementary Figure S9A). Patients that achieved remission showed a reduced dissimilarity to healthy controls, albeit not reaching statistical significance, compared to the no remission patients. This trend is also represented in the smaller average distance of the remission samples to the healthy cohort in the principal coordinate projection (Supplementary Figure S9B). In summary, the success of anti-TNF therapy in CD patients is reflected both in a specific alteration in the phenotypic signature and to some extent also in the taxonomic composition of the intestinal microbiota.

Identification of a microbiota signature predicting response of CD patients to anti-TNF response

We tested whether the microbiota profiles could distinguish therapy-induced remission already at baseline, prior to advent of therapy. For this, the baseline samples were retrospectively categorized according to the ground truth and subjected to the corresponding machine-learning-based feature selection resulting in 18 clusters and significant Bray-Curtis dissimilarity ($R^2 = 0.15$, $p = 0.011$, Figure 6a). With the selected 18 clusters, random-

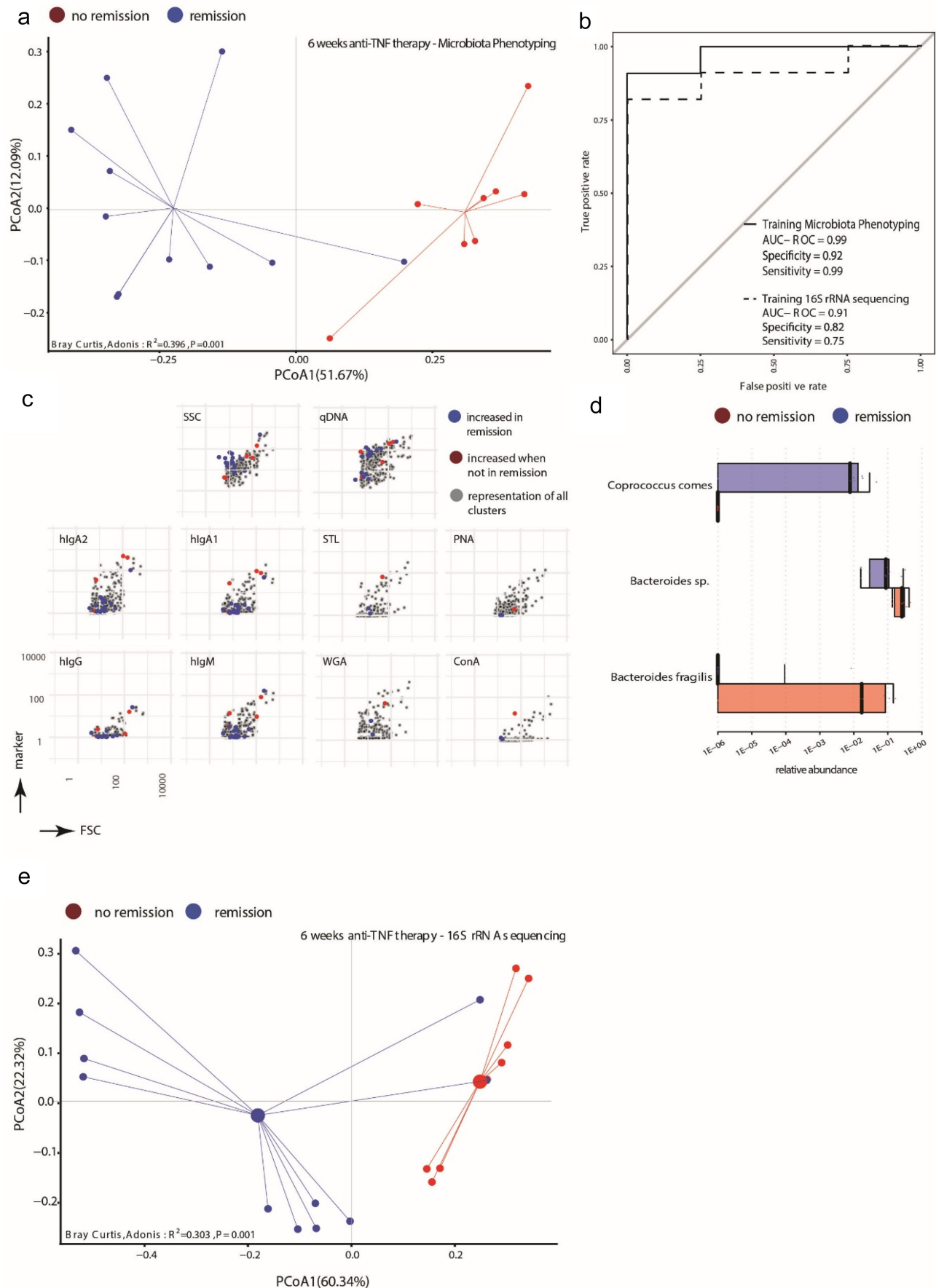


Figure 5. Phenotypic properties of the microbiota and several taxa correlate with therapy success for Crohn's disease patients after 6 weeks of anti-tnf therapy. 19 CD patients (CD cohort 2) were sampled before initiation of treatment with adalimumab and 6 weeks later, at which timepoint the patients were stratified into those achieving remission (Harvey-Bradshaw-index, HBI < 5, $n = 11$) and

forest classification achieved an AUROC = 0.93 (specificity = 1, sensitivity = 0.86) (Figure 6b, solid line). Among the 18 clusters, three were derived from the immunoglobulin panel and 15 from the lectin panel (Supplementary Figure S10A). The clusters predicting remission corresponded to bacterial cells displaying increased staining with all four lectins, while no-remission was predicted by clusters lacking the target sugar moieties of the used lectins (N-Acetyl glucosamine, lactose, mannose) (Figure 6c, Supplementary Figure S10A). At the 6 weeks' time point, the predictive biosignature was lost and no longer separated remission from no-remission patients (Supplementary Figure S10B, $R^2 = 0.083$, $p = 0.133$). From the 16S rRNA gene sequencing dataset, we identified five taxa in total, which resulted in a significant separation between remission and no-remission patients (Figure 6d,e, $R^2 = 0.308$, $p = 0.005$). *UCG-002 sp.*, *Lachnospiraceae NK4A136 group sp.*, *unclassified [Eubacterium] coprostanoligenes group* and *unclassified Clostridia UCG-014* were elevated in abundance in the group of patients that achieved remission during anti-TNF therapy 6 weeks later. In 8 out of the 11 patients achieving remission, any combination of two or more of the remission-associated taxa was detectable. The remission-associated taxa were not found in any of the no-remission patients, but also not in 3 of the remission patients. In the no-remission group, *Bacteroides fragilis* was more abundant and detectable in 7 out of the 8 individuals (Figure 6d). The classification based on the selected taxa performed slightly worse than phenotyping (AUROC = 0.8, specificity = 0.73, sensitivity = 1, Figure 6b, dotted line). Combining microbiota phenotyping and microbiome profiling (22 features: 2 taxa and 20 clusters) increased the Bray-Curtis dissimilarity

between remission and no-remission (Supplementary Figure S8B, $R^2 = 0.237$, $p = 0.003$). Our data thus indicate that the microbial signature has the potential to predict the response to anti-TNF therapy in CD patients.

