



# Multi-biome analysis identifies distinct gut microbial signatures and their crosstalk in ulcerative colitis and Crohn's disease

Received: 27 January 2024

Accepted: 19 November 2024

Published online: 27 November 2024

Check for updates

Shintaro Akiyama <sup>1</sup>, Suguru Nishijima <sup>2,11</sup>✉, Yasushi Kojima <sup>3</sup>, Moto Kimura <sup>4</sup>, Mitsuru Ohsugi <sup>5</sup>, Kohjiro Ueki <sup>6</sup>, Masashi Mizokami <sup>7</sup>, Masahira Hattori <sup>8</sup>, Kiichiro Tsuchiya <sup>1</sup>, Naomi Uemura <sup>9</sup>, Takashi Kawai <sup>10</sup>, Peer Bork <sup>2</sup> & Naoyoshi Nagata <sup>10,11</sup>✉

The integrative multi-kingdom interaction of the gut microbiome in ulcerative colitis (UC) and Crohn's disease (CD) remains underinvestigated. Here, we perform shotgun metagenomic sequencing of feces from patients with UC and CD, and healthy controls in the Japanese 4D cohort, profiling bacterial taxa, gene functions, and antibacterial genes, bacteriophages, and fungi. External metagenomic datasets from the US, Spain, the Netherlands, and China were analyzed to validate our multi-biome findings. We found that *Enterococcus faecium* and *Bifidobacterium* spp. were enriched in both diseases. Enriched *Escherichia coli* was characteristic of CD and was linked to numerous antibiotic resistance genes involved in efflux pumps and adherent-invasive *Escherichia coli* virulence factors. Virome changes correlated with shifts in the bacteriome, including increased abundances of phages encoding pathogenic genes. *Saccharomyces paradoxus* and *Saccharomyces cerevisiae* were enriched in UC and CD, respectively. *Saccharomyces cerevisiae* and *Escherichia coli* had negative associations with short-chain fatty acid (SCFA)-producing bacteria in CD. Multi-biome signatures and their interactions in UC and CD showed high similarities between Japan and other countries. Since bacteria, phages, and fungi formed multiple hubs of intra- or trans-kingdom networks with SCFA producers and pathobionts in UC and CD, an approach targeting the interaction network may hold therapeutic promise.

The bacteriome in inflammatory bowel disease (IBD) has been extensively studied<sup>1–7</sup> and is characterized by a reduction in the bacteriome diversity and short-chain fatty acid (SCFA)-producing bacteria, and an enrichment of several pathobionts, including *Enterobacteriaceae* spp. and *Enterococcus faecium*<sup>8,9</sup>. Regarding bacterial gene functions, adherent-invasive *Escherichia coli* (AIEC), originally isolated from patients with CD, has the ability to penetrate the mucus layer and attach to intestinal epithelial cell lines via type 1 fimbrial adhesion<sup>8</sup>, suggesting a crucial role of this bacterial function in inducing intestinal inflammation. Additionally, the *Escherichia coli* associated with ileal CD

often possesses multidrug resistance<sup>10,11</sup>. A clade of *Klebsiella pneumoniae* strains featuring unique antibiotic resistance genes (ARGs) has been shown to be associated with disease exacerbation and severity of IBD<sup>12</sup>. These findings imply that several species of the family *Enterobacteriaceae* contribute to the pathogenesis of IBD.

Bacteriophages and fungi may also be implicated in the pathogenesis of IBD, as they are involved in the acquisition and spread of functional genes and ARGs<sup>13</sup>, as well as determining bacterial colonization and shaping its structures and functions<sup>14</sup>. The gut virome of IBD is significantly diminished, likely due to bacteriome changes<sup>15</sup> or

A full list of affiliations appears at the end of the paper. ✉ e-mail: [nishijima.suguru@gmail.com](mailto:nishijima.suguru@gmail.com); [nnagata\\_ncgm@yahoo.co.jp](mailto:nnagata_ncgm@yahoo.co.jp)

mucosal inflammation<sup>16</sup>. Fungi also regulate the assembly of co-residing gut bacteria<sup>17</sup>, and *Saccharomyces* has been reported to have a positive correlation with several SCFA-producing bacteria depleted in IBD<sup>18</sup>. Furthermore, the anti-*Saccharomyces cerevisiae* antibody is a serologic marker for CD<sup>19</sup>, indicating a heavy exposure of fungi to the immune system of the patients. These findings underscore the importance of integrative analysis of the bacteriome, mycobiome, and virome in IBD. Although interactions between different biomes have been inferred in healthy individuals<sup>20</sup> or some diseases<sup>21,22</sup>, it remains unclear how these biomes are involved in IBD pathology, and whether their interactions differ between UC and CD.

