ORIGINAL ARTICLE

Uncoupling of mucosal gene regulation, mRNA splicing and adherent microbiota signatures in inflammatory bowel disease

Robert Häsler,[1] Raheleh Sheibani-Tezerji,[1] Anupam Sinha,[1] Matthias Barann,[1] Ateequr Rehman,[1] Daniela Esser,[2] Konrad Aden,[1] Carolin Knecht,[3] Berenice Brandt,[4] Susanna Nikolaus,[4] Sascha Schäuble,[5] Christoph Kaleta,[2] Andre Franke,[1] Christoph Fretter,[6] Werner Müller,[7] Marc-Thorsten Hütt,[6] Michael Krawczak,[3] Stefan Schreiber,[1,4] Philip Rosenstiel[1]

ABSTRACT

Objective An inadequate host response to the intestinal microbiota likely contributes to the manifestation and progression of human inflammatory bowel disease (IBD). However, molecular approaches to unravelling the nature of the defective crosstalk and its consequences for intestinal metabolic and immunological networks are lacking. We assessed the mucosal transcript levels, splicing architecture and mucosa-attached microbial communities of patients with IBD to obtain a comprehensive view of the underlying, hitherto poorly characterised interactions, and how these are altered in IBD.

Design Mucosal biopsies from Crohn's disease and patients with UC, disease controls and healthy individuals (n=63) were subjected to microbiome, transcriptome and splicing analysis, employing next-generation sequencing. The three data levels were integrated by different bioinformatic approaches, including systems biology-inspired network and pathway analysis.

Results Microbiota, host transcript levels and host splicing patterns were influenced most strongly by tissue differences, followed by the effect of inflammation. Both factors point towards a substantial disease-related alteration of metabolic processes. We also observed a strong enrichment of splicing events in inflamed tissues, accompanied by an alteration of the mucosa-attached bacterial taxa. Finally, we noted a striking uncoupling of the three molecular entities when moving from healthy individuals via disease controls to patients with IBD.

Conclusions Our results provide strong evidence that the interplay between microbiome and host transcriptome, which normally characterises a state of intestinal homeostasis, is drastically perturbed in Crohn's disease and UC. Consequently, integrating multiple OMICs levels appears to be a promising approach to further disentangle the complexity of IBD.

Significance of this study

What is already known on this subject?
► IBD is a complex disease that is characterised by chronic inflammation of the intestinal mucosa.
► Inadequate host responses to intestinal microbiota, changes of the host transcriptome and altered host splicing patterns are known to be associated with IBD.
► If and how the above factors interact to influence disease manifestation and progression is only poorly understood.

What are the new findings?
► For the first time, we illustrate a close disease-relevant interaction between microbiota, host gene regulation and splicing of the host mucosal transcriptome.
► A structured relationship between the three factors is most evident in healthy individuals and partially reduced in acute inflammation of the gut, but strikingly perturbed in patients with IBD.
► Only a few inflammation-associated biological processes appear to drive the observed disease-relevant host-microbiome interaction.

How might it impact on clinical practice in the foreseeable future?
► Our results emphasise that the microbiota and the host mucosa represent a joint intestinal 'meta-organism' that must be viewed as a single entity to attain a better understanding of the aetiology of IBD.

INTRODUCTION

Human chronic inflammatory bowel disease (IBD) (OMIM #266600) is an archetypical chronic immune-mediated disease, characterised by excessive T-lymphocyte-driven immune responses. IBD comprises a group of disorders with two main subentities, namely Crohn's disease (CD, OMIM #266600) and UC (OMIM #191390). Recently, advances have been made towards the identification of genetic IBD risk factors (by genome-wide association studies) and the molecular dissection of IBD pathophysiology (eg, by the detection of aberrant inflammatory signal transduction).[1–4] However, why IBD preferentially manifests in a specific time

window, spanning only a few decades in life and how the disease becomes chronic over time in a given individual is still poorly understood.

The human host together with their gut microbes forms a functional unit, called 'meta-organism'.[5] The commensal microbiota is essential for shaping the immune system and is involved in many host physiological functions including, among others, the digestion of nutrients.[6] Host and microbiota are usually connected tightly in a resilient state believed to be important for the survival of the organism as a whole, thus qualifying as *bona fide* 'symbiosis'. In contrast to the fixed repertoire of germline genetic variants inherited by a given individual, however, this symbiosis is influenced by a number of environmental factors as well, including lifestyle and hygiene. We may further assume that the interplay of gene and environment is much more complex than hitherto described by simple models[7–9] even though defective host-microbiota crosstalk evidently contributes to the manifestation and progression of IBD.

The transcriptome of patients with IBD has been studied extensively since genome-wide approaches became feasible and popular, unravelling gene expression patterns of IBD in general, or CD or UC in particular.[10–14] Various efforts have been made to characterise the transcriptional changes associated with IBD in more detail, from array-based studies to more recent RNAseq experiments. This work laid the foundation for a first map of mRNA expression changes characterising the disease.[10–13 15] However, only a few studies so far have addressed the interaction between the human mucosal transcriptome and the gut microbiota[16–19] in the context of intestinal inflammation. The gut microbiota has been shown to potentially interact with genetic IBD risk factors,[20] the latter potentially influencing the response of the mucosa to the resident bacteria. Thus, it has been demonstrated that the NOD2 genotype likely influences[21] and reshapes the microbiota in a disease-dependent fashion,[22–25] a result further supported by observations of impaired mRNA expression in mouse models of colitis and colitis-associated carcinoma.

Although the relationship between variation of the human transcriptome and the human genome is increasingly better understood,[26] to the best of our knowledge, all previous studies on inflammatory diseases have neglected the complexity of the transcriptome (eg, sense/antisense transcription, alternative splicing). Currently, gene expression changes are mostly assessed at a gene-condensed level, as recently shown in paediatric IBD.[19 27] We hypothesise that studies focusing on the inference of functional changes from genetic data alone are inadequate in that they disregard the important layer of transcriptome information. In fact, a seminal publication recently emphasised that a variety of diseases can be caused by mutations that disrupt the splicing code,[28] yet studies assessing splicing alterations globally for a given biological state, such as inflammation, are still rare.

To highlight one specific functional aspect of IBD pathophysiology, we jointly analysed comprehensive host transcriptome data and mucosa-attached microbial community profiles from a cohort of patients with IBD, healthy and non-IBD intestinal inflammation controls. We employed next-generation sequencing (RNAseq and 16S rRNA amplicon sequencing) and systems biological-inspired data integration to unravel previously unrecognised, global metabolism-connected and immunity-connected expression patterns in the host mucosa and related these to IBD. Following this approach, we were able for the first time to identify the molecular elements of the host-microbiome 'meta-organism' that are likely affected by IBD and to demonstrate a rather specific disconnection of the human transcriptome from the microbiota in patients with IBD.

METHODS

Patient recruitment and biopsy selection

Biopsies from four groups of individuals were retrieved from the local hospital biobank at Kiel: healthy individuals without significant pathological findings, patients with acute non-IBD intestinal inflammation, patients with CD and UC. The samples were taken either from the terminal ileum or the sigmoidal colon. In the disease groups, both inflamed and non-inflamed material was collected. Overall, 63 biopsies from 41 individuals were available for further study (see tables 1, 2 and online supplement for further details). Patient identity was taken properly into account to allow for dependent samples in the statistical analyses. The study protocol was approved by the Ethics Committee of the Medical Faculty of Kiel University (reference #B231/98-1/13). All study participants gave written informed consent prior to sampling and data collection.

RNA extraction and sample preparation

Biopsies were obtained endoscopically during routine diagnosis and stored in liquid nitrogen. Total RNA was extracted using RNeasy (Qiagen, Germany) according to the manufacturer's protocols, including repeated tissue crushing using a Teflon head (Omnilab, Bremen, Germany) under liquid nitrogen to augment bacterial cell wall disruption. RNA was prepared as previously described,[29] with RNA integrity assessed using an Agilent Bioanalyzer following the manufacturer's instructions.