Discussion

In this study, we describe the phenotypic characterization of single bacterial cells in complex intestinal communities by multi-parameter microbiota flow cytometry (mMFC) motivated by the observation that immunoglobulin-coating and lectin-binding to intestinal microbiota implied taxonomically dependent and independent information correlating to CD. The IgA-coated fractions of the intestinal microbiota were taxonomically significantly different between CD patients and healthy controls, and were representative of the overall microbiome composition of both groups (Figure 1a,d). The lectin-stained fractions, both WGA to PNA, show a decreasingly distinct profile between CD and healthy controls (Figure 1b, c) and enriched some of the CD-related potentially colitogenic taxa but also taxa that were not visibly different between the cohorts in the total microbiome (Supp. 3A,B). Overall we hypothesize that in disease, in this case CD, the intestinal microbiota adapts to the altered condition of the host, which is reflected by the altered patterns of immunoglobulin coating and surface sugar expression.

To explore this, we simultaneously analyzed quantitative DNA staining, light scattering, host-immunoglobulin coating, and expression of certain sugar residues on the surface of the bacteria. Using dimensionality reduction, machine learning for feature selection and random forest modeling for automated sample classification, we demonstrate the feasibility of using single-cell-based microbiota

those not achieving remission criteria ($HBI \geq 5$, $n = 8$) (a) Principal coordinate projection representing the Bray-Curtis dissimilarity between samples of remission (blue) and no remission (red) according to 24 selected cluster 6 weeks after initiation of adalimumab treatment. (b) A random forest model classifier to distinguish between CD patients in remission and patients not in remission after 6 weeks of adalimumab treatment was trained with the 24 selected phenotypic clusters (solid line) or with 3 taxa (dotted line) identified by 16S rRNA amplicon sequencing. The performance of the respective classifier model is illustrated by the AUROC curves. (c) Projection of the selected clusters containing increased abundance of bacterial cells in CD patients in remission or not in remission visualizing the mean fluorescence intensity in each stained parameter in relation to all clusters. (d) Abundance of the bacterial taxa significantly differentially abundant between patients in remission or not in remission following 6 weeks of adalimumab treatment. Indicated are the median abundance, 95% confidence interval and coefficient of variation. (e) Principal coordinate projection representing the Bray-Curtis dissimilarity between samples of remission (blue) and no remission (red) according to the 3 taxa 6 weeks after initiation of adalimumab treatment.

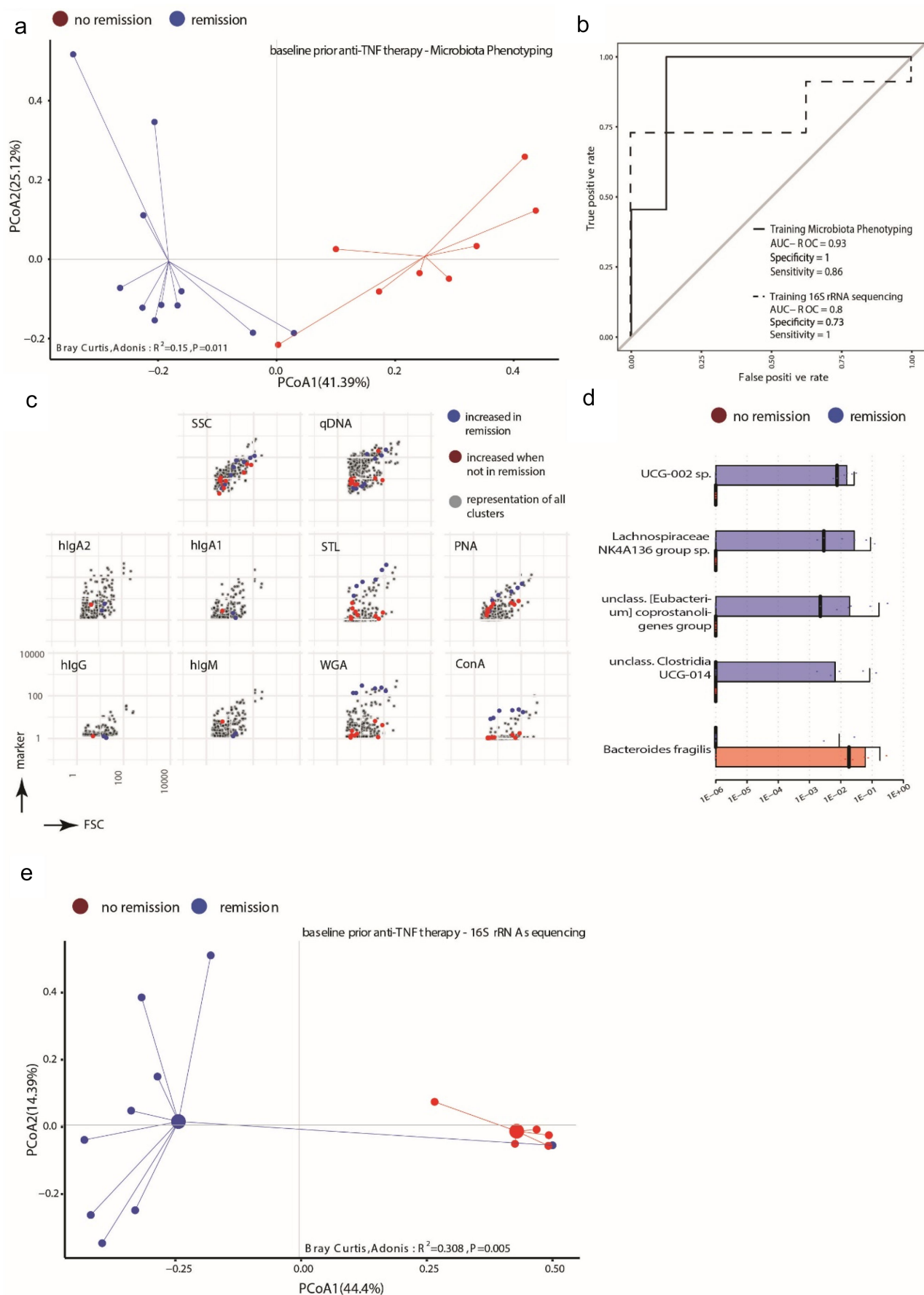


Figure 6. The baseline microbial signature can predict the achievement of remission induced by anti-tnf therapy in Crohn's disease patients. (a) Principal coordinate projection representing the Bray-Curtis dissimilarity between patients at baseline before initiation of adalimumab treatment stratified into those reaching remission (blue) and those not-fulfilling remission criteria (red) according to 18

characterization to identify disease-specific biosignatures for patient stratification and monitoring. The mMFC-based microbiota phenotyping reliably classified CD patients and healthy donors. Thus, in addition to the microbiome alterations on the taxonomic level described for CD, we here show that surface characteristics of bacteria are presumably impacted by the inflammatory environment in CD and distinct from healthy controls.

