



## OPEN ACCESS

## ORIGINAL ARTICLE

## Gut mucosal virome alterations in ulcerative colitis

Tao Zuo,<sup>1,2,3</sup> Xiao-Juan Lu,<sup>4</sup> Yu Zhang,<sup>5</sup> Chun Pan Cheung,<sup>2,3</sup> Siu Lam,<sup>2,6</sup> Fen Zhang,<sup>2,3</sup> Whitney Tang,<sup>3</sup> Jessica Y L Ching,<sup>3</sup> Risheng Zhao,<sup>2,3</sup> Paul K S Chan,<sup>1,6</sup> Joseph J Y Sung,<sup>2,3</sup> Jun Yu,<sup>2,3</sup> Francis K L Chan,<sup>1</sup> Qian Cao,<sup>5,7</sup> Jian-Qiu Sheng,<sup>4</sup> Siew C Ng<sup>1,2,3</sup>

► Additional material is published online only. To view please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2018-318131>).

For numbered affiliations see end of article.

## Correspondence to

Professor Siew C Ng, Medicine and Therapeutics, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China; [siewchienng@cuhk.edu.hk](mailto:siewchienng@cuhk.edu.hk)

Received 17 December 2018

Revised 4 February 2019

Accepted 15 February 2019

## ABSTRACT

**Objective** The pathogenesis of UC relates to gut microbiota dysbiosis. We postulate that alterations in the viral community populating the intestinal mucosa play an important role in UC pathogenesis. This study aims to characterise the mucosal virome and their functions in health and UC.

**Design** Deep metagenomics sequencing of virus-like particle preparations and bacterial 16S rRNA sequencing were performed on the rectal mucosa of 167 subjects from three different geographical regions in China (UC=91; healthy controls=76). Virome and bacteriome alterations in UC mucosa were assessed and correlated with patient metadata. We applied partition around medoids clustering algorithm and classified mucosa viral communities into two clusters, referred to as mucosal virome metacommunities 1 and 2.

**Results** In UC, there was an expansion of mucosa viruses, particularly *Caudovirales* bacteriophages, and a decrease in mucosa *Caudovirales* diversity, richness and evenness compared with healthy controls. Altered mucosal virome correlated with intestinal inflammation. Interindividual dissimilarity between mucosal viromes was higher in UC than controls. *Escherichia phage* and *Enterobacteria phage* were more abundant in the mucosa of UC than controls. Compared with metacommunity 1, metacommunity 2 was predominated by UC subjects and displayed a significant loss of various viral species. Patients with UC showed substantial abrogation of diverse viral functions, whereas multiple viral functions, particularly functions of bacteriophages associated with host bacteria fitness and pathogenicity, were markedly enriched in UC mucosa. Intensive transkingdom correlations between mucosa viruses and bacteria were significantly depleted in UC.

**Conclusion** We demonstrated for the first time that UC is characterised by substantial alterations of the mucosa virobiota with functional distortion. Enrichment of *Caudovirales* bacteriophages, increased phage/bacteria virulence functions and loss of viral-bacterial correlations in the UC mucosa highlight that mucosal virome may play an important role in UC pathogenesis.

## INTRODUCTION

UC, one subtype of IBD, is a remitting and relapsing inflammatory disease affecting the entire large intestine, which usually begins in the rectum and spreads proximally.<sup>1</sup> The incidence of UC is increasing in the 21st century, especially in newly industrialised countries.<sup>2,3</sup> While the aetiology of IBD remains

## Significance of this study

## What is already known on this subject?

- Alterations in faecal bacteria and faecal virome have been reported in patients with IBD.
- Patients with UC showed an expansion of *Caudovirales* bacteriophages and *Caudovirales* species richness in the stool.

## What are the new findings?

- We demonstrated for the first time that UC is characterised by alterations in the mucosa virobiota with functional distortion.
- UC mucosa showed a high abundance of DNA viruses, particularly *Caudovirales* bacteriophages, but a low *Caudovirales* diversity, richness and evenness. Gut mucosal inflammation was associated with these alterations.
- Abundance of *Escherichia phage* and *Enterobacteria phage* was significantly higher in the mucosa of UC than controls.
- Bacteriophage functions associated with host bacteria fitness and pathogenicity were markedly enriched in UC mucosa.
- Intricate interkingdom correlations between mucosa viruses and bacteria were significantly depleted in UC.

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

- Alterations of mucosal virome in UC may contribute to disease pathogenesis.
- Therapeutic approaches involving diminishing the abundance of mucosal *Caudovirales* bacteriophages, including *Escherichia phage* and *Enterobacteria phage*, and reconstituting the gut microbial homeostasis between mucosa bacteria and viruses should be explored.

unclear, it is thought to arise from an aggravated immune response towards the gut microbiota in genetically susceptible individuals.<sup>4</sup>

The gut microbiota is important for human health<sup>5–8</sup> and has been implicated in IBD.<sup>9</sup> Compelling evidence suggests that gut microbes play a critical role in disease pathogenesis, while geographic, dietary and ethnic factors impact the microbial composition. Most microbiota studies in IBD have investigated the bacterial microbiota and



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

**To cite:** Zuo T, Lu X-J, Zhang Y, et al. Gut Epub ahead of print: [please include Day Month Year]. doi:10.1136/gutjnl-2018-318131

their relationship with the host in disease pathogenesis. With advance in high-throughput sequencing technologies, the unculturable viral component of the gut microbiota is recognised as important habitants of the gut microbial ecosystem, where the number of viral biological entities is believed to far outcompetes that of the bacterial populations.<sup>10–11</sup> The healthy human faecal virome consisted of eukaryote-infecting viruses and bacteriophages that infect bacterial cells, with a predominance of the double-strand DNA *Caudovirales*.<sup>12–15</sup> In health, gut bacteriophages show substantial diversity between individuals and are temporally stable.<sup>12–14</sup> They latently infect their bacterial hosts, coevolve with gut bacteria which facilitates drifts in the genetic diversity and infectivity of their phages,<sup>16</sup> and when under stress they generate progenies that can infect and kill other bacteria.<sup>11</sup>

To date, we know very little of the role of viruses in IBD pathogenesis. Others have reported alterations in the faecal virome with an increase of bacteriophages, particularly *Caudovirales*, in mice and humans with IBD.<sup>17–18</sup> To address whether gut viruses play an important role in IBD, understanding the composition and function of mucosal virus in health and disease is critical. In this study, we postulate that viruses populating the gut mucosal lining are significantly altered in UC and they play an important role in mucosal inflammation. Here we enriched virus-like particles (VLP) from the rectum of healthy individuals and patients with UC and performed ultradeep virome metagenomics sequencing and 16S rDNA sequencing to determine mucosal virome, bacteriome and viral-bacterial correlations in subjects with UC and healthy individuals. Sequencing analysis of such broader microbiome will elucidate related microbial changes that allow the expression of protective or causative inflammatory organisms. To the best of our knowledge, this study is the first and largest to date to characterise the mucosal virome in health and UC.

## RESULTS

### Alterations in the diversity of mucosal virome in UC

We compared the mucosal virome of 63 patients with UC and 48 healthy subjects in Hong Kong. On average,