Mucosa-attached microbiota analysis

Transcriptionally active mucosa-attached bacterial profiles were generated from mucosal biopsy RNA (converted to cDNA using random hexamers, Qiagen). The 16S rRNA variable region V3-V4 was amplified using bacterial 16S rRNA gene-specific composite primers (319F and 806R) as described.[30] Pooled amplicon libraries were sequenced employing an Illumina MiSeq platform (2×300 bp) and processed further as described in the online supplementary methods section. Unprocessed 16S rRNA reads (fastq) have been submitted in the European Nucleotide Archive under accession number PRJEB14288.

Host transcriptome analysis

Host transcriptome analysis was performed by RNA sequencing of the primary samples described above. Strand-specific sequencing libraries were prepared with the TruSeq stranded Total RNA kit (Illumina) from 1 µg total RNA of each sample and sequenced on an Illumina HiSeq2000 (100-nucleotide paired-end reads). Reads were quality-controlled, aligned and processed further using Bioconductor package DeSeq2. The Kyoto Encyclopedia of Genes and Genomes-GenomeNet (KEGG) enrichment and transcription factor binding site analyses were carried out using the Innate DB database.[31] For further details, see online supplementary methods.

Table 1 Characteristics of biopsies

Subgroup (female; male; age range)	Sigma (non-inflamed; inflamed)	Terminal ileum (non-inflamed; inflamed)
Healthy individuals (7; 5; 18–72)	6 (6; 0)	6 (6; 0)
Disease controls (13; 2; 24–68)	8 (4; 4)	7 (5; 2)
Crohn's disease (15; 4; 18–47)	8 (4; 4)	11 (6; 5)
UC (9; 8; 24–55)	9 (5; 4)	8 (6; 2)
Total (44; 19; 18–72)	31 (19; 12)	32 (23; 9)

Table 2 Individual medication overview

Subgroup	5-Aminosalicylic acid	Azathioprine	Biologics	Corticosteroids	Treatment naive
Healthy individuals	0/7	0/7	0/7	0/7	7/7
Disease controls	0/8	0/8	2/8	2/8	6/8
Crohn's disease	3/13	3/13	4/13	3/13	2/13
UC	5/13	2/13	5/13	2/13	0/13
Total	5/41	5/41	11/41	7/41	15/41

Numbers listed are patients with medication/all patients.

Analysis of splicing

RNA splicing products were identified and quantified using vast-tools.[32] The level of alternative product inclusion was defined as either the percentage spliced in (PSI), for exons, the percentage intron retention (PIR), for introns, or the percentage splice site usage (PSU) for alternative splice site acceptor choice (ALTA) and alternative splice site donor choice (ALTD), as appropriate. A splicing event was defined by fulfilling the following criteria: (1) it had an alternative product inclusion level ≥ 10 and <100 and (2) it had sufficient read coverage in at least 90% of all the samples.[32] For further details, see online supplementary methods.

Analysis of host-microbiome crosstalk

To obtain a quantitative measure of host-microbe crosstalk, the Spearman's rank correlation coefficient was calculated, for all differentially expressed genes and all operational taxonomical units (OTUs) present in at least 50% of all samples, between the respective gene expression level and the respective OTU abundance. To determine the statistical significance of individual correlation coefficients, the false discovery rate (FDR) was estimated employing a modified Westfall and Young permutation approach.[33] Group-wise differences in Spearman's rank correlation coefficient were assessed for statistical significance employing a Mann-Whitney U test.

Reconstruction of context-specific metabolic models and analysis of metabolic network coherence

To create context-specific metabolic models of transcriptional activity, the gene expression data were mapped to the human metabolic model Recon 2.04[34] using the intensity modulated arc therapy (iMAT)-implementation of the constraint-based reconstruction and analysis (COBRA)-toolbox.[35] Metabolic network coherence was then computed based upon dichotomised gene expression data as described[36] with modifications facilitating an application to human data.[37] For details, see online supplementary methods.

RESULTS

The IBD-associated mucosal transcriptome

We used deep RNA sequencing to identify transcriptome alterations that were associated with IBD. Filtering and trimming of RNA sequencing data from 63 mucosal biopsies excluded between 1% and 4% of reads (adapter contamination and low quality), leaving 89–167 million read pairs for subsequent mapping. The proportion of reads mapping to the human reference genome varied between 89% and 98% over biopsies. The raw read counts are provided in online supplementary table S1.

Some 4576 genes were found to be significantly differentially expressed between cases and controls (see online supplementary table S2). Using global scaling methods, most of the variation in expression level was explained by tissue type, followed by

inflammation status. In contrast, clinical diagnosis did not discriminate well between expression profiles, as was highlighted by principle component analysis (figure 1A–C). The first two dimensions explained 19% of the overall variance in expression level (figure 1D). Of the differentially expressed genes, 2466 were upregulated and 2110 were downregulated in patients with IBD (figure 2A: unique and shared features in differentially gene expression; figure 2B: top 200, ranked by fold change after applying a cutoff of $p \leq 0.01$). The top 50 genes (ranked by fold change after applying a p value cutoff of 0.01) were differentially regulated exclusively in either of the two IBD subtypes and are displayed in figure 2C (CD) and 2D (UC).

The KEGG pathway analysis identified several functional categories comprising differentially expressed genes, including overexpression of adhesion molecules and of genes related to fatty acid biosynthesis and natural killer cell-mediated cytotoxicity (figure 3A) as well as underexpression of genes related to spliceosome assembly and nicotinate (NA)/nicotinamide (NAM) metabolism (figure 3B).

Transcription factor analysis inferred signal transducer and activator of transcription 1 (STAT1) (figure 3A) and nuclear factor-κB (NF-κB) reticuloendotheliosis proto-oncogene, NF-kB subunit (REL) binding motifs as major inflammation-associated sites, regardless of diagnosis, whereas heat shock transcription factor 1 (HSF1) and Sox5 binding motifs were the most prevalent sites for CD and UC, respectively. Binding sites for aryl hydrocarbon receptor (AhR) and fetal Alzheimer antige clone 1 (FAC1) were significantly depleted in genes downregulated by non-specific inflammation. For further information, see figure 3B.

Reconstruction of context-specific metabolic networks and host metabolic network coherence in IBD

Models comprising between 2235 and 2668 metabolic reactions were generated and analysed using classical data mining and a previously described metabolic network coherence approach.[36] [37] Comparing the number of reactions found in networks reconstructed from samples of non-inflamed versus inflamed tissues, we observed a significant decrease of inferred metabolic activity in the latter (FDR-corrected p value=4.0×10^{-2}), with 8 of 12 altered pathways displaying a significant decrease of inferred activity (table 3). Furthermore, metabolic coherence values (M) measuring the connectedness of expression level alterations in a predefined network (Recon2.04) differed substantially between groups defined by tissue type, inflammation status and disease type (see online supplementary figure S1). The differential expressed genes originating from the comparison of inflamed and non-inflamed regions of the terminal ileum for CD, resulted in a positive metabolic network coherence (M=0.74). The effective network consisted of a single large and dense component in addition to several smaller network fragments (see online supplementary figure S1B), pointing towards a rather specific change in metabolic activity.