In this work, we perform shotgun metagenomic analysis of fecal samples from patients with UC and CD and healthy control individuals, and extensively profile bacterial species, their genes, bacteriophages, and eukaryotic microbes. We characterize microbiome ecology through multi-biome analysis in UC and CD and identify intra- or trans-kingdom interactions in the diseases.

## Results

### Study characteristics

We performed shotgun metagenomic sequencing of fecal samples from 111 UC patients, 31 CD patients, and 540 healthy controls from the Japanese 4D (Disease, Drug, Diet, Daily life) cohort (aged 16–89 years, 59% male, Supplementary Data 1 and 2). All fecal samples were processed using an identical protocol to reduce batch effects caused by technical factors. Taxonomic and functional analyzes of the metagenomic data (Methods) identified a total of 441 genera, 4364 bacterial species, 403 antibiotic resistance genes (ARGs), and 10,689 Kyoto Encyclopedia of Genes and Genomes Orthologies (KOs). Additionally, we profiled the virome and mycobiome using a custom gene and genome database (Methods) and identified 1347 phages and 90 eukaryotic species. For internal validation, we conducted a propensity score matching analysis to balance baseline characteristics between IBD cases and new controls (1 case: 4 controls) such as age, sex, body mass index (BMI), alcohol consumption, current smoking status, diet, exercise, comorbidities, Bristol Stool Scale<sup>23</sup>, proton pump inhibitors, and antibiotics (139 IBD vs 523 controls, 110 UC vs 424 controls, and 31 CD vs 119 controls, Supplementary Data 3). For external validation, we collected and used public metagenomes from 287 IBD patients and 178 controls across four cohorts from previous studies: the United States (US) cohort (51 UC, 68 CD, and 34 control samples)<sup>2</sup>, the Netherlands (NL) cohort (23 UC, 20 CD, and 22 control samples)<sup>2</sup>, Spanish (ES) cohort (31 UC, 45 CD, and 62 control samples)<sup>6</sup>, and Chinese (CN) cohort (49 CD and 60 control samples)<sup>7</sup>.

### The bacteriome signature of commensals and pathobionts is distinct between UC and CD

We found that alpha-diversity of the bacteriome in the Japanese 4D cohort was significantly decreased in patients with UC and CD compared to controls and was significantly lower in CD than UC (Fig. 1a and Supplementary Fig. 1a–c). Multidimensional scaling (MDS) analysis showed significant differences in β-diversity for the gut bacteriome among the three groups (Fig. 1b). These findings were consistent with previous studies<sup>24</sup>.

With the age-, sex-, and BMI-adjusted regression model, we identified significant changes (false discovery rate [FDR] <0.1) in 263, 214, and 133 species in patients with IBD, UC, and CD compared with controls, respectively (Supplementary Data 4). We replicated previous results of bacterial alterations<sup>25</sup>, specifically the significant (FDR < 0.1) enrichment of *Bifidobacterium* spp. (*B. breve*, *B. longum*, and *B. dentium*), *Enterococcus* spp. (*E. faecium* and *E. faecalis*), and *Streptococcus salivarius* in IBD, and depletion of SCFA-producing bacteria such as *Faecalibacterium prausnitzii*, *Eubacterium* spp. (*E. rectale* and *E. ventriosum*), and *Anaerostipes hadrus*. No species belonging to *Archaea*, such as *Methanobrevibacter smithii*, were significantly associated with IBD.

To investigate the reproducibility of the results in the Japanese 4D cohort, we compared the coefficients of each bacterial species from the regression model with those in the external cohorts. The analysis showed significant positive correlations for all of the cohorts ( $P < 0.05$ ). The bacterial signatures between the Japanese 4D and US cohorts were similar in IBD, UC, and CD (Spearman's rho 0.62, 0.64, and 0.46, respectively, Fig. 1c), and also in the other cohorts (Spearman's rho = 0.16 to 0.53, Supplementary Fig. 2). Our bacteriome analysis showed that *Enterococcus faecium* was significantly enriched in UC and CD, whereas *Escherichia coli* was specifically associated with CD in the Japanese 4D cohort (Fig. 1d). Since this finding suggests that UC and CD may have distinct pathobionts, we next conducted a deeper analysis of identifying ESKAPE pathogens (Supplementary Methods), which are WHO-targeted pathogens known to cause multidrug-resistant nosocomial infections<sup>26</sup>. The results revealed that *Klebsiella pneumoniae*, *Staphylococcus aureus*, *Acinetobacter baumannii*, and *Pseudomonas aeruginosa* were significantly increased in CD, but not UC (FDR < 0.1, Fig. 1e). We confirmed again that *Enterococcus faecium* was significantly enriched in both UC and CD (FDR < 0.1). Since ESKAPE pathobionts, enriched in CD in our cohort, are known multidrug-resistant bacteria, there may be more multidrug-resistant bacteria in CD than in UC patients.