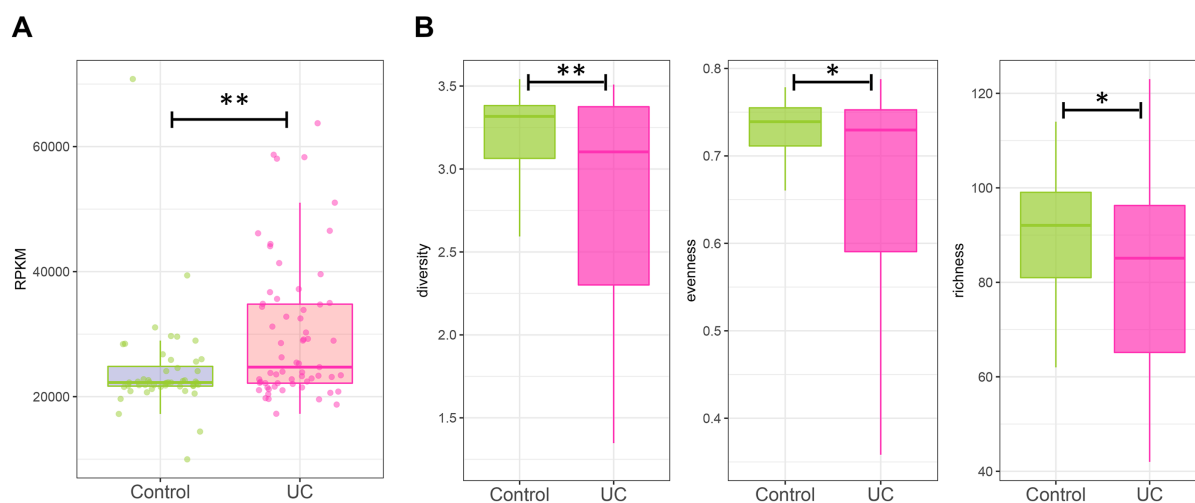
56 632 558 ± 14 330 713 clean paired-end reads were obtained from the enriched rectal VLP preparations. The mucosal virome composition was investigated at the order, genus and species levels in UC and controls. At the order level, *Caudovirales* were the predominant bacterial viruses (phages) detected in UC mucosa, while at the genus and species level, *Ascovirus* and *Streptococcus phage*, respectively, were the dominant viruses detected in UC mucosa (see online supplementary figure 1).

At the species level, UC subjects showed decreased mucosal virome diversity and richness compared with healthy subjects (t-test,  $p=0.031$  and  $0.006$ , respectively, see online supplementary figure 2). We next explored alterations in the most dominant mucosal viruses, *Caudovirales* bacteriophages, and found that *Caudovirales* abundance was significantly higher in UC (Mann-Whitney test,  $p=0.003$ , figure 1A), while species diversity, evenness and richness within the *Caudovirales* order were decreased, when compared with controls (t-test,  $p=0.009$ ,  $0.017$  and  $0.018$ , respectively, figure 1B). Collectively, these findings indicate dysbiosis of the mucosal virome in patients with UC.

Inflamed mucosa showed more enrichment of total viruses and *Caudovirales* bacteriophages than non-inflamed mucosa of patients with UC (Mann-Whitney test,  $p=0.015$  and  $0.017$ , see online supplementary figure 3A,C). Compared with healthy control mucosa, inflamed mucosa of patients with UC also showed a significant decrease in diversity, evenness and richness of total viruses and *Caudovirales* bacteriophages (Mann-Whitney test, all  $p<0.05$ , see online supplementary figure 3B,D). In contrast, there was no difference in the diversity of mucosal virome between non-inflamed mucosa of patients with UC and healthy control mucosa (see online supplementary figure 3B,D). Overall these findings suggest that gut mucosal inflammation is associated with exacerbated mucosal virome dysbiosis in UC.

### Distinct mucosal virome configurations in UC

Using non-metric multidimensional scaling (NMDS) analysis, we assessed the dissimilarity in the mucosal viral communities between UC and healthy controls. We found that mucosal viromes of UC subjects were significantly different from those



**Figure 1** Mucosal virome alterations in abundance and diversity in UC. (A) Comparison of *Caudovirales* abundance in the rectal mucosa of controls and UC subjects. The *Caudovirales* abundance was calculated as RPKM sum of *Caudovirales* contigs recruited reads normalised by sequence depth in each subject. The dots indicate individual values of the studied subjects. Statistical significance was determined by Mann-Whitney test, \* $p<0.01$ . (B) Comparison of the mucosal *Caudovirales*  $\alpha$  diversity based on Shannon diversity, evenness, Chao1 richness index in the mucosa of controls and UC subjects. Statistical significance was determined by t-test. \* $P<0.05$ ; \*\* $P<0.01$ . For box plots, the boxes extend from the first to the third quartile (25th to 75th percentiles), with the median depicted by a horizontal line. RPKM, reads per kilobase per million.

of healthy controls at the species level (based on Bray-Curtis dissimilarities, *adonis* test  $p=0.001$ , [figure 2A](#)). While disease explained 4.3% of the virome composition variance, discrepancies in virome structure between health and UC were both prominent in the two NMDS axes (t-test,  $p=0.005$  and  $0.0005$ , respectively, for comparisons at MDS1 and MDS2, [figure 2B](#)). In addition, viral community dissimilarity among patients with UC was significantly higher than that of within healthy individuals (t-test,  $p<2.2e-16$ , [figure 2C](#)). These data suggest that virome dysbiosis in UC is highly individualised.

We next performed *DESeq* analysis to identify differentially present taxa between UC and controls, at family, genus and species levels. More taxa were found to be enriched in UC than in control mucosa ([figure 3](#)). At the family level, Microviridae (single-stranded DNA phage), Myoviridae, Podoviridae (double-stranded DNA phages from the *Caudovirales* order) and Pneumoviridae (eukaryotic virus) were more abundant in UC than in controls, whereas Anelloviridae (eukaryotic virus) were higher in controls than in UC ([figure 3A](#), see online supplementary figure 4). At the genus level, *Phix174microvirus*, *P1virus*, *Lambdavirus*, *T4virus*, *P22virus* (all *Caudovirales* bacteriophages) and *Orthopneumovirus* were enriched in UC, whereas giant viruses *Coccolithovirus*, *Minivirus* and vertebrate-infecting virus *Orthopoxvirus* (all eukaryotic viruses) were enriched in controls ([figure 3B](#), see online supplementary figure 4). Consistent with these findings, there was higher abundance of *Caudovirales* bacteriophage species, including *Escherichia phage* and *Enterobacteria phage* in UC relative to control mucosa ([figure 3C](#)). Altogether, these findings indicate that dysbiosis in prokaryotic viruses, particularly *Caudovirales* bacteriophages, is prominent in UC mucosa, while some eukaryotic viruses are more abundant in healthy subjects, suggesting an association between bacteria dysbiosis and bacteriophage expansion in UC.

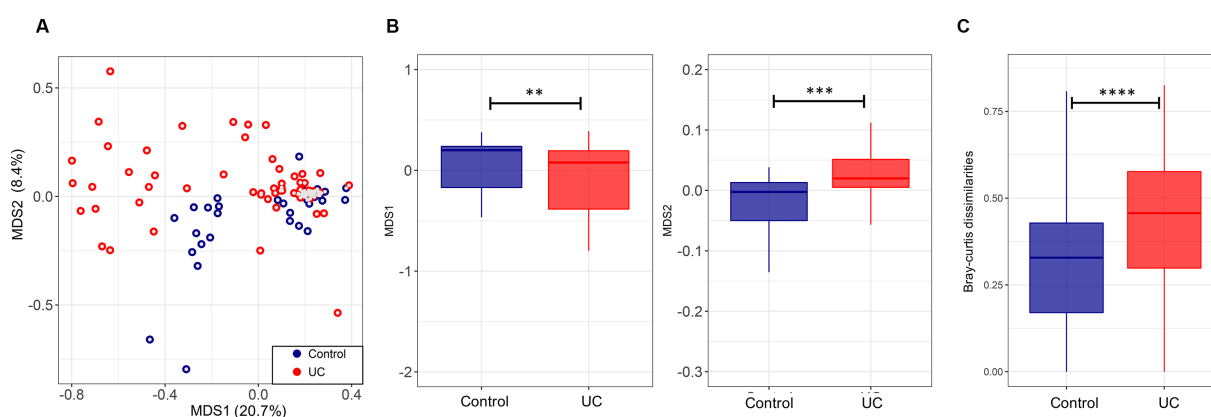
Recent studies have demonstrated geographical effects on the gut microbiome composition.<sup>19,20</sup> We included 56 additional ethnically Chinese subjects from two different regions in China (Beijing,  $n=40$ ; Xiangshan, Zhejiang province,  $n=16$ ; see online supplementary figure 5A), to investigate differences between mucosal viromes of healthy controls and UC subjects. Among the disparate mucosal virus species seen between UC and healthy subjects from the Hong Kong cohort, we were able to replicate two virus species in the additional cohort,