Our results indicate that mMFC could be an effective method alternative to or complementing the wide-spread sequencing approaches, for microbiome-based diagnostics and the analyses of disease- and therapy-specific alterations.¹³ The markers used in this study were suitable to differentiate CD patients from healthy individuals and to stratify CD patients. The inclusion of other markers may help or be necessary for the discrimination of other diseases. Technically, the mMFC approach maintains the flexibility to easily adapt the staining panel accordingly when required. We suspect that the phenotypic features, we assess by mMFC in high granularity, reflect adaptations of the immune system, the host-microbiome interaction and the bacteria and are less sensitive to taxonomic variations between individuals. Although both CD patients and healthy individuals have IgA-coated bacteria in their microbiota, the highly resolved phenotypic differentiation of the bacterial phenotype may allow to discriminate the homeostatic state from the disease-related state of IgA coating,²⁸ confirming that immunoglobulin coating of microbes in IBD identifies colitogenic, potentially pathogenic bacteria representative for the taxonomic dysbiosis.^{7,9,29} In a cohort of 19 patients with CD receiving anti-TNF therapy, we observed that remission of CD patients induced by anti-TNF therapy was primarily reflected by changes in bacteria coated with host immunoglobulins, perhaps indicating an alteration

of the mucosal humoral immune response by therapy. This is in line with previous findings that during anti-TNF therapy the expression and secretion of mucosal antibodies is changing.^{29,30} Taxonomically, we observed that *Coprococcus comes* positively and *Bacteroides* species negatively correlated with anti-TNF-induced remission. For *Coprococcus* this has been observed before.³¹

Changes in functionality rather than taxonomy have been shown to better characterize the dysbiosis in IBD.³² Correspondingly, we have observed that altered lectin-binding patterns do not correlate to altered taxonomic composition of the overall microbiome but have relevant impact on CD-specific microbiota phenotype signatures, suggesting that altered lectin-binding indicates cellular adaptations of certain bacteria rather than compositional changes of the community. Accordingly, we have observed that the staining with lectins is different in axenic cultures depending on growth state, pH and medium composition (data not shown). For the initial prediction of the remission potential of individual patients to anti-TNF therapy prior to therapy, the expression of sugar moieties of individual bacteria was very predictive, while surface coating with immunoglobulins was not relevant for the prediction of therapy response. Previous findings that the metabolic capacity of a microbiome and metabolite exchange between bacteria correlates with a positive therapy response³¹ support our assumption that the single-cell resolved analysis of surface sugars mirrors metabolic activity and intercellular crosstalk. The presence of a phenotypic signature that predicted anti-TNF-induced remission before start of the therapy highlights the potential of mMFC for patient stratification. However, further studies will be needed to clarify the molecular link between surface sugars and metabolism. The taxonomic profiling identified

cluster selected from the baseline samples. (b) A random forest model classifier to stratify CD patients at baseline into those reaching remission and patients not reaching remission after 6 weeks of adalimumab treatment was trained with 18 selected cluster (solid line) or with 5 taxa (dotted line) identified by 16S rRNA amplicon sequencing. The performance of the respective classifier model is illustrated by the AUROC curves. (c) Projection of the selected clusters containing increased abundance of bacterial cells in CD patients reaching remission or not of the baseline analysis visualizing the mean fluorescence intensity in each stained parameter in relation to all clusters. (d) Abundance of the bacterial taxa significantly differentially abundant between patients reaching remission or not before initiation of adalimumab treatment. Indicated are the median abundance, 95% confidence interval and coefficient of variation. (e) Principal coordinate projection representing the Bray-Curtis dissimilarity between samples of remission (blue) and no remission (red) according to the 5 taxa at baseline before initiation of adalimumab treatment.

UCG-002 sp., *Lachnospiraceae NK4A136 group sp.*, *unclass. Clostridia UCG-014* and *unclass. Eubacterium coprostanoligenes group* which, when detected at baseline, predicted a positive therapy outcome, nevertheless the classification performance by a random forest model was limited. In such a scenario, we showed, that mMFC and 16S rRNA sequencing could complement each other to increase predictive power.

While at this point we cannot draw conclusions about any causal relationship between microbiota phenotype and disease pathogenesis *per se*, the flow cytometric approach offers the tools to do so, by giving access to the bacteria with a particular phenotype directly by cell sorting,³³ although adaptations in the technology need to be considered to gain access to strictly anaerobic bacteria.³⁴

In conclusion, we here demonstrate that single-cell phenotyping of bacteria by multi-parametric microbiota flow cytometry (mMFC) is a potent tool for the investigation and characterization of complex microbial communities, such as the intestinal microbiota, for the identification of disease-specific bacterial signatures and generates robust results compared to 16S rRNA sequencing. Additionally, mMFC captures parameters e.g. sugar-moiety composition that were previously neglected in microbiome profiling but allowed for prediction of therapy outcome. Considering the widespread use of flow cytometry in diagnostics, mMFC presents a viable alternative and supplementation to “conventional” microbiota profiling with the potential to be applied for point-of-care diagnostics and therapy monitoring, but also to investigate the role of defined bacteria in disease pathogenesis.

Material and methods

Stool samples

Stool samples were provided by all donors under approval of the local ethics committee of the Charité Berlin (approval reference: EA4/014/20; EA4/247/20) and in accordance to the Helsinki II Declaration. The samples were taken with stool sampling tubes (Sterilin®, VWR Cat. No. 215–0327) and transferred to 4°C instantly upon arrival.