Figure 1 Principal component analysis (PCA) of mRNA expression levels. The first two components shown explain the largest part of the variation in mRNA expression. Individual insets are colour-coded by tissue (A), inflammation status (B) and diagnosis (C). (D) Variance explained by the first 20 PCA components.
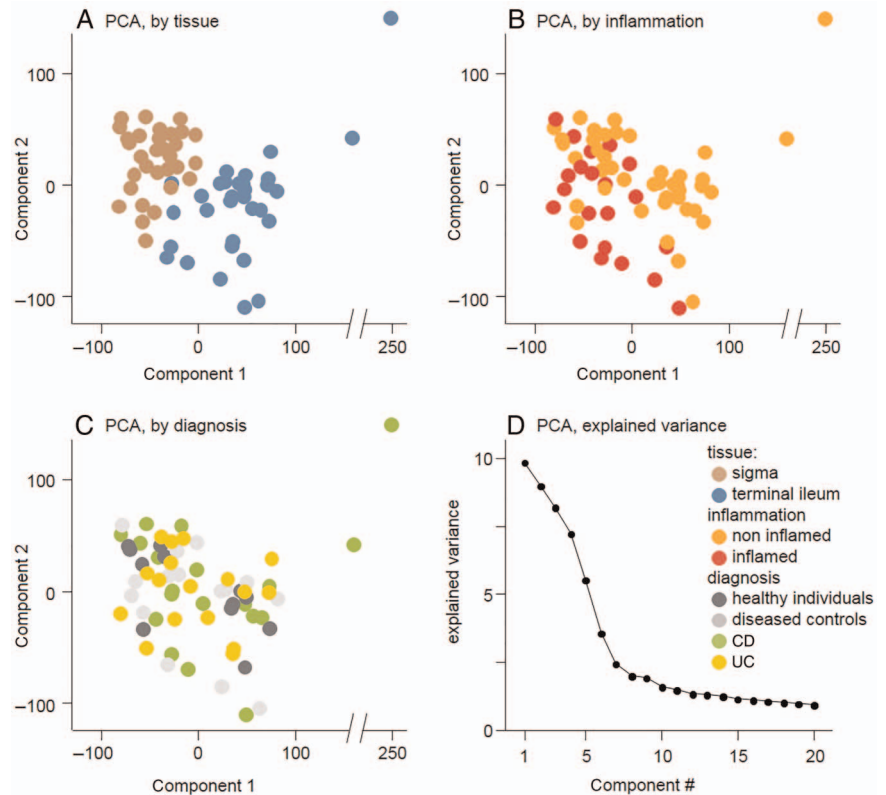
Similarly, differentially expressed genes originating from the comparison of samples from inflamed and non-inflamed sigma regions in cases of UC yielded a high metabolic network coherence (M=1.44). Again, one cluster dominated the effective network (see online supplementary figure S1C). The contextualisation of these metabolic networks by way of gene ontology analysis highlighted the associated metabolic functions and processes (see online supplementary figure S2).

Alternative splicing in IBD
The proteomes of both healthy and diseased tissues attain an extra level of complexity by alternative splicing that may fall into one of four major categories (figure 4A): intron retention, exon skipping (ES), ALTA and ALTD. To reflect the distribution of splicing events, we further categorised them with respect to their relative occurrence in the corresponding sample groups (events occurring in either 100%, 90%, 50% or 10% of all samples, figure 4B). Only high-quality splicing events that occurred in at least 90% of all samples per experimental group were subjected to further analysis. Note that the sensitivity of alternative splicing event detection by way of RNA sequencing is heavily dependent upon the read coverage across samples.

The nature of the alternative splicing events detected in individual sample groups is summarised in figure 4C, highlighting unique and shared events for all pair-wise group comparisons. Briefly, tissue difference accounted for 25 292 alternative splicing events, while 25 283 events were attributable to general inflammation status. At the individual disease level, 17 534 alternative splicing events were associated with inflammation in CD and 17 928 events occurred in UC. Across all samples, ES was found to be the most prevalent event (figure 4D). Simultaneous consideration of tissue type, inflammation stage and diagnosis identified 320 shared alternative splicing events (see figure 5A and online supplementary table S3). The KEGG pathway

analysis of genes linked to these 320 events revealed a significant enrichment of a number of different categories (figure 5B), including bacterial invasion of epithelial cells. Spearman's correlation analysis of PSI/PIR/PSU values and the respective gene expression levels revealed only weak relationships, suggesting that splicing occurs largely independent of gene regulation (figure 5C).

Subtle alteration of the mucosa-attached microbiome associated with IBD
Analysis of mucosa-attached active bacterial communities resulted in a total of 1 918 235 16S rRNA amplicon reads, with an average of 32 517 sequence reads per sample. For quality reasons, three samples were excluded from further analysis. Some 19 phyla and 3120 OTUs were detected, including the most dominant phyla Firmicutes (62.23%), Bacteroidetes (26.62%), Proteobacteria (8.71%) and Actinobacteria (1.51%) (see online supplementary figure S3; for the full dataset and online supplementary table S4). Similar to other studies at the RNA level, bacterial community richness and diversity did not differ significantly between patients with CD or UC, respectively, and controls. Only a small number of OTUs and low Shannon diversity were observed in patients with CD (see online supplementary figure S4A: terminal ileum and online supplementary figure S4B: sigmoidal colon), while multivariate analysis Permutational Multivariate Analysis of Variance (PERMANOVA) of abundance-based (Bray-Curtis) and non-abundance-based (Jaccard) distance matrices revealed significant differences between patient and control microbial communities. Regardless of tissue type (terminal ileum or sigma), inflammation status was significantly associated with altered microbiota. Principal coordinate analysis of the two distance matrices showed seemingly distinct clusters, based upon diagnosis, while clustering based upon tissue or inflammation status was not
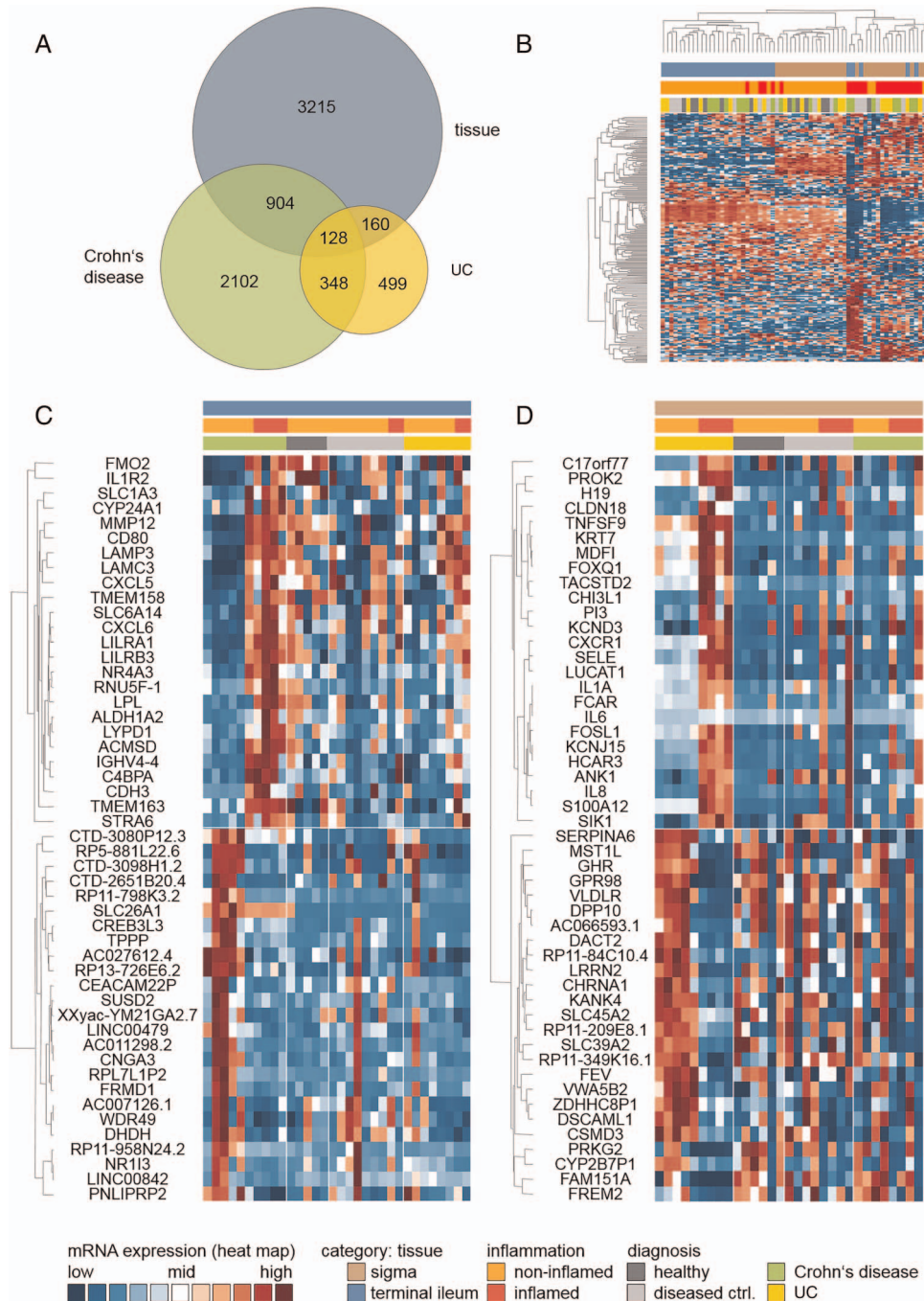
Figure 2 Hierarchical clustering of genes differentially expressed in IBD, Crohn's disease (CD) or UC. Hierarchical clustering was based upon relative expression levels. Each row corresponds to a single gene, whereas each column corresponds to a single sample. (A) Venn diagram illustrating overlaps and unique features identified in different pair-wise comparisons (tissue: terminal ileum vs sigmoidal colon; CD inflamed vs non-inflamed; UC inflamed vs non-inflamed); (B) Clustering of top 200 genes differentially expressed between tissue type, inflammation status or IBD subtype; (C) Clustering of top 50 genes differentially expressed in CD only and (D) clustering of top 50 genes differentially expressed in UC only. Genes were ranked by fold change after applying a p value cutoff of 0.01. Genes are labelled with gene symbols.