*Chrysochromulina ericina virus* and *Mimivirus* (see online supplementary figure 5B,C), suggesting a geographical effect on the mucosal virome variation. The overpresence of algae-infecting virus *C. ericina virus* in UC may indicate a detrimental effect of alga-related diet on IBD. Meanwhile, the high presence of the giant virus *Mimivirus* in UC, a potential pneumonia-causing agent,<sup>21,22</sup> is in line with the observation that patients with IBD may be at an increased risk of opportunistic infections including pneumonia.<sup>23</sup>

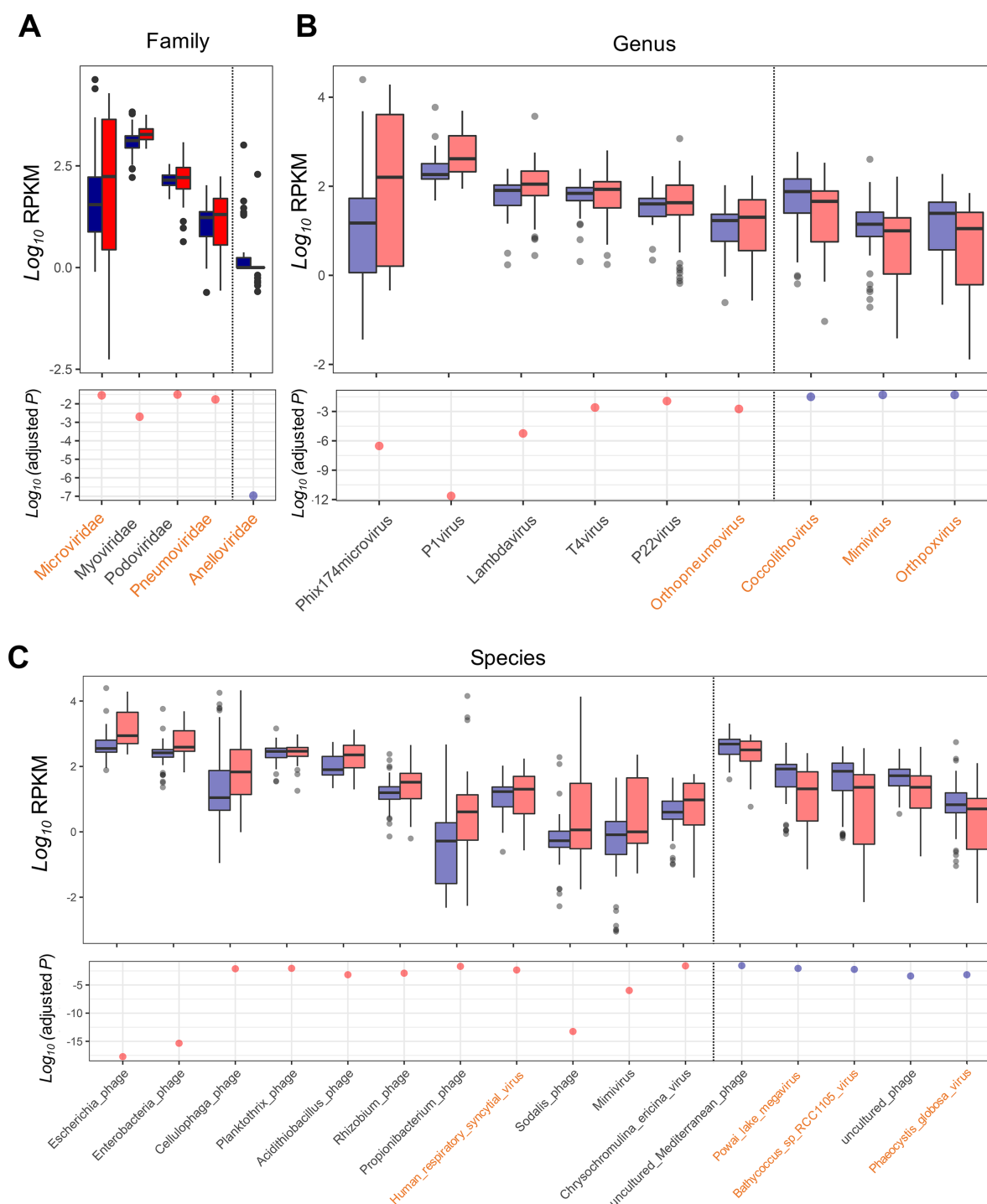
### Mucosal virome metacommunities in health and UC

Studies have shown bacterial microbiomes clustering in relation to IBD.<sup>24,25</sup> With the observed high degree of dissimilarity among mucosal viral communities, we investigated clustering of mucosal viromes in health and UC. By applying partition around medoids clustering algorithm on the species abundance profile, we found that the viral communities converged into two clusters (referred to as mucosal virome metacommunities hereafter, [figure 4A](#)). Among the 85 metacommunity 1 clustered subjects, 40 (47%) were patients with UC. In contrast, among the 26 metacommunity 2 clustered subjects, 23 (88%) were patients with UC ([figure 4A](#)), indicating a significant association between patients with UC and metacommunity 2 virome configuration ( $\chi^2$  test,  $p=0.0002$ ). We then implemented *DESeq* and *Random Forest* analyses to delineate the viral species discriminating these two metacommunities. Forty-six species were discerned between metacommunity 1 and metacommunity 2 viromes, among which 20 species were highly present, whereas 26 were significantly depleted in metacommunity 2 ([figure 4B](#)). In metacommunity 2, *Escherichia phage*, *Enterobacteria phage* and *uncultured Caudovirales phage* were the most abundant distinctive species with large effect sizes.

To explore the differences in the mucosal virome composition between healthy individuals and different UC metacommunities, we quantified the presence ratio of viral core species, common species and unique species<sup>26</sup> (corresponding to viral species shared among  $>50\%$ ,  $20\%$ – $50\%$  and  $<20\%$  of the individuals, respectively) in all study subjects. While no significant differences were seen in the ratio of core species, common species or unique species between healthy and UC metacommunity 1 individuals, the core species accounted for a significant higher proportion



**Figure 2** Between-group and within-group viral community structure difference in health and UC. (A) Mucosal viral community structure difference between controls and UC by NMDS (non-metric multidimensional scaling) plotting based on Bray-Curtis dissimilarities at the viral species level. (B) Between-group viral community structure difference reflected on each axis of NMDS plotting. (C) Comparison of within-group mucosal virome Bray-Curtis dissimilarities between healthy individuals and UC subjects. Statistical significance was determined by t-test. \*\* $P<0.01$ ; \*\*\* $P<0.001$ ; \*\*\*\* $P<0.0001$ . For box plots, the boxes extend from the first to the third quartile (25th to 75th percentiles), with the median depicted by a horizontal line.



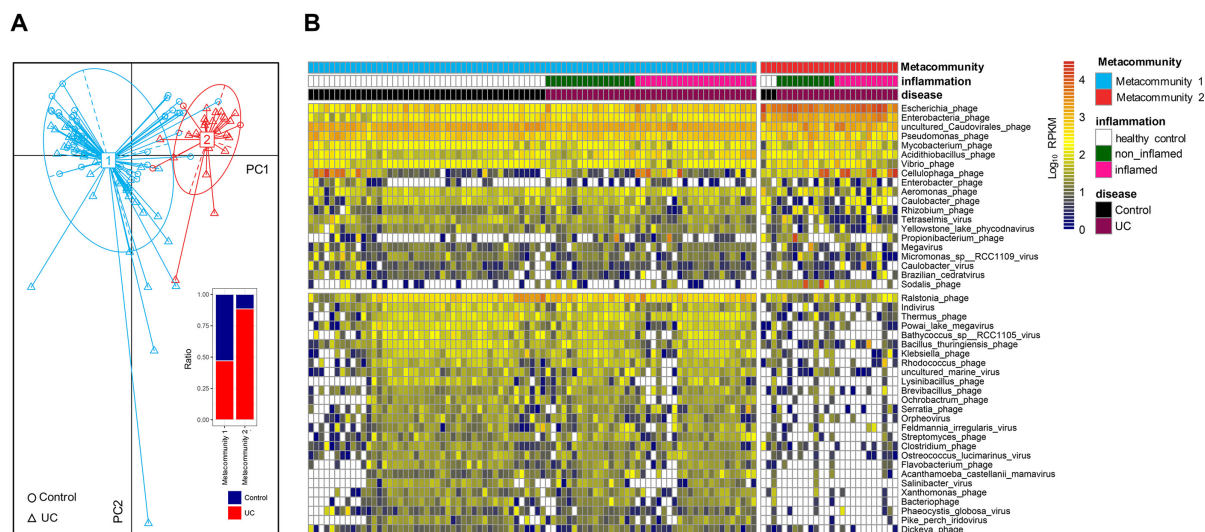
**Figure 3** Differential viral taxa between health and UC mucosa at the family, genus and species levels. Differentially enriched viral families (A), genera (B) and species (C) between health and UC mucosa were determined by *DESeq* analysis with false discovery rate (FDR) correction (only those differential taxa with adjusted  $p < 0.05$  and  $|\log_2(\text{between-group fold change})| > 2$  are shown). For viral taxon names, taxa colour coated by black denote prokaryotic viruses, while those colour coated by orange denote eukaryotic viruses. For viral abundance box plots, the boxes extend from the first to the third quartile (25th to 75th percentiles), with the median depicted by a horizontal line. RPKM, reads per kilobase per million mapped reads.