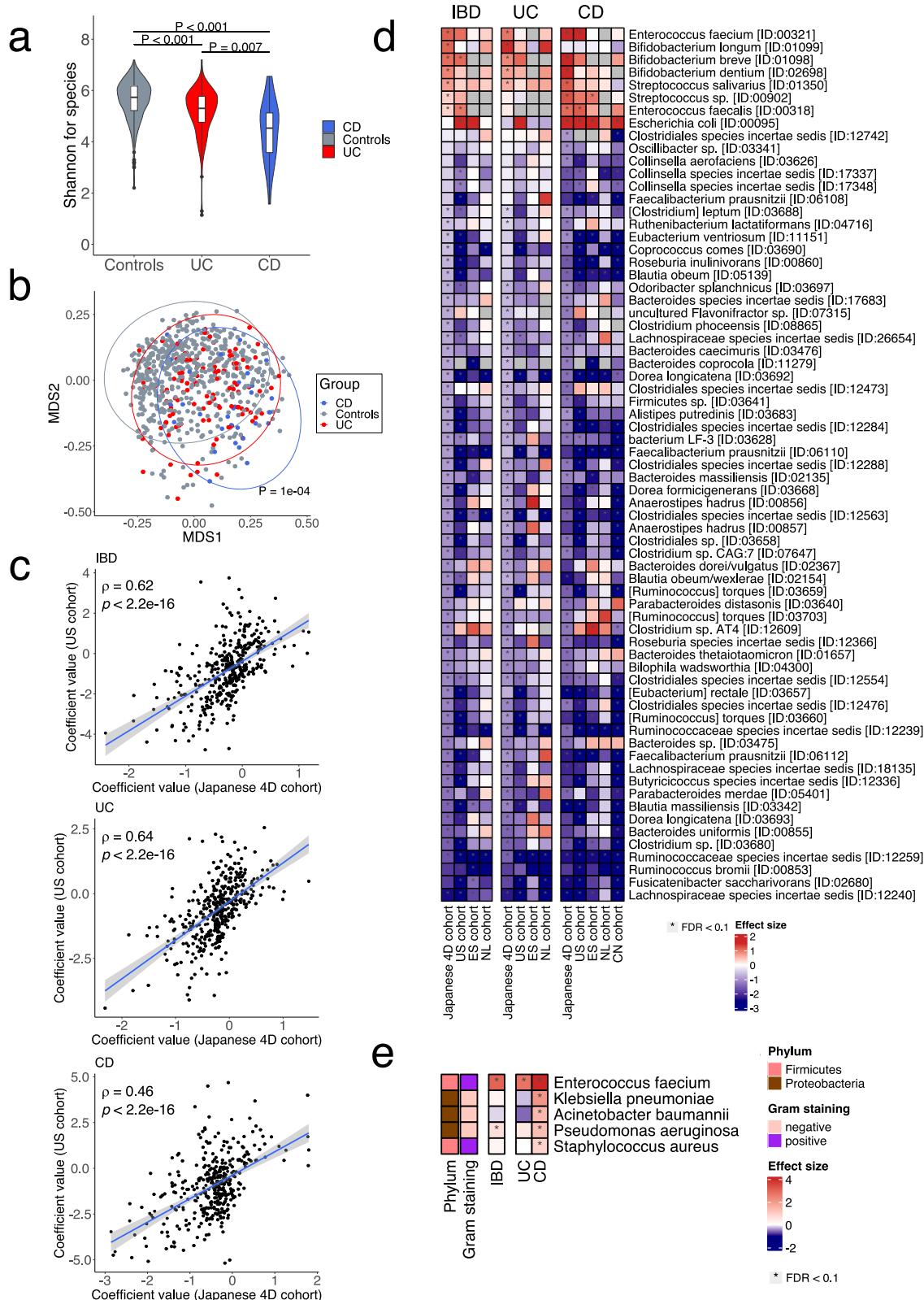
LASSO-based machine learning models (Supplementary Methods) distinguished between the patients and controls with high accuracy (area under receiver-operating characteristics [AUC] = 0.88, 0.88, and 0.87 for IBD, UC, and CD, respectively) (Supplementary Fig. 3a–c). The majority of the bacterial species selected in these models overlapped with the significant species identified by the regression models (Supplementary Fig. 3d–f and Supplementary Data 5). External validation of these models in independent cohorts showed moderate accuracy in the US cohort (AUC = 0.90, 0.84, and 0.77 for IBD, UC, and CD, respectively), ES cohort (AUC = 0.62, 0.70, and 0.64 for IBD, UC, and CD, respectively), NL cohort (AUC = 0.75, 0.85, and 0.54 for IBD, UC, and CD, respectively), and CN cohort (AUC = 0.82 for CD) (Supplementary Fig. 3g–i).