evident (see online supplementary figure S5). These observations support the view that active mucosa-attached microbiota are different in healthy individuals and patients with IBD, and that they are also significantly altered by inflammation status.

Co-abundance analysis of microbiome and host transcriptome
As a measure of host-microbiome crosstalk, we correlated the expression level of differentially expressed genes with the abundance of bacterial OTUs, revealing 5988 significantly correlated pairs (FDR≤0.05). These pairs resulted from six different

correlation analyses, namely for terminal ileum samples of (i) healthy individuals, (ii) disease control individuals and (iii) patients with CD, and for sigmoidal colon samples of (iv) healthy individuals, (v) disease control individuals and (vi) patients with UC. Inspection of the top 500 correlations for each tissue type (ranked by a combination of Spearman's rank correlation coefficient and FDR) revealed a trend from healthy individuals (318 significantly correlated pairs) via disease controls (155 pairs) to patients with CD (terminal ileum only: 31 pairs). With the exception of sigmoidal colon from patients with
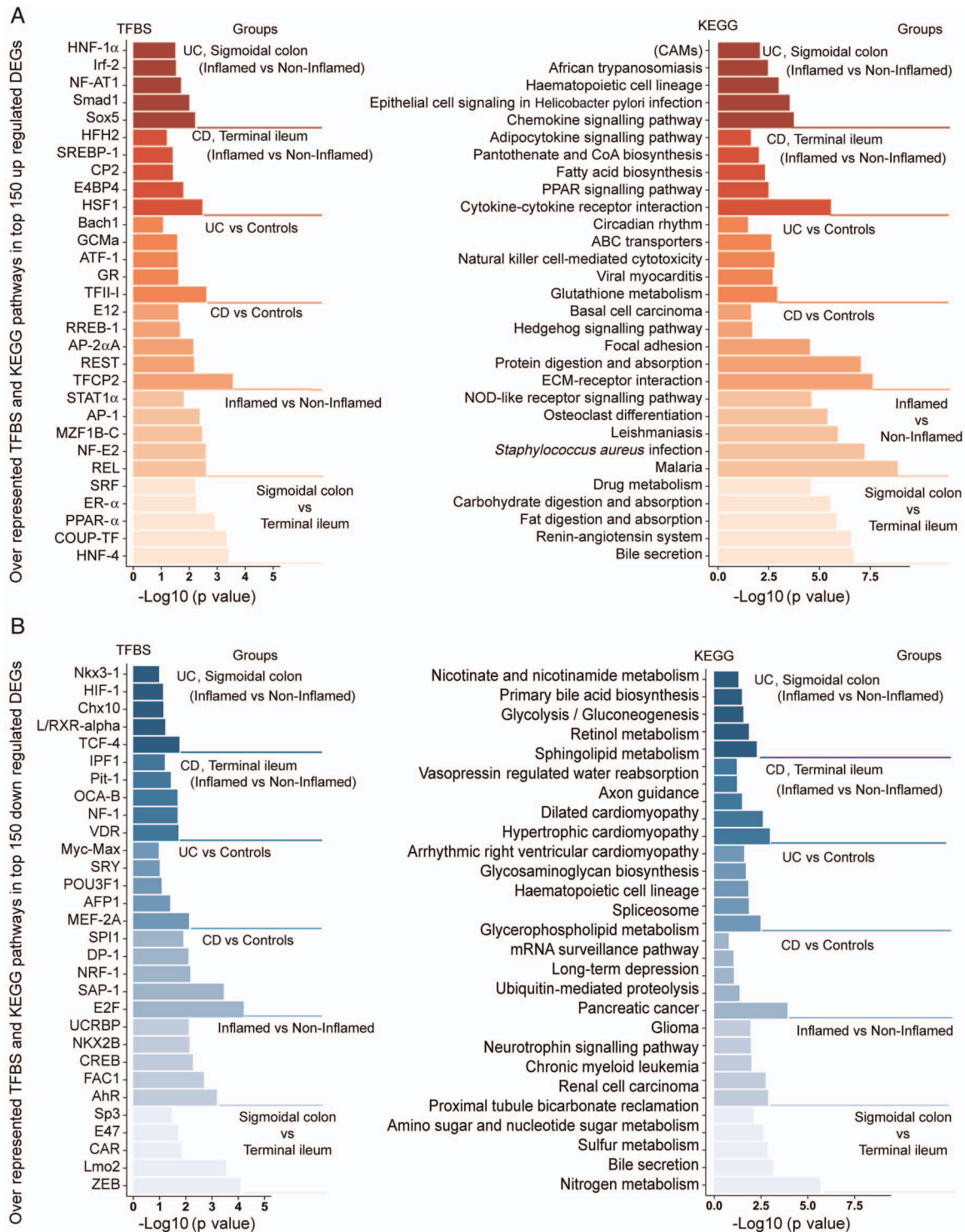
Figure 3 Transcription factor binding side and the KEGG pathway analysis of genes differentially expressed in IBD. Transcription factor binding sites (left panel) and the KEGG pathway categories (right panel) were identified for the top 150 genes upregulated (A) or downregulated (B) in IBD. The differentially expressed genes included origin from the pair-wise comparisons of the tissue type (sigmoidal colon vs terminal ileum), inflammatory stage (inflamed vs non-inflamed) and diagnostic status (Crohn's disease (CD) terminal ileum inflamed vs non-inflamed, UC sigmoidal colon inflamed vs non-inflamed).