of the constituent viral species in UC metacommunity 2 individuals than controls and UC metacommunity 1 individuals (all  $p < 0.0001$ , one-way analysis of variance [ANOVA], [figure 5A,B](#)). However, both common and unique species proportion were diminished in UC metacommunity individuals compared with controls ( $p < 0.0001$  and  $p < 0.01$ , respectively, one-way ANOVA, [figure 5A,B](#)). We then assessed the abundance of these viral

species, and found that while several top core species were more abundant in UC metacommunity 2 individuals, common species and unique species were decreased or abrogated in UC metacommunity 2 individuals, compared with healthy controls and UC metacommunity 1 individuals ([figure 5c](#)).

Exploiting *Random Forest*, we performed classification on the mucosal viromes of the subjects from Beijing and Xiangshan

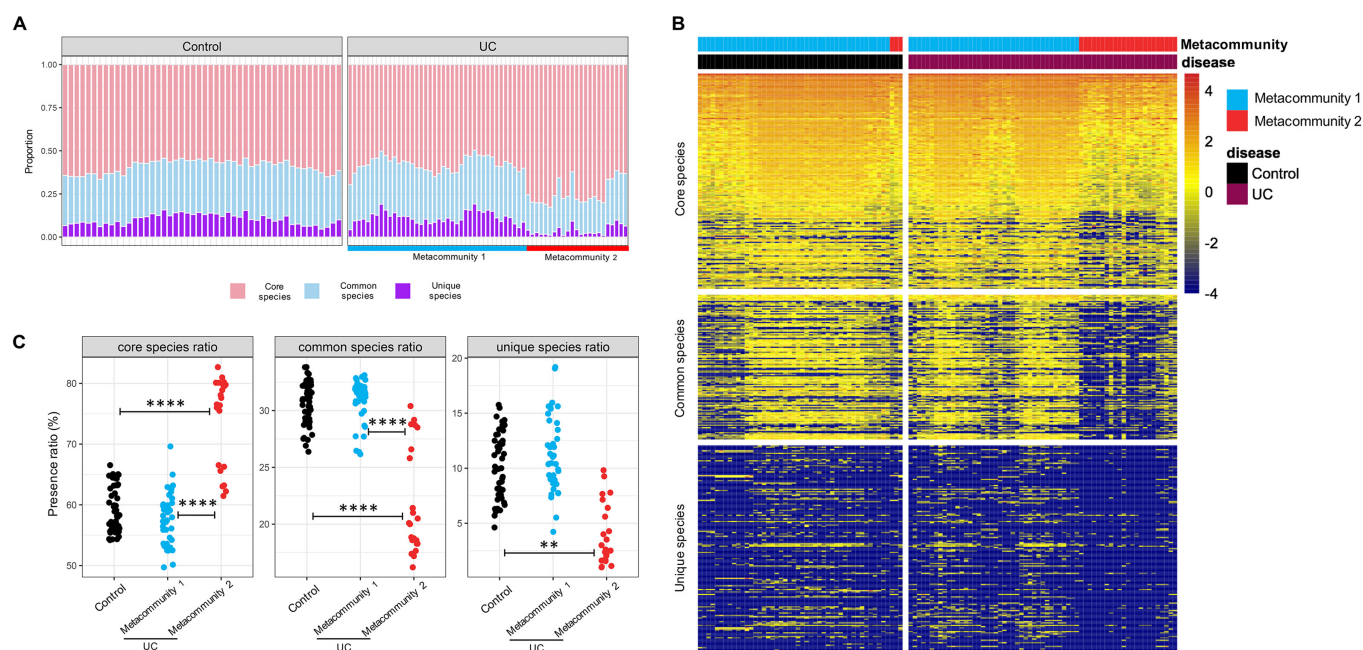




**Figure 4** Mucosal virome metacommunities in health and UC. (A) Virome metacommunity clustering based on partition around medoids (PAM) algorithm and principal coordinates analysis (PCoA) on the viral community structures of healthy individuals and UC mucosa. The inset shows the ratio of healthy individuals and UC subjects within each metacommunity population. (B) Heatmap of the presence of differential viral species contributing to clustering of the two mucosal virome metacommunities. Discriminative species were identified by concordant *DESeq* and *Random Forest* analyses. Viral species abundances are colour intensified according to  $\log_{10}$  RPKM values. Only those species concordantly determined by *DESeq* and *Random Forest* algorithm with effect size more than 2 and FDR-adjusted  $p < 0.05$  are shown. RPKM, reads per kilobase per million mapped reads.

cohorts. All subjects were classified as metacommunity 1. Differential analysis on all metacommunity 1 subjects between healthy controls and UC in Hong Kong cohort identified a panel of distinctive virus species (see online supplementary figure 6A). However, only two species were replicated, algae-infecting *Feldmannia species virus*, significantly higher in UC metacommunity

1 subjects relative to control metacommunity 1 subjects in Xiangshan cohort, and *Pseudomonas virus*, significantly higher in control metacommunity 1 subjects relative to UC metacommunity 1 subjects in Beijing cohort (see online supplementary figure 6B). Altogether, these data highlight a geographical effect on the mucosal virome structure, which may lead to intercohort



**Figure 5** Significant altered presence of core, common and unique viral species in different UC metacommunities. (A) The proportion of core, common and unique viral species in each subject. The core species, common species and unique species correspond to viral species shared among  $>50\%$ ,  $20\%–50\%$  and  $<20\%$  of studied subjects, respectively. (B) Quantification of the presence ratio of core, common and unique viral species in the viral communities of healthy controls and UC subjects with two respective mucosal virome metacommunities. Statistical significance was determined by one-way analysis of variance (ANOVA).  $**P < 0.01$ ;  $****P < 0.0001$ . (C) Heatmap of the abundances of the most abundant 100 core, common and unique species in healthy controls and UC subjects with two respective mucosal virome metacommunities. Viral species abundances are colour intensified according to  $\log_2$  reads per kilobase per million (RPKM) values.

variations in mucosal virome composition and likely relates to external factors including diet and living environment.

### Functional alterations of mucosal virome in UC

To assess functions of mucosal virome, we performed HUMANN2 analysis on the viral sequences against Gene ontology (GO) and Pfam protein family databases. Due to the largely underexplored nature of gut mucosal virome functions, we were only able to characterise the function of a small proportion of the viral reads. Among the annotated functions, integral component of membrane, DNA binding (GO terms), ATP-binding cassette (ABC) transporter and integrase core domain (Pfam proteins) were the prominent proteins/functions (see online supplementary figure 7), suggesting an active role for viruses in virus-host biology and significant interplay between gut viruses and bacteria.

We then compared the predicted mucosal virome functions between health and UC. Though the most abundant functions did not differ in abundance between health and UC, healthy individuals showed a richer virome function, in terms of GO and Pfam protein functions, whereas UC subjects exhibited a significant abolishment of functions (figure 6). Nonetheless, several molecular functions were determined to be more abundant in UC than in healthy mucosal virome, including DNA template negative regulation of transcription, beta-lactamase, glutamine amidotransferase, glycosyl hydrolases, type II/IV secretion system and multicopper oxidase, all of which were linked to phage lysis of bacteria host as well as functions of bacteria origin. This result implies that the enriched mucosal viral functions in UC were associated with bacteria fitness, pathogenicity and antibiotics resistance.