Patient and public involvement

During patient recruitment, the study was openly and transparently communicated by informed patient consent form. Processed study data was made available at all stages of the study to the participants. Healthy controls completed a questionnaire assessing potential confounders such as age, gender, smoking status, and nutritional habits. The control group did not report any chronic intestinal inflammation. The patients’ disease state and entity were indicated by the examining clinicians. The sample collection logistics was designed to enable easy participation in the study, e. g. from home. Members of the research team are actively involved in science communication events to explain the methods and goals of this study to the broad public and to receive feedback regarding patients’ needs and concerns.

Cohorts

For this study two independent CD cohorts were recruited (characteristics summary in Table 1). Cohort 1 ($n = 57$) was recruited from the IBD ambulance ward at Charité-Universitätsmedizin Berlin (CCM) without selection of the patients in regards to inflammation severity, disease duration or therapy. This cohort laid the groundwork to characterize the microbiota of CD patients and was applied in the comparisons to healthy controls ($n = 44$, age- and sex-matched, Table 1) to define the CD biosignature for each of the comparisons (training sets). Cohort 2 consisted of CD patients ($n = 19$) who participated in an independent study that investigated the predictability of TNF-therapy efficacy. The patients of cohort 2 were selected for active disease (HBI >5) and samples were collected prior to anti-TNF therapy (CD cohort 2 baseline) as well as after 6 weeks of therapy (CD cohort 2 therapy). Cohort 2 was applied to validate the biosignatures of CD when compared to healthy controls ($n = 10$) which were not originally used for model training, and to define new biosignatures to predict and evaluate the success of anti-TNF-therapy in these patients.

Stool sample processing

Stool samples were kept at 4°C for a maximum of 96 h before processing. When longer storage was

required, samples were frozen directly at -20°C or -80°C . Each sample was diluted in autoclaved and $0.2\ \mu\text{m}$ sterile-filtered PBS (in-house, Steritop[®] Millipore Express[®]PLUS $0.22\ \mu\text{m}$, Cat. No: 2GPT05RE) to a concentration of $100\ \text{mg/ml}$ and homogenized by vortexing. The suspension was sequentially filtered through $70\ \mu\text{m}$ (Falcon, Cat. No. 352350) and $30\ \mu\text{m}$ filters (CellTrics[®], Sysmex, Cat. No. 04-0042-2316). $10\ \mu\text{l}$ were stored directly at -20°C for 16S rRNA gene sequencing. For each sample stocks at $\text{OD} = 0.4$ determined at $690\ \text{nm}$ with a VIS-spectrometer (Multiskan[™] FC, Thermo Scientific[™]) were prepared for long-term storage by re-suspending the respective volume from the microbiota suspension in $1\ \text{ml}$ 40% glycerol/LB freezing medium and immediate transfer to -80°C .

Staining microbiota for multi-parameter microbiota flow cytometry (mMFC)²¹

Frozen microbiota stocks (OD 0.4) were topped up with $1\ \text{mL}$ of autoclaved and sterile-filtered PBS and centrifuged at $13,000 \times g$ for $10\ \text{min}$, 4°C . The pellet was incubated in $500\ \mu\text{l}$ blocking solution containing $20\ \mu\text{g/ml}$ mIgG1 (clone: IS5-21F5, Miltenyi Biotech Cat. No.: 130-106-545) and $10\ \mu\text{g/ml}$ mIgG2a (clone: S43.10, Miltenyi Biotech Cat. No.: 130-106-546) in PBS for $5\ \text{min}$ at RT. The suspension was topped with $1.5\ \text{ml}$ PBS and subjected to another centrifugation step ($13,000 \times g$, $10\ \text{min}$, 4°C). The pellets were re-suspended in PBS containing 0.2% BSA (v/w) and $25\ \mu\text{g}/\mu\text{l}$ DNase (Sigma Aldrich Cat. No. 10104159001), which was also used as staining buffer. Cell density was adjusted to $0.02\text{--}0.04\ \text{OD}_{690}/\text{ml}$. $100\ \mu\text{L}$ of the cell suspension was used for one test, e.g. stained with the immunoglobulin panel. The antibodies used were anti-human IgM-Brilliant Violet 650 (clone: MHM-88, Biolegend[®] Cat. No. 314526), anti-human IgG-PE/Dazzle[™] 594 (clone: HP6017, Biolegend[®] Cat. No. 409324), anti-human IgA1-Alexa Fluor 647 (clone: B3506B4, Southern Biotech Cat. No. 9130-31), anti-human IgA2-Alexa Fluor 488 (clone: A9604D2, Southern Biotech Cat. No. 9140-30). The lectins used for staining were $0.5\ \mu\text{g}/\text{test}$ Peanut Agglutinin-CF[®]488 (PNA, Biotium Cat. No.29060), $0.5\ \mu\text{g}/\text{test}$ Concanavalin A-CF[®]680 (Con A, Biotium Cat. No. 29020-1) and $0.25\ \mu\text{g}/\text{test}$ Wheat Germ Agglutinin-CF[®]555 (WGA, Biotium Cat.

No.29076-1); $0.5\ \mu\text{g}/\text{test}$ of biotinylated Solanum Tuberosum Agglutinin (STL, Vector Laboratories/Biozol Cat. No. B-1165) were shortly pre-incubated with $2\ \mu\text{L}$ (1:50, v/v) anti-Biotin-PerCP antibody (clone: Bio3-18E7, Miltenyi Biotech Cat. No. 130-133-293) before adding to the residual reagents. The tests were incubated for $30\ \text{min}$ at 4°C and subsequently topped up with $1\ \text{ml}$ $5\ \mu\text{M}$ Hoechst solution (Hoechst 33,342, Thermo Fisher Scientific Cat. No. 62249) for another $30\ \text{min}$ at 4°C . After incubation the tests were washed with $900\ \mu\text{l}$ PBS/BSA at $13,000 \times g$ and re-suspended in fresh PBS/BSA for acquisition.

Microbiota flow cytometry

BD Influx[®] cell sorter was used for all cytometric measurements. The sheath buffer (PBS) for the instrument was autoclaved and sterile filtered (Steritop[®] Millipore Express[®]PLUS $0.22\ \mu\text{m}$, Cat. No: 2GPT05RE) before each fluidics start up. The quality and reproducibility of each acquisition was controlled by the alignment of lasers, laser delays and laser intensities by Sphero[™] Rainbow Particles (BD Biosciences Cat. No. 559123) and control of scatter properties by Megamix-Plus FSC beads (BioCytex Cat. No7802). Samples were acquired with an event rate below $15,000\ \text{events/second}$. For each sample, 3×10^5 Hoechst 33,342-stained events (mean fluorescence intensity > 10) were recorded. We controlled the staining procedure throughout the study by including a standardized microbiota sample (anchor sample) comprising a pool of different donors.