UC and non-IBD controls (p=0.523), all other Spearman's rank correlation coefficients were found to differ statistically significantly between patient/tissue groups ($p < 10^{-5}$). The majority of the top 200 differentially expressed genes (ranked by fold change after applying a p value cutoff of 0.01) also showed a correlation to bacterial OTUs (186 of 200 in CD and 194 of 200 in UC). With both diseases, only three bacterial families accounted for over 70% of all disease-specific significant correlations between abundance and transcription level, namely Lachnospiraceae, Bacteroidaceae and Ruminococcaceae. Assigning the involved genes to functional categories showed that, in patients with CD, the majority of the correlations (12 of

Table 3 Metabolic pathways significantly differentially regulated in inflamed versus non-inflamed tissues

Pathway	Direction	p Value
Dietary fibre binding	Repressed	9.8×10^{-3}
Citric acid cycle	Repressed	8.2×10^{-3}
Cytochrome metabolism	Repressed	8.2×10^{-3}
Reactive oxygen species detoxification	Repressed	7.5×10^{-4}
Bile acid synthesis	Repressed	4.9×10^{-2}
Fatty acid oxidation	Repressed	4.1×10^{-4}
Network size	Repressed	4.0×10^{-2}
Steroid metabolism	Repressed	4.0×10^{-2}
Pyrimidine catabolism	Induced	4.0×10^{-2}
Nucleotide salvage pathway	Induced	4.0×10^{-2}
Tryptophan metabolism	Induced	4.0×10^{-2}
Propanoate metabolism	Repressed	4.0×10^{-2}
Glycine, serine, alanine and threonine metabolism	Induced	3.3×10^{-2}

p Values were adjusted for multiple testing employing a false discovery rate-based method (Benjamini-Hochberg).

31) were attributable to one of two categories: chemotaxis and inflammation. In patients with UC, a single bacterial OTU, which originated from the Bacteroidetes phyla, accounted for 40% of these correlations and paired with host transcripts from all functional categories (figure 6). Finally, functional classification of those genes that were involved in at least one significant correlation revealed that 208 of 1507 genes were functionally related to splicing and RNA processing. On the other hand, a total of 390 genes actually showed signs of alternative splicing. Similarly, for UC, 431 of 2286 categorised genes were functionally related to splicing and RNA processing, 617 genes were identified actually to be alternatively spliced. Taken together, these findings indicate a notable disease-associated loss of interaction between the host transcriptome and the mucosal microbiota.

DISCUSSION

In our study, we jointly investigated three closely related OMICs layers, namely host gene expression level, host transcript splicing pattern and mucosa-associated active microbiota, in the context of IBD. We described for the first time how these three components interact quantitatively and how the interactions are altered in the presence of IBD. Our key finding was a significant uncoupling of host gene expression and microbial signature in patients with IBD that was accompanied by substantial changes in splicing patterns. These findings seem to be an IBD-specific phenomenon because, in the disease control group available to us (mainly infectious acute non-IBD intestinal inflammation), a much weaker uncoupling was observed than in patients with IBD.

Before alluding in more detail to the genome-wide aspects of our study, we will discuss several selected host gene sets in order to put our results in the context of previous findings. Of course, this highlighting of individual genes does not mean to suggest that the respective results are necessarily of higher relevance than the network-based findings discussed later.

Several of the genes found to be upregulated or downregulated in our samples from patients with IBD have been addressed in previous studies. Many of them were thus shown to be of pathophysiological importance, including interleukin-1 receptor type II, interleukin 1α, interleukin-6 (*IL6*) and interleukin-8, all of which are key regulators of inflammation in IBD.[38] Similarly, we observed an upregulation of claudin 18 in UC samples, which corroborated similar findings obtained before irrespective of disease severity, thereby suggesting an epithelial defect underlying the disease.[39] The KEGG pathway enrichment analyses pointed towards a functional role of differentially regulated genes associated with chemokine signalling, natural killer cell-mediated cytotoxicity, NOD-like receptor signalling, drug metabolism or spliceosome assembly, all of which are highly relevant in the context of microbiota-associated inflammation and the development of IBD. At the same time, we also observed many known IBD-relevant genes to be alternatively spliced in IBD. However, this does not mean that alternative splicing is causally related to changes in gene expression. In fact, in concordance with a previous study,[40] we found only a weak correlation between alternative splicing and differential gene expression. Genes found in our study to be alternatively included autophagy-related 16-like 1, forkhead box P1, various interleukins (*IL22RA1*, *IL10RA/B*, *IL17RA/B/C/E*, *IL1R1/2*, *IL2RG*, *IL4R*, *IL6R*, *IL6ST*, *IL18*, *IL32*) and dual oxidase 2 (*DUOX2*). Interestingly, mutations in DUOX2 and in splicing elements of other NAM adenine dinucleotide phosphate oxidase genes have been linked to very early onset IBD before, suggesting a direct link between altered splicing of a risk gene and susceptibility to IBD.[41] Transcription factors reflect key regulatory elements of transcriptional response. The inferred transcription factor binding sites we identified here point towards inflammation pathways that overlap between IBD subentities, for example, NF-κB, activator protein 1 and STAT1-α, and also invoke potentially relevant unique targets, for example, Nuclear factor of activated T-cells 1 (NF-AT1) and SMAD (small mothers against decapentaplegic) family member 1 (SMAD1) (enriched in upregulated transcripts in UC inflamed tissue only) or vitamin D receptor (uniquely enriched in downregulated transcripts in CD inflamed tissue).

We next tried to interpret the complex individual relationships between different data layers from a systems biology perspective by mapping our transcriptome data to a predefined model of cellular metabolism. Although a metabolic component of mucosal inflammation can be assumed to be present,[42] metabolome-wide approaches to contextualise gene expression profiles in IBD are still missing. Importantly, the employed model contains information on the directionality of flux, which may be used to infer cause-effect relationships without knowing exact metabolite levels. The role of cellular metabolism in regulating the gut microbiome is receiving increasing attention and has recently been suggested to play a more important role in shaping the gut microbiome than the immune system itself.[43] Coherence analysis showed a striking 'connectedness' (ie, high coherence) of transcriptome changes in IBD, particularly for subnetworks highlighting only downregulated genes. At the level of individual pathways, notable regulatory changes included a decreased biosynthesis of bile acids and an inferred downregulation of the short-chain fatty acid propanoate metabolism. Both metabolites are key components of host-microbiome crosstalk. The inferred decrease in propanoate metabolism indicates that this metabolite may be produced in smaller quantities by the microbiome, which could drive inflammation due to its anti-inflammatory effect.[44] Bile acids, in turn, have previously been implied as key regulators of microbial composition.[45] Hence, a reduction of their synthesis would point towards a potential causal mechanism underlying the breakdown of microbiota-host interactions as observed in our IBD samples. The upregulation of tryptophan metabolism represents another hallmark of inflammation.[46] Our results indicate that the previously reported reduction of serum tryptophan levels in IBD[47]
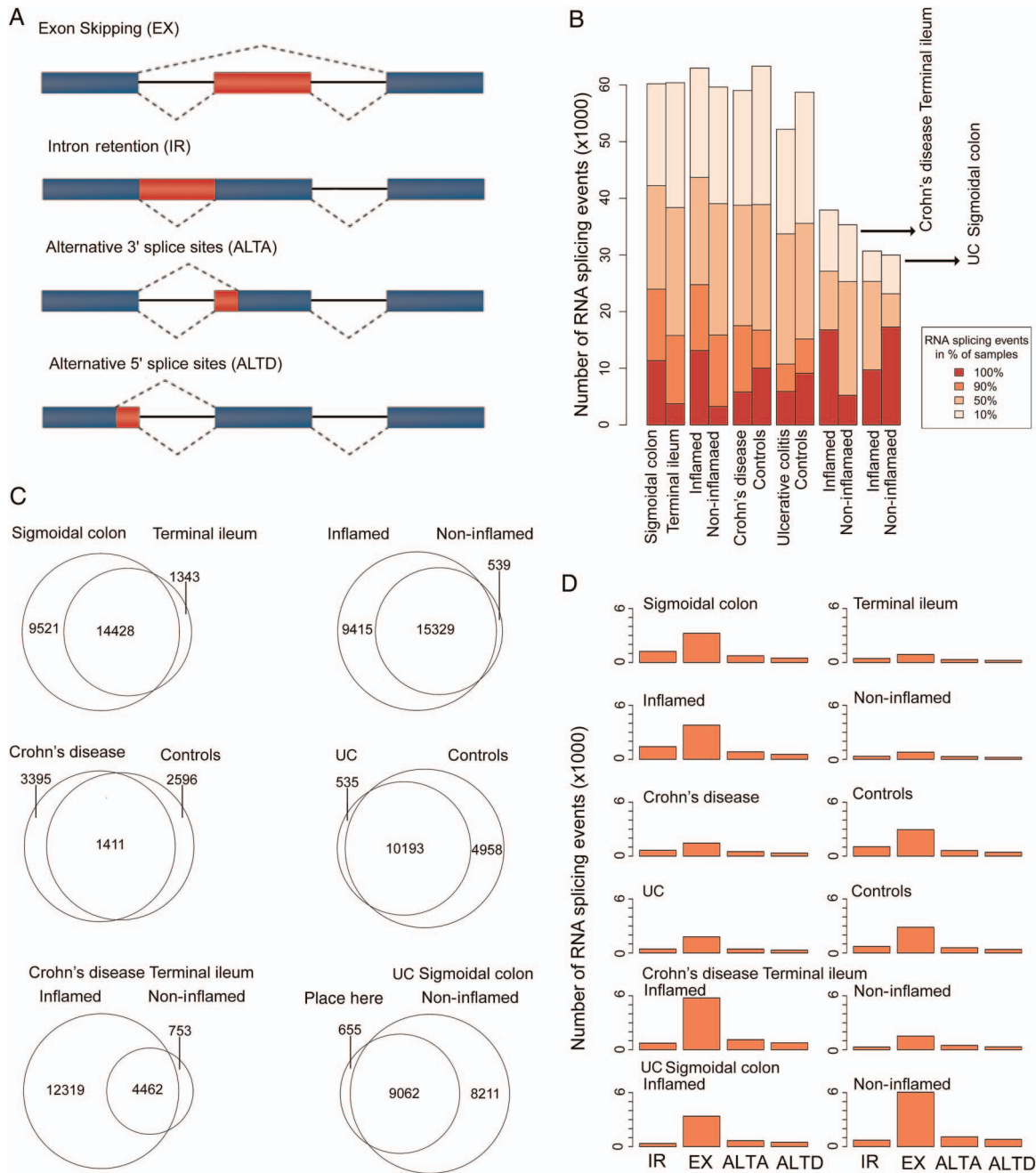
Figure 4 Alternative splicing in relation to tissue type, inflammation status and IBD subtype. (A) Types of alternative splicing events; (B) distribution of alternative splicing events of different prevalence, stratified by group-wise comparison type (tissue, diagnostic status, inflammation); (C) Venn diagram of alternative splicing events shared or unique in different group-wise comparisons (only high-quality events detected in at least 90% of IBD biopsies included). (D) Distribution by alternative splicing type detected in different group-wise comparisons (only high-quality events that occurred in at least 90% of all samples included). IR, intron retention; EX, exon skipped; ALTA/ALTD, alternative splice-site acceptor/donor.