### Mucosal bacteriome alterations and transkingdom network between virome and bacteriome at the mucosa

We further assessed the bacterial microbiome alteration in UC mucosa compared with healthy control mucosa. Bacterial diversity and richness were significantly increased (t-test,  $p=0.024$  and  $0.007$ , respectively, see online supplementary figure 8A), indicating an expansion of diverse bacteria in the mucosa of patients with UC. UC mucosal bacteriome showed distinctive structure to that of controls, at the phylum, family and genus levels (see online supplementary figure 8B–D). Quantitative differential analysis identified a number of bacterial taxa distinguished between health and UC mucosa (see online supplementary figure 8E). Among them, Firmicutes (phylum), Pseudomonadaceae, Ruminococcaceae, Thermaceae, unclassified Clostridiales, Veillonellaceae (family) and *Coprococcus* (genus) were significantly enriched in UC than in controls.

To characterise the relationship between bacteriome and virome in the mucosa, we evaluated the correlation between the  $\alpha$  diversity (diversity, evenness and richness) of the bacteriome and virome. In controls, there were significant correlations between intrakingdom  $\alpha$  diversity metrics (figure 7A). However, correlations of within-bacteria-kingdom  $\alpha$  diversity seen in controls were lost in UC (figure 7A), suggesting a more dysbiotic state of the bacterial microbiota than the virobiota in UC mucosa. We then assessed the correlation between bacterial genera and viral species in control and UC mucosa. The alterations of viral-bacterial correlations in UC could be ascribed to both losses of viral-bacterial associations and appearances of novel viral-bacterial associations. A larger number of bacterial-viral correlations were seen in UC than in controls (figure 7B). Among these bacterial-viral correlations, most were negative

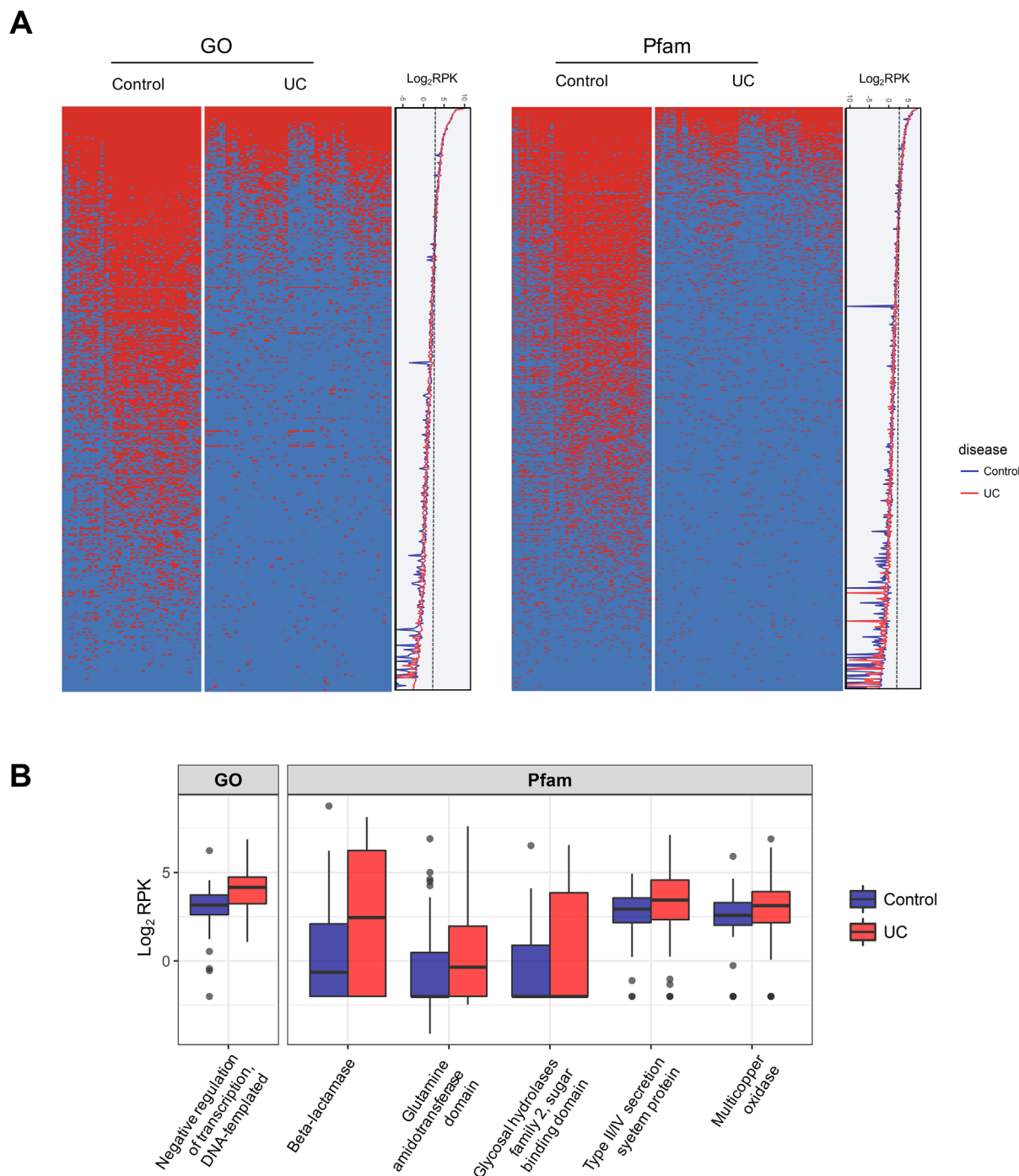
correlations. However, a substantial contraction in the number of significant bacterial-viral correlation was observed in UC mucosal samples, meanwhile the correlations in UC were weakened compared with that in controls (figure 7B). In particular, the correlations between the virus species *Cellulophaga phage*, *Bacillus phage*, *Burkholderia phage*, *Acidithiobacillus phage* (all *Caudovirales* bacteriophages), *Hokovirus* and mucosal bacteria were extensively intensive in controls, whereas they were significantly dampened in UC. This corresponded with an expansion in the abundance of *Cellulophaga phage* and *Acidithiobacillus phage* in patients with UC, along with a depletion of correlations between these two species and mucosal bacteria. These results suggest an altered virobiota–bacterial microbiota relationship in UC, where viruses and bacteria became less intertwined and more specialised, further implicating the importance of interkingdom equilibrium among mucosal viruses and bacteria for human health.

### DISCUSSION

To our knowledge, this study represents the first and most in-depth human mucosal virobiota study in health and UC using a dedicated metagenomics approach based on enriched virome preparation. We showed that patients with UC exhibited a dysbiotic mucosal virobiota, characterised by an increase in virus abundance, particularly *Caudovirales* bacteriophages, and a decreased viral diversity, richness and evenness. Given the underexplored nature of virome for which the viruses database remains small and incomplete, unknown viruses can influence the estimation of diversity, limiting our effort to disentangle mucosal viromes in health and UC. We also observed significant functional alterations in mucosal virome in UC, which were associated with bacteriophage lytic cycle, and bacteria host features, including bacteria adaption, antibiotics resistance and pathogenicity. Inflammation was associated with shifts in virome composition and alga-infecting viruses were more abundant in UC than controls. We hypothesise that host inflammation and environmental factors such as medication or diet may be significant contributors to the altered mucosal viromes, which are intertwined with the bacterial microbiome and host immune responses.

In mice, it has been shown that augmented bacteriophage populations during colitis were associated with intestinal inflammation in the pathobiont host,<sup>18</sup> and inflammation can boost free phage production and subsequent transfer in mice.<sup>27</sup> In humans, *Caudovirales* were shown to be present in high abundance in patients with IBD<sup>17</sup> and *Clostridium difficile* infection.<sup>28</sup> In another study, a healthy core phageome was identified in healthy individuals and decreased numbers of these core phages in the faeces were reported in patients with UC.<sup>26</sup> More recently, an increased faecal eukaryotic virome richness has been described in patients with UC from a Belgian cohort.<sup>29</sup> Increased abundance of *Escherichia phage* and *Enterobacteria phage* and decreased core viral species numbers in UC mucosa in this study are consistent with findings from the faeces of colitic mice<sup>18</sup> and patients with IBD.<sup>17</sup> In this cross-sectional study, we were unable to discern whether alterations in the virome are the cause or consequence of mucosal inflammation and denudation of the epithelial barrier. Future animal studies will be necessary to investigate this further.