As a sensitivity analysis, we compared cases with new controls matched by propensity score (Supplementary Data 3) and found that the results were very consistent with those obtained with the original healthy controls (Supplementary Fig. 4 and Supplementary Data 6). To also test the influence of tumor necrosis factor (TNF) inhibitors and azathioprine (AZA) on the result, we selected only IBD patients without these drugs (86 UC and 14 CD) and compared them with healthy controls, confirming the consistent results (Supplementary Fig. 5 and Supplementary Data 7). Furthermore, comparisons between 96 IBD patients without any clinical symptoms (e.g., diarrhea, abdominal pain, bloody stools) (74 UC and 22 CD) and healthy controls again yielded similar results (Supplementary Fig. 6 and Supplementary Data 8). These findings underscore the robustness of the IBD-associated bacterial signature regardless of confounding factors, use of TNF inhibitors and azathioprine, and clinical symptoms.

### The resistome analysis reveals specific associations between *Escherichia coli* and antibiotic efflux pumps in CD, but not UC

To investigate the antibiotic resistance potential of the microbiome, we profiled ARGs in the metagenomes (the resistome) in the Japanese 4D cohort and compared them across the groups. Alpha-diversity of the gut resistome in CD was slightly higher than in controls and UC, but not significantly (Fig. 2a and Supplementary Fig. 1d–f). MDS analysis showed significant differences in β-diversity among the three groups, indicating the compositional dissimilarity of the resistome (Fig. 2b). The number of ARGs and their total abundances were significantly higher in CD than in UC or controls (Fig. 2c, d), suggesting the gut commensals in CD patients would more likely possess ARGs.

A comparative analysis with controls identified 26, 20, and 77 ARGs with significant changes in IBD, UC, and CD, respectively



( $FDR < 0.1$ , Supplementary Data 9). Specifically, efflux pumps, such as *acrB*, *emrR*, and *acrA*, were significantly enriched in CD whereas they were not significantly changed in UC. Marked enrichment of efflux pumps specifically in CD suggests that they lead to the emergence of multiple-antibiotic-resistant-bacteria<sup>27</sup>.

Next, we examined the correlation between ARGs and bacterial species and found that *msrC*, enriched in UC and CD, had a positive

correlation (Spearman's rho = 0.73) with *Enterococcus faecium*, which was also enriched in UC and CD (Fig. 2e). Notably, 38 antibiotic efflux pumps specifically enriched in CD showed positive correlations (Spearman's rho = 0.22–0.92,  $FDR < 0.1$ ) with *Escherichia coli* and *Klebsiella pneumoniae*. Interestingly, these species also showed significant positive correlations with the total number of ARGs detected in the sample (Spearman's rho = 0.82 for *Escherichia coli* and

**Fig. 1 | Gut bacteriome signatures of UC and CD.** **a** Violin plots showing the Shannon index of the gut bacteriome in controls ( $n = 540$  individuals), ulcerative colitis (UC) ( $n = 111$  individuals), and Crohn's disease (CD) ( $n = 31$  individuals). Violin plots and box-and-whisker plots were depicted using the ggplot2 package in R. In boxplots, boxes represent the interquartile range (IQR), and the lines inside show the median. Whiskers denote the lowest and highest values within 1.5 times the IQR. Values were compared using the Kruskal–Wallis test with a post hoc analysis of multiple comparisons using Dunn's test with the Holm correction (two-sided).  $P$ -value (controls vs. UC): 2.38E-7,  $P$ -value (controls vs. CD): 8.78E-9. **b** Multidimensional scaling (MDS) plots showing the similarity of samples. Gray, red, and blue circles represent controls, UC, and CD, respectively.  $P$ -values were calculated by permutational analysis of variance. **c** A scatterplot showing coefficient values for bacterial species signatures for inflammatory bowel disease (IBD), UC, and CD identified in our Japanese cohort (JP, x-axis) and the United States cohort (US, y-axis). Coefficient values of bacterial species alterations between IBD and controls were estimated using MaAsLin2. Concordance values of 388, 403, and