Graphical and statistical data analysis

Statistical analyses were implemented through R (v. 4.0.3 or later versions, Figure 2c), unless stated otherwise (supplementary Methods). Computation of β -diversity with Bray-Curtis dissimilarity was computed using `vegdist(data, method="bray")` function from `vegan` package.³⁵ The Bray-Curtis dissimilarity is a statistical metric to quantify the difference in composition between two cohorts. In our case it is the abundance of cells in each cluster or the presence or absence of bacterial taxa and their abundance. The Bray-Curtis dissimilarity is defined by a number between 0 and 1, 0 indicating that two samples/cohorts are

completely similar and 1 indicating that two samples/cohorts do not share anything. The PCoA was computed by the R base function `cmdscale()` on the respective distance matrix followed by `Adonis` test `adonis()` from `vegan` package to evaluate the variance within groups. Graphical representation of the dissimilarity of all samples by Principal Coordinates Analysis (PCoA) was plotted using `ggplot2` package. Correlation analyses data was evaluated by Pearson's r e. g. in `ggscatter(...)`, `cor.method="pearson"`) and the statistical comparisons of paired data points by `stat_compare_means()` using `ggpubr` package.

Clustering of flow cytometric data

The outline of the computational steps are shown in [Figure 2c](#). Raw FCS-files from flow cytometry were imported to R without any transformation using `FlowCore's read.flowSet()`³⁶ and was gated in the environment of the `flowWorkspace`³⁷ environment to reduce instrument noise by including only events > 1 for forward (FSC) and side scatter (SSC). A self-organizing map (SOM, `kohonen` package)^{38,39} was previously trained on a data set comprising 3×10^5 cells subsampled and concatenated from approx. 300 samples (comprising a mixture of patients with different chronic inflammatory diseases and healthy controls) stained for the immunoglobulin coating and the surface sugar expression to capture a widespread diversity of mMFC patterns by the `som()` function of the `kohonen` package for a 5×5 hexagonal map describing the data based on gaussian neighborhood. For SOM training on the 3×10^5 cells subsample clusters should contain at least 0.05% (i.e. 150 cells) of the cells which yielded 2025 cluster for each staining panel. To characterize the phenotypic diversity of the samples used in this study, we then mapped the pre-determined cluster to each sample by `kohonen:map()` after downsampling to 3×10^4 cells per individual in favor of computing power and time, as we found 3×10^4 cell to represent a sample sufficiently. Each sample is now described by an individual abundance of cells in each cluster. To assess the information of immunoglobulin coating and surface sugar moieties presence simultaneously we combine both cluster sets. Thus the same bacteria are described by two different

SOMs and the number of clusters adds up to 4050 which we termed the microbiota phenotype.

Feature selection & modelling

The outline of the computational steps is shown in [Figure 2c](#). Cellular frequencies within each mMFC cluster and relative abundances of bacterial genera, respectively, were used as inputs for random forest machine learning models for sample classification. Briefly, all data were pre-filtered to exclude non-significant features with $p > 0.05$, Wilcoxon rank-sum test (`vegan` package,³⁵ removing all mMFC clusters and 16S rRNA sequencing-derived taxonomic units not contributing to the discrimination between the cohorts. In a second filtering step, Recursive Feature Elimination (RFE) with 10-fold cross-validation was applied to remove weak features for classification, with `rfeControl()` function in `caret` package.⁴⁰ The importance of the selected features was obtained within the RFE according to the consensus ranking through the 10-fold cross-validation. The resulting features were used to train the random forest model involving 10 times repeated 10-fold cross-validation with function of `train()` and `trainControl()`, to mitigate overfitting (`caret` and `MLeval` package).^{40,41} 10 randomly selected samples each from the healthy controls and all samples from CD cohort 2 originally excluded from all steps of feature selection were used to evaluate the models' predictive performance in one modeling attempt. All procedures related to model construction were performed in `caret` package,⁴⁰ and evaluation of predictive ability (performance metrics, AUROC, sensitivity, specificity, and confusion matrix at default threshold 0.5) was implemented using `MLeval` package.⁴¹

Fluorescence-activated cell sorting of bacteria

For cell sorting, the BD Influx cell sorter was used. The event rate was adjusted to below 10^4 events/sec. The drop delay and deflection accuracy was set with `Accudrop` (BD Biosciences, Cat No. 345249). The sorter was operated with 1-drop pure settings. A minimum of 10^5 cells per sample and phenotype were sorted into buffer-pre-coated 2 ml reaction tubes. Samples were stored on ice until the complete sort was finished. The fractions were then spun down for 15 min at $17,000 \times g$, the

supernatant was removed and the cells left in a volume of approx. 50 µl were stored at -20°C until further processing.

16S rRNA sequencing (illumina MiSeq platform)

For 16S rRNA gene sequencing, we amplified the V3/V4 region of the 16S rRNA gene (for: TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGCCTACGGGnGGCWGCAG, rev: GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACTACHVGGGTATCTAATCC⁴²; TIB MOLBIOL Syntheselabor GmbH) directly from a microbiome sample with a prolonged initial heating step of 5 min. After the amplicon PCR the genomic DNA was removed by AmPure XP Beads (Beckman Coulter Life Science Cat. No. A63881) with a 1:1.25 ratio of sample to beads (v/v). The amplicons were checked for their size and purity on a 1.5% agarose gel, and if suitable, subjected to the index PCR using the Nextera XT Index Kit v2 Set C (Illumina, FC-131-2003). After Index-PCR, the samples were cleaned again with AmPure XP Beads (Beckman Coulter Life Science Cat. No. A63881) in a 1:0.8 ratio of sample to beads (v/v). Samples were analyzed by capillary gel electrophoresis (Agilent Fragment Analyser 5200) for correct size and purity with the NGS standard sensitivity fragment analysis kit (Agilent Cat. No. DF-473). Of all suitable samples a pool of 2 nM was generated and loaded to the Illumina MiSeq 2500 system.

Full-length 16S rRNA sequencing (PacBio platform)

The PacBio 16S rRNA sequencing protocol was conducted as described by PacBio's instructions for the SMRTbell® Express Template PrepKit 2.0 (Pacific Biosciences of California, Inc; Cat. No. 101-685-400). Briefly, bacterial full-length 16S rRNA gene was amplified with barcoded primers (fwd.: AGRGTTYGATYMTGGCTCAG; rev.: RGYTACCTTGTTACGACTTT). 5 µL from each sorted sample were used as input DNA. After successful amplification, 500 ng of each barcoded amplicon were pooled for library construction. Before library construction the quality of the pooled input DNA was control by capillary gel electrophoresis (Agilent Fragment Analyser 5200), followed by cleanup from non-PCR product by AmPure XP Beads (Beckman Coulter Life Science Cat. No.