A marginal inverse correlation between *Escherichia phage* and *Escherichia* genus bacteria in UC mucosa was observed (figure 7B). The enhanced *Escherichia phage* abundance and antagonistic relationship between *Escherichia phage* and the



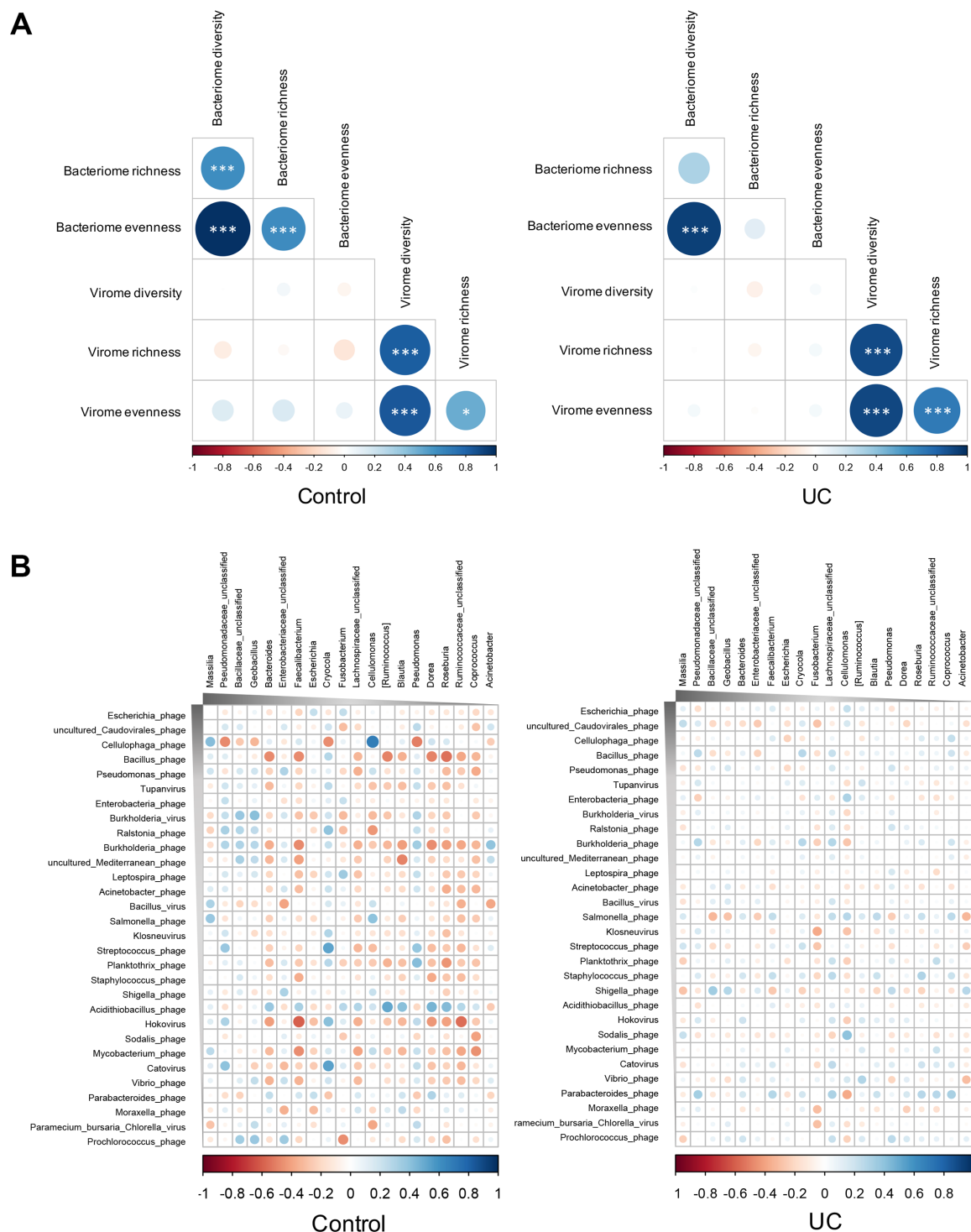
**Figure 6** Significant loss of diverse viral functions in UC mucosa with concomitant increases in bacteria-pathogenicity associated functions. (A) Presence-absence heatmap of the classified viral functions in controls and UC. Viral functions were predicted and classified via HUMANN2 pipeline, exploiting the sophisticated Gene ontology (GO) and Pfam protein family databases. Functions with reads per kilobase (RPK) >10 were considered present in individuals. The abundance distribution of classified viral functions is plotted in the line chart, with abundance values expressed as  $\text{Log}_2\text{RPK}$ . (B) Differentially enriched viral functions between health and UC mucosa. Differential viral functions were determined by *DESeq* analysis with FDR correction. Only those functions with adjusted  $p < 0.05$  and  $|\text{Log}_2(\text{between-group fold change})| > 2$  are shown. For box plots, the boxes extend from the first to the third quartile (25th to 75th percentiles), with the median depicted by a horizontal line.

bacteria genus *Escherichia* in UC mucosa, shown in our study, along with reported expansion in both *Escherichia phage* and *Escherichia* in UC faeces by others,<sup>17,18</sup> may indicate that *Escherichia phage* contention at the mucosa is linked to coexpansion of both in the faeces, suggesting prophage induction and complex bacteriophage-bacteria dynamics in the gut. It is postulated that apart from the empirical ‘predator-prey’ phage-bacteria

dynamics, there are also ‘kill-the-winner’ and ‘arms-race’ phage-bacteria dynamics at play in the gut milieu where coexpansion and coevolution of phages and bacteria are expected, particularly when host is under stress.<sup>30</sup>

Alterations of bacteriophage composition in the gut mucosa can influence the bacterial microbiota ecology. Bacteriophages are primary drivers of bacterial fitness and diversity.<sup>31</sup> Lysis of





**Figure 7** Mucosal bacterial-viral correlation patterns in health and UC. (A) Correlations between the  $\alpha$  diversity (diversity, evenness and richness) of mucosal bacteria and viruses in healthy controls and UC, respectively. Spearman's correlation coefficient was calculated, while statistical significance was determined for all pairwise comparisons. Significant correlations ( $p < 0.05$ ) are displayed with asterisk. \* $P < 0.05$ ; \*\*\* $P < 0.001$ . (B) Correlations between the most abundant 30 virus species and the most abundant 20 bacteria genera in health and UC mucosa. Spearman's correlation coefficient was calculated, while statistical significance was determined for all pairwise comparisons. Only statistically significant correlations were plotted, where blue circles indicate positive correlations and red circles indicate inverse correlations. The size and shading indicate the magnitude of the correlation where darker shades denote more intensive correlations than light ones.

bacteria could result in a release of nucleic acids, proteins and lipids that serve as pathogen-associated molecular patterns that trigger inflammatory responses to induce cellular infiltration, cytokine production and even tissue damage. In the GI tract, bacteriophages are responsible for horizontal transfer of genetic

elements among bacterial populations, including those for antibiotic resistance and disease pathogenesis.<sup>32–35</sup>

The human virome evolves rapidly due to varying conditions and stressors conferred by the host and the environment. A multitude of factors contributing to the hypervariation in virome relative to the



bacteriome among populations have been reported.<sup>12–36–39</sup> Disaccordance in discriminative viral species identified between Hong Kong cohort and other China cohorts implies high degree of variation in mucosal virome configuration, likely a region-population-specific effect on the virome structure. However, our study is limited by a modest sample size of the China cohort and future large population-based studies will be important to confirm universal and population-specific virome in health and IBD.