may be due to an active upregulation of tryptophan degradation in the intestine during inflammation. A critical pathway downstream of tryptophan breakdown, which constitutes further processing of NAM/NA, is downregulated in the colonic mucosa of patients with UC and CD, which may explain the low level of NAM vitamin observed in the serum of patients with IBD.[48] Indeed, countering tryptophan/NAM depletion by dietary supplementation of NAM has been found to reduce inflammation in a murine model of IBD.[49]

Finally, we made the striking observation that the vast majority (>90%) of genes differentially expressed in IBD and controls showed a significant correlation of their expression level to the abundance of microbial taxa only in the healthy state. The correlation was slightly lower in cases of non-IBD inflammation but was almost completely lacking in patients with IBD. This gradient may point towards an uncoupling of the host and intestinal microbiota, which has already been suggested before by a small study of monozygotic twins discordant for IBD.[18] Likewise, the disease-associated gradient in crosstalk may also assign functional relevance to the correlated bacterial OTUs. The finding that three bacterial families accounted for over 70% of the residual host-microbiota correlations in patients with IBD, namely Lachnospiraceae, Bacteroidaceae and Ruminococcaceae, supports their potential role in disease pathogenesis.
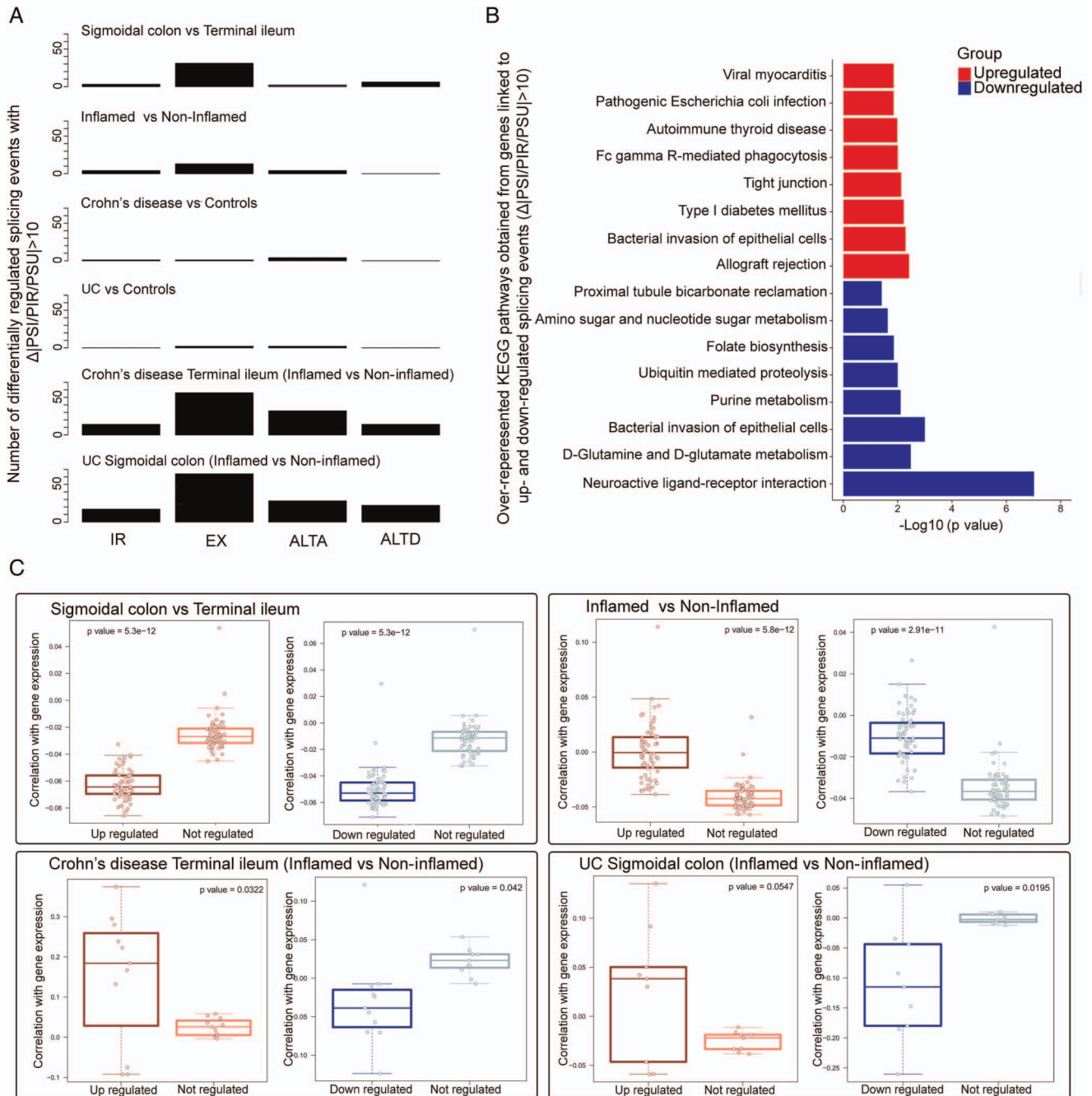
Figure 5 Differentially regulated splicing events between tissue types, inflammatory stages and IBD subtypes. (A) Distribution of differentially regulated splicing events found in different comparisons (tissue type: sigmoidal colon vs terminal ileum; inflammatory stage: inflamed vs not-inflamed; diagnosis status: Crohn's disease, terminal ileum inflamed vs non-inflamed and UC, sigmoidal colon inflamed vs non-inflamed) in IBD biopsy samples (differentially regulated splicing events were considered to have $\Delta|(PSI/PIR/PSU)|\geq10$ between two conditions in each pair-wise comparison). (B) The KEGG pathway enrichment categories for all genes linked to splicing events in (A) that were observed to be differentially upregulated/downregulated in IBD biopsy samples. The x-axis indicates their associated enrichment p value for each category. (C) The distribution of Spearman's rank correlation coefficients between PSI/PIR/PSU of differentially regulated splicing events and their corresponding gene expression values that are upregulated/downregulated or not regulated across IBD biopsy samples. p Values were generated employing a Wilcoxon rank-sum test. PSI, percentage spliced in; PIR, percentage intron retention; PSU, percentage splice site usage; IR, intron retention; EX, exon skipped; ALTA/ALTD, alternative splice-site acceptor/donor.