A63881). Afterwards, DNA repair and A-tailing as well as adapter ligation with subsequent bead purification were performed according to the PacBio protocol. The prepared library was subjected to full-length 16S rRNA sequencing by the PacBio Sequel I system using SMRT Link Version 8 including sequencing primer annealing and polymerase binding.

Sequence alignment illumina MiSeq data

Paired-end reads generated by Illumina MiSeq 16S rDNA sequencing were filtered and trimmed using Trimmomatic (Version 0.39).⁴³ 7 leading bases with qualities below 35 were trimmed and reads shorter than 180 bases were filtered out. Using the DADA2 (Version 1.22.0) software package,⁴⁴ forward and reverse reads were truncated at 260 and 210 bases respectively and filtered with a minimum quality score of 12 and a maximum of 0 ambiguous nucleotides. Amplicon sequence variants (ASVs) were identified using the default settings of the DADA2 algorithm and ASVs were classified using the Silva 138.1 prokaryotic SSU taxonomic training data formatted for DADA2.⁴⁵ After alignment of the sequences with DECIPHER (Version 2.24.0),⁴⁶ a phylogenetic tree was computed using FastTree (Version 2.1.11).⁴⁷ For analysis the ASVs were matched with the respective phylogenetic information, the data was processed to genus level and normalized prior to any further calculations.

Sequence alignment PacBio SMRT link data

After sequencing, samples were demultiplexed and circular consensus sequence (CCS) reads were constructed using PacBio's software followed by filtering with a 99% base call precision (quality score of 20). Further processing was done using the DADA2 pipeline (version 1.24.0) functions. First, sequencing primers were removed and reads without primers were discarded. Unique sequences were then identified and dereplicated, i.e. collapsed into a set of clustered ASVs. After learning and removing errors from amplicon reads, the ASV counts

observed in each sample were recorded in a summary table. Chimeras were removed from the sequences and taxonomy assignments to ASVs were achieved by implementing the Ribosomal Database Project naive Bayesian classifier algorithm as described by.⁴⁸ Species-level annotation were added to taxonomic assignments by exact matching against the Silva 138.1 reference FASTA database.⁴⁹ For analysis the ASVs were matched with the respective phylogenetic information, the data was processed to genus level and normalized prior to any further calculations.

Abbreviations

CD	Crohn's disease
mMFC	multi-parameter microbiota flow cytometry
PCoA	principal coordinate analysis
SOM	self-organizing map
AUROC	area under the receiver operator curve
RFE	recursive feature elimination
HBI	Harvey-Bradshaw index

Acknowledgments

The execution of this study relies on the participation of clinical partners from Charité-Universitätsmedizin Berlin; physicians and study nurses who provided access to sample material and recruited patients to the study. Further this study benefits from the technical/instrumental infrastructure of the German Rheumatism Research Centre Berlin and the implementation of a sample coordination/clinical data handling office. All major partners of the study are listed as co-authors.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was supported by the Rolf M. Schwiete Foundation grant (to H.D.C.), the EFRE-Project grant 1.6./01: „BacFlow - Analyse bakterieller Gemeinschaften durch Mikrobiota-Zytometrie” (to H.D.C.), the DFG Project-ID 375876048 TRR241 B03 (to H.D.C. and A.R.), TRR241 B01 (to B.S. and

C.W.), Z02 (to B.S.), CRC1449 B04, Z02 (to B.S.), CRC1340 project number 372486779 B06 (to B. S.), 491069896 (to H.D. C.) and the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement no. 777357 (RTCure) and no. 831434 (3TR), The state of Berlin and the “European Regional Development Fund”: ERDF 2014-2020, EFRE 1.8/11 (to M.-F. M., Deutsches Rheuma-Forschungszentrum); The Leibniz Association (Leibniz Collaborative Excellence, TargArt) (to T. K. and M.-F. M.), Deutsche Morbus Crohn/Colitis ulcerosa Vereinigung e. V. (DCCV e.V., to B. J. and M. T.).

ORCID

Lisa Budzinski  <http://orcid.org/0000-0003-4120-6032>
 Marcell T. Tordai  <http://orcid.org/0009-0006-4905-5637>
 Britta Siegmund  <http://orcid.org/0000-0002-0055-958X>

Author's contributions

L. B., H.D. C. designed and carried out the study; L. B., T. S., L. L., R. M., J. C., R. K., T.M. processed clinical samples, developed the method protocol and conducted the flow cytometric measurements; L. B., G-U. K., R. R., A. S. developed the data processing pipelines and processed the data; G. A. H., K. L., M.-F. M. supported the 16S rRNA sequencing, Aa. S., A. A. mapped sequencing data, J. B., B. B., K. N., S. B., A. E. B., M. R., R. B., T. K., T. A. provided clinical samples, clinical data and relevant ethics for their cohorts; A. R., J. L. K. discussed the study design and supported the project development; L. B., C. W., B. S., U. H., H.D. C. wrote the ethics, L.B., H.D. C., G. K., wrote the manuscript which all coauthors proof read.

Availability of data and materials

10.5281/zenodo.13375351

Consent for publication

All personal data of included donors is pseudonymised as outlined in the ethics application. All contributing authors are informed about the publication and consent on the publication of the data.

Ethics approval and consent to participate

All donors included in this study were informed and about the study concept, aim and the experiments conducted with their biological samples, they approved their participation by

signature. The study is approved by the local ethics committee of the Charité Berlin (approval reference: EA4/014/20; EA2/113/20).