Although diverse bacteriophages were present in high abundance in UC, some eukaryotic viruses were more abundant in healthy individuals. Chronic infections with viruses are common in humans.<sup>11</sup> The history of infection with acute and chronic viruses contributes to host immunity education and maturation,<sup>40–42</sup> especially when it occurs in early life.<sup>2–43</sup> Viral pathogen-driven selection assists in shaping the immune system, improving immune tolerance and memory, thereby preventing against autoimmune disease development.<sup>41–44–46</sup>

Microbiota composition has been reported to differ between faecal and mucosal samples. Most IBD microbiome studies have focused on faecal samples. There has been controversy over the degree of dysbiosis in IBD, particularly at the mucosal level.<sup>47</sup> Studies reporting changes in the diversity, richness and composition of the bacteriome in UC mucosa have produced inconsistent results. Multiple studies have shown decreased bacterial diversity and richness in IBD mucosa,<sup>48–49</sup> several other studies however indicated similar levels of bacterial diversity and richness between UC and control mucosa<sup>50–51</sup> while some on the contrary showed increased bacterial richness in the non-inflamed tissues from IBD subjects compared with healthy subject tissues.<sup>52</sup> Though some studies reported overrepresentation of Proteobacteria in UC mucosa compared with healthy mucosa,<sup>48–53</sup> some reported insignificant difference.<sup>50</sup> Additionally, while some studies reported decrease in Firmicutes in UC mucosa compared with healthy mucosa,<sup>48–49</sup> some other groups showed higher abundance of Firmicutes in UC mucosa than healthy mucosa,<sup>50–53–54</sup> the latter findings are consistent with our data. The heterogeneity observed is likely to be multifactorial including the site of sampling, medication use, disease phenotype, active versus inactive disease, oxidative stress level as well as differences in population, geography and diet, all of which have been documented to influence mucosal bacteria composition. Further large-scale in-depth investigations and validation are necessary.

In conclusion, our findings highlight the importance of mucosal virome in the pathogenesis of UC. A healthy mucosal virome is characterised by a low abundant but diversified viral community, which rests in homeostasis with the host and other microbial communities. Alterations in the mucosal virome in UC may contribute to disease pathogenesis and restoration of the mucosa viral community represents promising therapeutics for UC.

## METHODS

### Cohort description and study subjects

A clinical cohort of 63 patients with UC versus 48 healthy subjects from Hong Kong was included in this study. In addition, 20 healthy subjects versus 20 patients with UC from Beijing, and eight healthy subjects versus eight patients with UC from Xiangshan, Zhejiang Province (all Chinese), were also enrolled into this study, with informed consent. Patient inclusion criteria include subjects aged  $\geq 18$  with a diagnosis of UC defined by endoscopy, radiology and histology. Consecutive patients with UC who consented to have colonic samples taken during surveillance colonoscopy or colonoscopy as part of assessment for disease activity were included. Tissue biopsies were obtained from the rectum. Majority of patients were in clinical remission. Controls comprise individuals undergoing

colonoscopy for screening for polyp, colorectal cancer, or symptoms of rectal bleeding due to haemorrhoids, and friends and spouses or partners of patients at local hospitals, or any individuals who were interested to participate in this study. Consecutive healthy individuals underwent screening colonoscopy and were found to have a normal macroscopic colon and no evidence of microscopic inflammation after colonoscopy.

Rectal biopsies from the study subjects were collected via endoscopy and then stored at  $-80^{\circ}\text{C}$  for downstream mucosal virome and bacterium analysis. Patient metadata were shown in online supplementary table 1.

### VLP enrichment and sequencing

VLPs were enriched from the rectal biopsies of patients with UC and healthy subjects, using a modified protocol according to previously described methods.<sup>17–28–32</sup> The detailed procedure is included in the online supplementary appendix.

VLP DNAs were quantified (Nanodrop) and  $1\ \mu\text{g}$  of DNA was randomly fragmented by ultrasonication (Covaris) followed by library construction. The qualified libraries were amplified on cBot to generate the cluster on the flow cell (TruSeq PE Cluster Kit V3-cBot-HS, Illumina). The amplified flow cell was sequenced pair end on the HiSeq Xten System (TruSeq SBS KIT-HS V3, Illumina) (BGI, Shenzhen, China; standard  $2\times 150\text{ bp}$  run), generating 20–60 million raw sequences (5–8G raw data) per sample. The detailed sequence statistics was included in the online supplementary table 2. Sequence processing and quality control, de novo contig assembly and taxonomy annotation were included in the online supplementary appendix.

### Virome data analysis

To estimate contig abundance and calculate sequence diversity, all reads were aligned to the resulting curated contigs using Bowtie2 (V.2.2.9).<sup>55</sup> The mapped sequence counts, contig lengths and total sequence counts were used to normalise the sequence counts and represent the reads per kilobase per million of each sample to the contigs. These values were used to generate viral abundance tables at various taxonomy levels.

The virome abundance data were imported into R V.3.2.3. Diversity, evenness and richness calculation were performed using *phyloseq* in R. Spearman's correlation and their significance were calculated using the *cor* and *cor.test* functions in R, respectively. For the viruses–bacteria correlations in  $\alpha$  diversity and taxa abundance, Spearman's correlations were calculated. Correlation plots were generated using the *corrplot* R package. Heatmaps were generated using the *pheatmap* R package. *DESeq*, *Random Forest* and *LEfSe* linear discriminant analysis were included in the online supplementary appendix.

### Accession codes

Sequence data have been deposited to the NCBI Sequence Read Archive under BioProject accession numbers PRJNA504921 and PRJNA506811.

### Author affiliations

<sup>1</sup>Center for Gut Microbiota Research, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

<sup>2</sup>State Key Laboratory for digestive disease, Institute of Digestive Disease, Li Ka Shing Institute of Health Science, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

<sup>3</sup>Department of Medicine and Therapeutics, Faculty of Medicine, The Chinese University of Hong Kong, Shatin, Hong Kong, China

<sup>4</sup>Department of Gastroenterology, The General Hospital of the People's Liberation Army, Beijing, China

<sup>5</sup>Faculty of Medicine, Zhejiang University, Hangzhou, China

<sup>6</sup>Department of Microbiology, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China

<sup>7</sup>Department of Gastroenterology, School of Medicine, Sir Run Run Shaw Hospital, Zhejiang University, Hangzhou, China

**Contributors** TZ performed the experiments and metagenomics analysis and drafted the manuscript. XJL, YZ and WT performed study subject inclusion and collection of specimens, helped in virome preparation and analysis. CPC, SL, FZ and RZ collected the clinical samples and data, helped in virus DNA extraction and provided significant intellectual contribution. WT and JYLC managed the clinical practice and clinical database. PC, JJYS, JY, FKLC, QC and JQS provided intellectual contribution and critical comments on the manuscript. SCN designed and supervised the study and revised the manuscript.

**Funding** PROCORE-France/Hong Kong Joint Research Scheme (F-CUHK402/15); ANR/RGC Joint Research Scheme (A-CUHK402/17); the ENIGMA study supported by Helmsley Charitable Trust; and a seed fund for Gut Microbiome Research provided by the Gut Microbiota Research Center, Faculty of Medicine (The Chinese University of Hong Kong).

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Ethics approval** The study was approved by The Joint Chinese University of Hong Kong, New Territories East Cluster Clinical Research Ethics Committee (The Joint CUHK-NTEC CREC, CREC Ref No: 2014.026).