It is clearly not possible to draw safe functional conclusions from our rather preliminary observations. Naturally, clinical samples exhibit a high degree of variation. In addition to inter-individual differences such as genotype and medication, altered cell compositions in biopsies can also contribute to this variation. We have tried to allow for these imponderabilities by appropriate sampling and a choice of robust statistical methods. However, it cannot be excluded, for example, that the proportion of one or the other type of immune cells differed between our inflamed IBD and non-IBD disease control biopsies or some signals were even lost due higher sample-to-sample variability introduced by the different medications. This notwithstanding, taken together, our results highlight a close interaction between microbiota, gene regulation and splicing architecture of the host mucosal
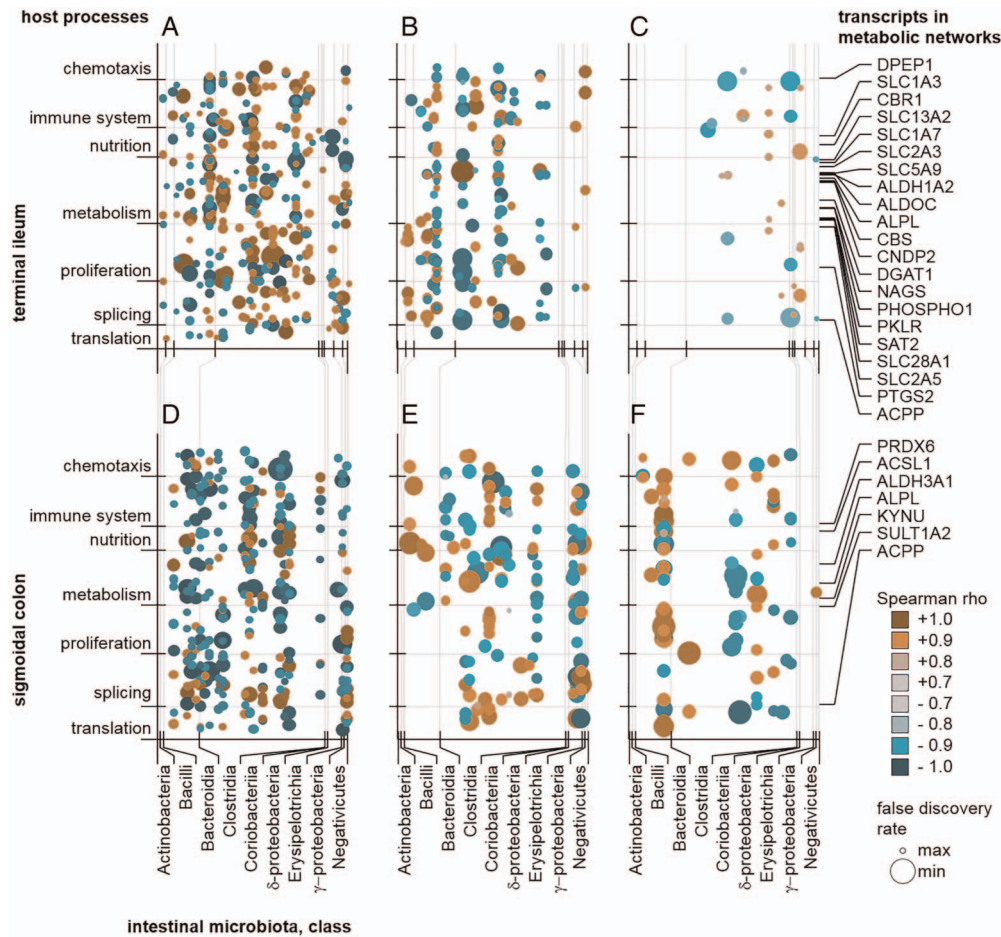
Figure 6 Loss of host-microbiome interaction. Host-microbiome interaction was quantified by Spearman's rank correlation coefficient between relative host transcript amount (mRNA) and operational taxonomical unit (OTU) count. Only mRNA-OTU pairs with a significant correlation (false discovery rate (FDR)≤0.05) are shown. Spearman's rank correlation coefficients are colour-coded, whereas the circle size depicts the FDR. (A) Terminal ileum in healthy individuals; (B) terminal ileum in diseased controls; (C) terminal ileum in patients with Crohn's disease; (D) sigmoidal colon in healthy individuals; (E) sigmoidal colon in diseased controls; (F) sigmoidal colon in patients with UC. Host gene transcripts are grouped by functional category based upon gene ontology terms. Genes that are part of a metabolic network are labelled with by the corresponding gene symbol. OTUs were categorised by associated bacterial class.

transcriptome, which is only partially impeded in acute non-IBD inflammation, but significantly reduced in chronic IBD. Unravelling the precise nature of this interaction will be a challenge for future research, but the findings presented here emphasise that the intestinal meta-organism comprising microbiota and host mucosa has to be regarded as a joined entity to allow a better understanding of the aetiology of IBD.

Author affiliations
[1]Institute of Clinical Molecular Biology, Christian Albrechts University of Kiel, Kiel, Germany
[2]Institute for Experimental Medicine, Christian Albrechts University of Kiel, Kiel, Germany
[3]Institute of Medical Informatics and Statistics, Christian Albrechts University of Kiel, Kiel, Germany
[4]Department of General Internal Medicine, University Hospital Schleswig-Holstein Campus Kiel, Kiel, Germany
[5]Language and Information Engineering Lab, Friedrich-Schiller-University Jena, Jena, Germany
[6]Department of Life Sciences and Chemistry, Jacobs University, Bremen, Germany
[7]Faculty of Life Sciences, University of Manchester, Manchester, UK

REFERENCES
1. Hampe J, Schreiber S, Shaw SH, et al. A genomewide analysis provides evidence for novel linkages in inflammatory bowel disease in a large European cohort. Am J Hum Genet 1999;64:808–16.
2. Stoll M, Corneliussen B, Costello CM, et al. Genetic variation in DLG5 is associated with inflammatory bowel disease. Nat Genet 2004;36:476–80.

3 Franke A, Balschun T, Karlsen TH, *et al*. Replication of signals from recent studies of Crohn's disease identifies previously unknown disease loci for ulcerative colitis. *Nat Genet* 2008;40:713–15.

4 Franke A, McGovern DPB, Barrett JC, *et al*. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet* 2010;42:1118–25.

5 Bik EM. Composition and function of the human-associated microbiota. *Nutr Rev* 2009;67(Suppl 2):S164–171.

6 Doré J, Blottière H. The influence of diet on the gut microbiota and its consequences for health. *Curr Opin Biotechnol* 2015;32:195–9.