References

- Clemente JC, Ursell LK, Parfrey LW, Knight R. The impact of the gut microbiota on human health: an integrative view. *Cell*. 2012;148(6):1258–1270. doi:10.1016/j.cell.2012.01.035.
- Duvallet C, Gibbons SM, Gurry T, Irazary RA, Alm EJ. Meta-analysis of gut microbiome studies identifies disease-specific and shared responses. *Nat Commun*. 2017;8(1):1784. doi:10.1038/s41467-017-01973-8. [published Online First: 20171205].
- Durack J, Lynch SV. The gut microbiome: relationships with disease and opportunities for therapy. *J Exp Med*. 2019;216(1):20–40. doi:10.1084/jem.20180448. [published Online First: 20181015].
- Knox NC, Forbes JD, Van Domselaar G, Bernstein CN. The gut microbiome as a target for IBD treatment: are we there yet? *Curr Treat Options Gastroenterol*. 2019;17(1):115–126. doi:10.1007/s11938-019-00221-w.
- Human Microbiome Project C. Structure, function and diversity of the healthy human microbiome. *Nature*. 2012;486(7402):207–214. doi:10.1038/nature11234. [published Online First: 20120613].
- Lloyd-Price J, Arze C, Ananthakrishnan AN, Schirmer M, Avila-Pacheco J, Poon TW, Andrews E, Ajami NJ, Bonham KS, Brislawn CJ, et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature*. 2019;569(7758):655–662. doi:10.1038/s41586-019-1237-9. [published Online First: 20190529].
- Palm NW, de Zoete MR, Cullen TW, Barry N, Stefanowski J, Hao L, Degnan P, Hu J, Peter I, Zhang W, et al. Immunoglobulin a coating identifies colitogenic bacteria in inflammatory bowel disease. *Cell*. 2014;158(5):1000–1010. doi:10.1016/j.cell.2014.08.006.
- Frehn L, Jansen A, Bennek E, Mandic AD, Temizel I, Tischendorf S, Verdier J, Tacke F, Streetz K, Trautwein C, et al. Distinct patterns of IgG and IgA against food and microbial antigens in serum and feces of patients with inflammatory bowel diseases. *PLOS ONE*. 2014;9(9):e106750. doi:10.1371/journal.pone.0106750. [published Online First: 20140912].
- Masu Y, Kanazawa Y, Kakuta Y, Shimoyama Y, Onodera M, Naito T, Moroi R, Kuroha M, Kimura T, Shiga H, et al. Immunoglobulin subtype-coated bacteria are correlated with the disease activity of inflammatory bowel disease. *Sci Rep*. 2021;11(1):16672. doi:10.1038/s41598-021-96289-5. [published Online First: 20210817].
- Choudari CP, O'Mahony S, Brydon G, Mwantembe O, Ferguson A. Gut lavage fluid protein concentrations: objective measures of disease activity in inflammatory bowel disease. *Gastroenterol*. 1993;104(4):1064–1071. doi:10.1016/0016-5085(93)90275-h.
- Tytgat HLP, de Vos WM. Sugar coating the envelope: glycoconjugates for microbe–host crosstalk. *Trends Microbiol*. 2016;24(11):853–861. doi:10.1016/j.tim.2016.06.004. [published Online First: 20160630].
- Matute JD, Duan J, Flak MB, Griebel P, Tascon-Arcila JA, Doms S, Hanley T, Antanaviciute A, Gundrum J, Mark Welch JL, et al. Intelectin-1 binds and alters the localization of the mucus barrier–modifying bacterium *akkermansia muciniphila*. *J Exp Med*. 2023;220(1). doi:10.1084/jem.20211938. [published Online First: 20221122].
- Budzinski L, von Goetze V, Chang H-D. Single-cell phenotyping of bacteria combined with deep sequencing for improved contextualization of microbiome analyses. *Eur J Immunol*. 2023; e2250337. doi:10.1002/eji.202250337. [published Online First: 20231020].
- Koch C, Gunther S, Desta AF, Hübschmann T, Müller S. Cytometric fingerprinting for analyzing microbial intracommunity structure variation and identifying subcommunity function. *Nat Protoc*. 2013;8(1):190–202. doi:10.1038/nprot.2012.149. [published Online First: 20130103].
- Ruben Props PM, Mysara M, Clement L, Boon N, Boon N. Measuring the biodiversity of microbial communities by flow cytometry. *Methods Ecol Evol*. 2016;7(11):1376–1385. doi:10.1111/2041-210X.12607.
- Zimmermann J, Hübschmann T, Schattenberg F, Schumann J, Durek P, Riedel R, Friedrich M, Glauben R, Siegmund B, Radbruch A, et al. High-resolution microbiota flow cytometry reveals dynamic colitis-associated changes in fecal bacterial composition. *Eur J Immunol*. 2016;46(5):1300–1303. doi:10.1002/eji.201646297.
- Schmiester M, Maier R, Riedel R, Durek P, Frentsch M, Kolling S, Mashreghi M-F, Jenq R, Zhang L, Peterson CB, et al. Flow cytometry can reliably capture gut microbial composition in healthy adults as well as dysbiosis dynamics in patients with aggressive B-cell non-hodgkin lymphoma. *Gut Microbes*. 2022;14(1):2081475. doi:10.1080/19490976.2022.2081475.
- Kupschus J, Janssen S, Hoek A, Kuska J, Rathjens J, Sonntag C, Ickstadt K, Budzinski L, Chang H-D, Rossi A, et al. Rapid detection and online analysis of microbial changes through flow cytometry. *Cytometry A*. 2022;103(5):419–428. doi:10.1002/cyto.a.24704. [published Online First: 20221110].
- Krause JL, Schaepe SS, Fritz-Wallace K, Engelmann B, Rolle-Kampczyk U, Kleinstaub S, Schattenberg F, Liu Z, Mueller S, Jehmlich N, et al. Following the community development of SIHUMix – a new intestinal in vitro model for bioreactor use. *Gut Microbes*. 2020;11(4):1116–1129. doi:10.1080/19490976.2019.1702431. [published Online First: 20200110].
- Rubbens P, Props R, Kerckhof FM, Boon N, Waegeman W. Cytometric fingerprints of gut microbiota predict Crohn's disease state. *ISME J*. 2021;15(1):354–358.