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

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

## REFERENCES

- Kaistha A, Levine J. Inflammatory bowel disease: the classic gastrointestinal autoimmune disease. *Curr Probl Pediatr Adolesc Health Care* 2014;44:328–34.
- Zuo T, Kamm MA, Colombel JF, et al. Urbanization and the gut microbiota in health and inflammatory bowel disease. *Nat Rev Gastroenterol Hepatol* 2018;15:440–52.
- Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 2018;390.
- Podolsky DK. Inflammatory bowel disease (1). *N Engl J Med* 1991;325:928–37.
- Subramanian S, Blanton LV, Frese SA, et al. Cultivating healthy growth and nutrition through the gut microbiota. *Cell* 2015;161:36–48.
- Zhao L, Zhang F, Ding X, et al. Gut bacteria selectively promoted by dietary fibers alleviate type 2 diabetes. *Science* 2018;359:1151–6.
- Marchesi JR, Adams DH, Fava F, et al. The gut microbiota and host health: a new clinical frontier. *Gut* 2016;65:330–9.
- Subramanian S, Huq S, Yatsunenko T, et al. Persistent gut microbiota immaturity in malnourished Bangladeshi children. *Nature* 2014;510:417–21.
- Zuo T, Ng SC. The Gut Microbiota in the Pathogenesis and Therapeutics of Inflammatory Bowel Disease. *Front Microbiol* 2018;9.
- Cadwell K. The virome in host health and disease. *Immunity* 2015;42:805–13.
- Virgin HW. The virome in mammalian physiology and disease. *Cell* 2014;157:142–50.
- Minot S, Bryson A, Chehoud C, et al. Rapid evolution of the human gut virome. *Proc Natl Acad Sci U S A* 2013;110:12450–5.
- Minot S, Grunberg S, Wu GD, Gd W, et al. Hypervariable loci in the human gut virome. *Proc Natl Acad Sci U S A* 2012;109:3962–6.
- Reyes A, Haynes M, Hanson N, et al. Viruses in the faecal microbiota of monozygotic twins and their mothers. *Nature* 2010;466:334–8.
- Waller AS, Yamada T, Kristensen DM, et al. Classification and quantification of bacteriophage taxa in human gut metagenomes. *ISME J* 2014;8:1391–402.
- De Sordi L, Khanna V, Debarbieux L. The gut microbiota facilitates drifts in the genetic diversity and infectivity of bacterial viruses. *Cell Host Microbe* 2017;22:801–8.
- Norman JM, Handley SA, Baldridge MT, et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* 2015;160:447–60.
- Duerkop BA, Kleiner M, Paez-Espino D, et al. Murine colitis reveals a disease-associated bacteriophage community. *Nat Microbiol* 2018;3:1023–31.
- Yatsunenko T, Rey FE, Manary MJ, et al. Human gut microbiome viewed across age and geography. *Nature* 2012;486:222–7.
- He Y, Wu W, Zheng HM, et al. Regional variation limits applications of healthy gut microbiome reference ranges and disease models. *Nat Med* 2018;24:1532–5.
- Wessner DR. Discovery of the giant mimivirus. *Nature Education* 2010;3:61.
- La Scola B, Marrie TJ, Auffray JP, et al. Mimivirus in pneumonia patients. *Emerg Infect Dis* 2005;11:449–52.
- Long MD, Martin C, Sandler RS, et al. Increased risk of pneumonia among patients with inflammatory bowel disease. *Am J Gastroenterol* 2013;108:240–8.
- Arumugam M, Raes J, Pelletier E, et al. Enterotypes of the human gut microbiome. *Nature* 2011;473:174–80.
- Costea PI, Hildebrand F, Arumugam M, et al. Enterotypes in the landscape of gut microbial community composition. *Nat Microbiol* 2018;3:8–16.
- Manrique P, Bolduc B, Walk ST, et al. Healthy human gut phageome. *Proc Natl Acad Sci U S A* 2016;113:10400–5.
- Diard M, Bakkeren E, Cornuault JK, et al. Inflammation boosts bacteriophage transfer between *Salmonella* spp. *Science* 2017;355:1211–5.
- Zuo T, Wong SH, Lam K, et al. Bacteriophage transfer during faecal microbiota transplantation in *Clostridium difficile* infection is associated with treatment outcome. *Gut* 2018;67:634–43.
- Conceição-Neto N, Deboutte W, Dierckx T, et al. Low eukaryotic viral richness is associated with faecal microbiota transplantation success in patients with UC. *Gut* 2018;67:1558–9.
- Mirzaei MK, Maurice CF. Ménage à trois in the human gut: interactions between host, bacteria and phages. *Nat Rev Microbiol* 2017;15:397–408.
- Brüssow H, Canchaya C, Hardt WD. Phages and the evolution of bacterial pathogens: from genomic rearrangements to lysogenic conversion. *Microbiol Mol Biol Rev* 2004;68:560–602.
- Reyes A, Wu M, McNulty NP, et al. Gnotobiotic mouse model of phage-bacterial host dynamics in the human gut. *Proc Natl Acad Sci U S A* 2013;110:20236–41.
- Zhang Y, Leleune JT. Transduction of bla(CMY-2), tet(A), and tet(B) from *Salmonella enterica* subspecies *enterica* serovar Heidelberg to *S. Typhimurium*. *Vet Microbiol* 2008;129:418–25.
- Maïques E, Ubeda C, Campoy S, et al. beta-lactam antibiotics induce the SOS response and horizontal transfer of virulence factors in *Staphylococcus aureus*. *J Bacteriol* 2006;188:2726–9.
- Jerlström Hultqvist J, Warsi O, Söderholm A, et al. A bacteriophage enzyme induces bacterial metabolic perturbation that confers a novel promiscuous function. *Nat Ecol Evol* 2018;2:1321–30.
- Kim MS, Bae JW. Spatial disturbances in altered mucosal and luminal gut viromes of diet-induced obese mice. *Environ Microbiol* 2016;18:1498–510.
- Minot S, Sinha R, Chen J, et al. The human gut virome: inter-individual variation and dynamic response to diet. *Genome Res* 2011;21:1616–25.
- Lim ES, Zhou Y, Zhao G, et al. Early life dynamics of the human gut virome and bacterial microbiome in infants. *Nat Med* 2015;21:1228–34.
- Ogilvie LA, Jones BV. The human gut virome: a multifaceted majority. *Front Microbiol* 2015;6.
- Reese TA, Bi K, Kambal A, et al. Sequential infection with common pathogens promotes human-like immune gene expression and altered vaccine response. *Cell Host Microbe* 2016;19:713–9.
- Tao L, Reese TA. Making mouse models that reflect human immune responses. *Trends Immunol* 2017;38:181–93.
- Beura LK, Hamilton SE, Bi K, et al. Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature* 2016;532:512–16.
- Gensollen T, Iyer SS, Kasper DL, et al. How colonization by microbiota in early life shapes the immune system. *Science* 2016;352:539–44.
- Hayashi T, Hasegawa K, Ohta A, et al. Reduction of serum interferon (IFN)-gamma concentration and lupus development in NZBxNZWF(1)mice by lactic dehydrogenase virus infection. *J Comp Pathol* 2001;125:285–91.
- Hayashi T, Noguchi Y, Kameyama Y. Suppression of development of anti-nuclear antibody and glomerulonephritis in NZB x NZWF1 mice by persistent infection with lactic dehydrogenase virus: possible involvement of superoxide anion as a progressive effector. *Int J Exp Pathol* 1993;74:553.
- Inada T, Mims CA. Infection of mice with lactic dehydrogenase virus prevents development of experimental allergic encephalomyelitis. *J Neuroimmunol* 1986;11:53–6.
- Ni J, Wu GD, Albenberg L, et al. Gut microbiota and IBD: causation or correlation? *Nat Rev Gastroenterol Hepatol* 2017;14.
- Lepage P, Häslér 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.
- Walker AW, Sanderson JD, Churcher C, et al. High-throughput clone library analysis of the mucosa-associated microbiota reveals dysbiosis and differences between inflamed and non-inflamed regions of the intestine in inflammatory bowel disease. *BMC Microbiol* 2011;11:7.
- Hansen R, Reiff C, Russell RK, et al. Colonic mucosal bacterial diversity of de novo extensive paediatric ulcerative colitis by next-generation sequencing. *Gut* 2011;60:A146–7.
- Shah R, Cope JL, Nagy-Szakal D, et al. Composition and function of the pediatric colonic mucosal microbiome in untreated patients with ulcerative colitis. *Gut Microbes* 2016;7:384–96.
- Sepehri S, Kotlowski R, Bernstein CN, et al. Microbial diversity of inflamed and noninflamed gut biopsy tissues in inflammatory bowel disease. *Inflamm Bowel Dis* 2007;13:675–83.

- 53 Forbes JD, Van Domselaar G, Bernstein CN. Microbiome survey of the inflamed and noninflamed gut at different compartments within the gastrointestinal tract of inflammatory bowel disease patients. *Inflamm Bowel Dis* 2016;22:817–25.
- 54 Yilmaz B, Juillerat P, Øyås O, *et al.* Microbial network disturbances in relapsing refractory Crohn's disease. *Nat Med* 2019;25:323–36.
- 55 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357–9.