7 Rawls JF, Mahowald MA, Ley RE, *et al*. Reciprocal gut microbiota transplants from zebrafish and mice to germ-free recipients reveal host habitat selection. *Cell* 2006;127:423–33.

8 Fraune S, Bosch TCG. Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra. *Proc Natl Acad Sci USA* 2007;104:13146–51.

9 Kremer N, Philipp EER, Carpentier MC, *et al*. Initial symbiont contact orchestrates host-organ-wide transcriptional changes that prime tissue colonization. *Cell Host Microbe* 2013;14:183–94.

10 Dooley TP, Curto EV, Reddy SP, *et al*. Regulation of gene expression in inflammatory bowel disease and correlation with IBD drugs: screening by DNA microarrays. *Inflamm Bowel Dis* 2004;10:1–14.

11 Dieckgraefe BK, Stenson WF, Korzenik JR, *et al*. Analysis of mucosal gene expression in inflammatory bowel disease by parallel oligonucleotide arrays. *Physiol Genomics* 2000;4:1–11.

12 Costello CM, Mah N, Häsler R, *et al*. Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays. *PLoS Med* 2005;2:e199.

13 Rosenstiel P, Derer S, Till A, *et al*. Systematic expression profiling of innate immune genes defines a complex pattern of immunosenescence in peripheral and intestinal leukocytes. *Genes Immun* 2008;9:103–14.

14 Francescone R, Hou V, Grivennikov SI. Cytokines, IBD, and colitis-associated cancer. *Inflamm Bowel Dis* 2015;21:409–18.

15 Natividad JM, Pinto-Sanchez MI, Galipeau HJ, *et al*. Ecobiotherapy rich in Firmicutes decreases susceptibility to colitis in a Humanized Gnotobiotic Mouse Model. *Inflamm Bowel Dis* 2015;21:1883–93.

16 Ott SJ, Schreiber S. Reduced microbial diversity in inflammatory bowel diseases. *Gut* 2006;55:1207.

17 Nishikawa J, Kudo T, Sakata S, *et al*. Diversity of mucosa-associated microbiota in active and inactive ulcerative colitis. *Scand J Gastroenterol* 2009;44:180–6.

18 Lepage P, Häsler R, Spehlmann ME, *et al*. Twin study indicates loss of interaction between microbiota and mucosa of patients with ulcerative colitis. *Gastroenterology* 2011;141:227–36.

19 Haberman Y, Tickle TL, Dexheimer PJ, *et al*. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J Clin Invest* 2014;124:3617–33.

20 Ananthakrishnan AN. Epidemiology and risk factors for IBD. *Nat Rev Gastroenterol Hepatol* 2015;12:205–17.

21 Rehman A, Sina C, Gavrilova O, *et al*. Nod2 is essential for temporal development of intestinal microbial communities. *Gut* 2011;60:1354–62.

22 Petnicki-Ocwieja T, Hrncir T, Liu YJ, *et al*. Nod2 is required for the regulation of commensal microbiota in the intestine. *Proc Natl Acad Sci USA* 2009;106:15813–18.

23 Frank DN, Robertson CE, Hamm CM, *et al*. Disease phenotype and genotype are associated with shifts in intestinal-associated microbiota in inflammatory bowel diseases. *Inflamm Bowel Dis* 2011;17:179–84.

24 Couturier-Maillard A, Secher T, Rehman A, *et al*. NOD2-mediated dysbiosis predisposes mice to transmissible colitis and colorectal cancer. *J Clin Invest* 2013;123:700–11.

25 Ramanan D, Tang MS, Bowcutt R, *et al*. Bacterial sensor Nod2 prevents inflammation of the small intestine by restricting the expansion of the commensal Bacteroides vulgatus. *Immunity* 2014;41:311–24.

26 Lappalainen T, Sammeth M, Friedländer MR, *et al*. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* 2013;501:506–11.

27 Christodoulou K, Wiskin AE, Gibson J, *et al*. Next generation exome sequencing of paediatric inflammatory bowel disease patients identifies rare and novel variants in candidate genes. *Gut* 2013;62:977–84.

28 Sterne-Weiler T, Sanford JR. Exon identity crisis: disease-causing mutations that disrupt the splicing code. *Genome Biol* 2014;15:201.

29 Mah N, Thelin A, Lu T, *et al*. A comparison of oligonucleotide and cDNA-based microarray systems. *Physiol Genomics* 2004;16:361–70.

30 Fadrosh DW, Ma B, Gajer P, *et al*. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome* 2014;2:6.

31 Breuer K, Foroushani AK, Laird MR, *et al*. InnateDB: systems biology of innate immunity and beyond—recent updates and continuing curation. *Nucleic Acids Res* 2013;41:D1228–1233.

32 Irimia M, Weatheritt RJ, Ellis JD, *et al*. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* 2014;159:1511–23.

33 Westfall PH, Young S. *Resampling-Based Multiple Testing*. New York: Wiley-Interscience, 1993.

34 Thiele I, Swainston N, Fleming RMT, *et al*. A community-driven global reconstruction of human metabolism. *Nat Biotechnol* 2013;31:419–25.

35 Schellenberger J, Que R, Fleming RMT, *et al*. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* 2011;6:1290–307.

36 Sonnenschein N, Geertz M, Muskhelishvili G, *et al*. Analog regulation of metabolic demand. *BMC Syst Biol* 2011;5:40.

37 Sonnenschein N, Golib Dzib JF, Lesne A, *et al*. A network perspective on metabolic inconsistency. *BMC Syst Biol* 2012;6:41.

38 Rosenstiel P, Sina C, Franke A, *et al*. Towards a molecular risk map—recent advances on the etiology of inflammatory bowel disease. *Semin Immunol* 2009;21:334–45.

39 Zwiers A, Fuss IJ, Leijen S, *et al*. Increased expression of the tight junction molecule claudin-18 A1 in both experimental colitis and ulcerative colitis. *Inflamm Bowel Dis* 2008;14:1652–9.

40 Danan-Gotthold M, Golan-Gerstl R, Eisenberg E, *et al*. Identification of recurrent regulated alternative splicing events across human solid tumors. *Nucleic Acids Res* 2015;43:5130–44.

41 Dhillon SS, Fattouh R, Elkadri A, *et al*. Variants in nicotinamide adenine dinucleotide phosphate oxidase complex components determine susceptibility to very early onset inflammatory bowel disease. *Gastroenterology* 2014;147:680–689.e2.

42 Roediger WE. The colonic epithelium in ulcerative colitis: an energy-deficiency disease? *Lancet Lond Engl* 1980;2:712–15.

43 Schluter J, Foster KR. The evolution of mutualism in gut microbiota via host epithelial selection. *PLoS Biol* 2012;10:e1001424.

44 Tedelind S, Westberg F, Kjerrulf M, *et al*. Anti-inflammatory properties of the short-chain fatty acids acetate and propionate: a study with relevance to inflammatory bowel disease. *World J Gastroenterol* 2007;13:2826–32.

45 Ridlon JM, Kang DJ, Hylemon PB, *et al*. Bile acids and the gut microbiome. *Curr Opin Gastroenterol* 2014;30:332–8.

46 Moffett JR, Namboodiri MA. Tryptophan and the immune response. *Immunol Cell Biol* 2003;81:247–65.

47 Gupta NK, Thaker AI, Kanuri N, *et al*. Serum analysis of tryptophan catabolism pathway: correlation with Crohn's disease activity. *Inflamm Bowel Dis* 2012;18:1214–20.

48 Filippi J, Al-Jaouni R, Wiroth JB, *et al*. Nutritional deficiencies in patients with Crohn's disease in remission. *Inflamm Bowel Dis* 2006;12:185–91.

49 Hashimoto T, Perlot T, Rehman A, *et al*. ACE2 links amino acid malnutrition to microbial ecology and intestinal inflammation. *Nature* 2012;487:477–81.