- doi:10.1038/s41396-020-00762-4. [published Online First: 20200902].
21. Budzinski L, Radbruch A, Chang H-D. Single-cell multi-parametric characterization of microbiota by flow cytometry. *Methods Cell Biol: Academic Press*. 2024;186:91–106. doi:10.1016/bs.mcb.2024.02.023.
 22. Bak-Romaniszyn L, Swierzko AS, Sokolowska A, Durko Ł, Mierzwa G, Szala-Póździej A, Małecka-Panas E, Cedzyński M. Mannose-binding lectin (MBL) in adult patients with inflammatory bowel disease. *Immunobiol*. 2020;225(1):151859. doi:10.1016/j.imbio.2019.10.008. [published Online First: 20191031].
 23. Saba K, Denda-Nagai K, Irimura T. A C-type lectin MGL1/CD301a plays an anti-inflammatory role in murine experimental colitis. *Am J Pathol*. 2009;174(1):144–152. doi:10.2353/ajpath.2009.080235. [published Online First: 20081218].
 24. Morosi LG, Cutine AM, Cagnoni AJ, Manselle-Cocco MN, Croci DO, Merlo JP, Morales RM, May M, Pérez-Sáez JM, Girotti MR, et al. Control of intestinal inflammation by glycosylation-dependent lectin-driven immunoregulatory circuits. *Sci Adv*. 2021;7(25). doi:10.1126/sciadv.abf8630. [published Online First: 20210618].
 25. Harvey RF, Bradshaw JM. A simple index of Crohn's-disease activity. *Lancet*. 1980;315(8167):514. doi:10.1016/s0140-6736(80)92767-1.
 26. Limberg B. Diagnosis of acute ulcerative colitis and colonic Crohn's disease by colonic sonography. *J Clin Ultrasound*. 1989;17(1):25–31. doi:10.1002/jcu.1870170106. [published Online First: 1989/01/01].
 27. Jessen B, Tt M, Charite IBDG, Staudacher JJ, Haag LM, Kredel L, Schumann M, Weidinger C, Sonnenberg E, Siegmund B. Peripheral tumor necrosis factor production is a predictor for remission under adalimumab in Crohn's disease. *Clin Gastroenterol Hepatol*. 2024; doi:10.1016/j.cgh.2024.10.008. [published Online First: 20241102].
 28. Bunker JJ, Flynn TM, Koval JC, Shaw D, Meisel M, McDonald B, Ishizuka I, Dent A, Wilson P, Jabri B, et al. Innate and adaptive humoral responses coat distinct commensal bacteria with immunoglobulin A. *Immun*. 2015;43(3):541–553. doi:10.1016/j.immuni.2015.08.007. [published Online First: 2015/09/01].
 29. Shapiro JM, de Zoete MR, Palm NW, Laenen Y, Bright R, Mallette M, Bu K, Bielecka AA, Xu F, Hurtado-Lorenzo A, et al. Immunoglobulin A targets a unique subset of the microbiota in inflammatory bowel disease. *Cell Host Microbe*. 2021;29(1):83–93.e3. doi:10.1016/j.chom.2020.12.003. [published Online First: 20210101].
 30. Schierova D, Roubalova R, Kolar M, Stehlikova Z, Rob F, Jackova Z, Coufal S, Thon T, Mihula M, Modrak M, et al. Fecal microbiome changes and specific anti-bacterial response in patients with IBD during anti-tnf therapy. *Cells*. 2021;10(11):3188. doi:10.3390/cells10113188. [published Online First: 20211116].
 31. Aden K, Rehman A, Waschina S, Pan W-H, Walker A, Lucio M, Nunez AM, Bharti R, Zimmerman J, Bethge J, et al. Metabolic functions of gut microbes associate with efficacy of tumor necrosis factor antagonists in patients with inflammatory bowel diseases. *Gastroenterol*. 2019;157(5):1279–1292.e11. doi:10.1053/j.gastro.2019.07.025. [published Online First: 20190718].
 32. Serrano-Gomez G, Mayorga L, Oyarzun I, Roca J, Borruel N, Casellas F, Varela E, Pozuelo M, Machiels K, Guarner F, et al. Dysbiosis and relapse-related microbiome in inflammatory bowel disease: a shotgun metagenomic approach. *Comput Struct Biotechnol J*. 2021;19:6481–6489. doi:10.1016/j.csbj.2021.11.037. [published Online First: 20211202].
 33. Bondareva M, Budzinski L, Durek P, Witkowski M, Angermair S, Ninnemann J, Kreye J, Letz P, Ferreira-Gomes M, Semin I, et al. Cross-regulation of antibody responses against the SARS-CoV-2 spike protein and commensal microbiota via molecular mimicry. *Cell Host Microbe*. 2023;31(11):1866–1881.e10. doi:10.1016/j.chom.2023.10.007.
 34. Bellais S, Nehlich M, Ania M, Duquenoy A, Mazier W, van den Engh G, Baijer J, Treichel NS, Clavel T, Belotserkovsky I, et al. Species-targeted sorting and cultivation of commensal bacteria from the gut microbiome using flow cytometry under anaerobic conditions. *Microbiome*. 2022;10(1):24. doi:10.1186/s40168-021-01206-7. [published Online First: 20220203].
 35. Oksanen J, Blanchet F. G, Kindt R, Legendre P, Minchin P.R, O'hara R.B, Simpson GL, Solymos P, Stevens MHH, Wagner H. Package 'vegan'. *Community Ecol Package, Version 2*. 2013. <https://cran.r-project.org/web/packages/vegan/vegan.pdf>.
 36. Hahne F, LeMeur N, Brinkman RR, Ellis B, Haaland P, Sarkar D, Spidlen J, Strain E, Gentleman R. flowCore: a bioconductor package for high throughput flow cytometry. *BMC Bioinf*. 2009;10(1):106. doi:10.1186/1471-2105-10-106. [published Online First: 20090409].
 37. Finak GJM. flowWorkspace: infrastructure for representing and interacting with gated and ungated cytometry data sets. 2022. doi:10.18129/B9.bioc.flowWorkspace
 38. Wehrens RB, Buydens LMC. Self- and Super-organizing Maps in R: the kohonen Package. *J Stat Softw*. 2007;21(5). doi:10.18637/jss.v021.i05.
 39. Wehrens RK, Kruisselbrink J. Flexible self-organizing maps in kohonen 3.0. *J Stat Softw*. 2018;87(7). doi:10.18637/jss.v087.i07.
 40. Kuhn M. caret: classification and regression training. 2016. <https://cran.r-project.org/web/packages/caret/caret.pdf>

41. John CR. Mlevel: machine learning model evaluation. 2020. <https://cran.r-project.org/web/packages/MLevel/MLevel.pdf>
42. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. *Nucleic Acids Res.* 2013;41(1):e1. doi:10.1093/nar/gks808. [published Online First: 20120828].
43. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformat.* 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170. [published Online First: 20140401].
44. Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP. DADA2: high-resolution sample inference from illumina amplicon data. *Nat Methods.* 2016;13(7):581–583. doi:10.1038/nmeth.3869. [published Online First: 20160523].
45. McLaren MRC, J B. Silva 138.1 prokaryotic SSU taxonomic training data formatted for DADA2. Zenodo. 2021; doi:10.5281/zenodo.4587955.
46. Wright E. Using DECIPHER v2.0 to analyze big biological sequence data in R. *R J.* 2016;8(1):352–359. doi:10.32614/RJ-2016-025.
47. Price MN, Dehal PS, Arkin AP. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol Biol and Evol.* 2009;26(7):1641–1650. doi:10.1093/molbev/msp077. [published Online First: 20090417].
48. Wang Q, Garrity GM, Tiedje JM, Cole JR. Naïve bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol.* 2007;73(16):5261–5267. doi:10.1128/AEM.00062-07. [published Online First: 20070622].
49. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res.* 2013;41(Database issue):D590–6. doi:10.1093/nar/gks1219. [published Online First: 20121128].