374 species for IBD, UC, and CD, respectively, were calculated by Spearman's rank correlation coefficient and its statistical significance was calculated using the stat.cor function in the ggpubr package in R (two-sided). The blue line represents the regression line, and the gray shadow shows 95% confidence interval. **d** The heatmap shows enriched and depleted species in IBD patients by comparison to the healthy controls in the Japanese 4D cohort, United States (US), Spanish (ES), Netherlands (NL), and Chinese (CN) cohorts. The plot includes only bacterial species showing absolute values of coefficient  $> 1.0$  and false discovery rates (FDR)  $< 0.1$  in UC or CD (Japanese 4D cohort) compared with controls based on MaAsLin2. Effect size is the coefficient value obtained by MaAsLin2. Asterisks represent FDR  $< 0.1$ . Numbers in parentheses beside the species name indicate the mOTUs3 ID. **e** The heatmap shows enriched and depleted species of ESKAPE pathogens in IBD, UC, and CD compared with healthy controls in the Japanese 4D cohort. Effect size is the coefficient value obtained by MaAsLin2. Asterisks represent FDR  $< 0.1$ . Row annotations show the profiles of phylum and gram staining. Source data are provided as a Source Data file.

Spearman's rho = 0.39 for *Klebsiella pneumoniae*; Fig. 2e), suggesting the presence of multi-resistant bacteria in the gut of CD. Furthermore, consistent with the correlation analysis, the Comprehensive Antibiotic Resistance Database (CARD) suggests that family *Enterobacteriaceae* has these ARGs (Supplementary Fig. 7 and Supplementary Methods).

Taken together, our findings suggest that patients with IBD have a greater risk of infection with multiple-antibacterial-resistant pathogens, specifically *Escherichia coli* and *Klebsiella pneumoniae*, compared to healthy individuals.

#### Unique microbial functions inducing inflammation and resisting antimicrobial defense exist in IBD

Functional analysis based on the KEGG module (MO) levels identified 336, 322, and 341 MOs with significant changes in patients with IBD, UC, and CD compared with controls, respectively (FDR  $< 0.1$ , Supplementary Data 10). Specifically, C5 isoprenoid biosynthesis (the mevalonate pathway for terpenoid biosynthesis) and cationic antimicrobial peptide (CAMP) resistance were significantly (FDR  $< 0.1$ ) enriched in both UC and CD, whereas several two-component regulatory systems were significantly depleted (FDR  $< 0.1$ , Fig. 3a, b). The MOs responsible for N-glycosylation by oligosaccharyltransferase were significantly (FDR  $< 0.1$ ) depleted in UC, whereas degradation of glycosaminoglycans such as heparan, chondroitin, and dermatan sulfate were significantly depleted in CD (FDR  $< 0.1$ , Fig. 3a, b).

Next, we focused on gut microbial functionality in KEGG orthology (KO) levels and identified 3637, 3347, and 3995 KOs with significant changes in patients with IBD, UC, and CD compared with controls, respectively (FDR  $< 0.1$ , Supplementary Data 11). Consistent with the results based on MOs (Fig. 3a, b), KOs involved in terpenoid backbone biosynthesis were significantly enriched in both UC and CD (FDR  $< 0.1$ , Fig. 3c). KOs responsible for N-glycan biosynthesis (KO7252 and KO7151) and glycosaminoglycan degradation were significantly depleted in both diseases (FDR  $< 0.1$ , Fig. 3d, e). These findings suggest that bacterial glycan metabolism is down-regulated in UC and CD. Meanwhile, some KOs belonging to CAMP resistance were significantly (FDR  $< 0.1$ ) increased in both diseases. Of note, genes for CAMP resistance (e.g., genes belonging to the OmpR family) were specifically enriched in CD but not in UC (Fig. 3f). Our data indicate that unique mechanisms to resist antimicrobial defense exist in IBD, and the species involved in the mechanism differ distinctly between UC and CD.

We examined microbial functions involved in pathogen-associated molecular pattern molecules<sup>28</sup> as well as potential virulence factors of AIEC<sup>8</sup>, which can activate inflammatory signal cascades and may play an important role in IBD. KOs for flagellar assembly and lipopolysaccharide (LPS) biosynthesis were significantly (FDR  $< 0.1$ ) enriched in CD, and these KOs were positively correlated with *Escherichia coli* and *Klebsiella pneumoniae* (Spearman's rho = 0.13–0.85, FDR  $< 0.1$ , Fig. 3g and Supplementary Fig. 8a). In

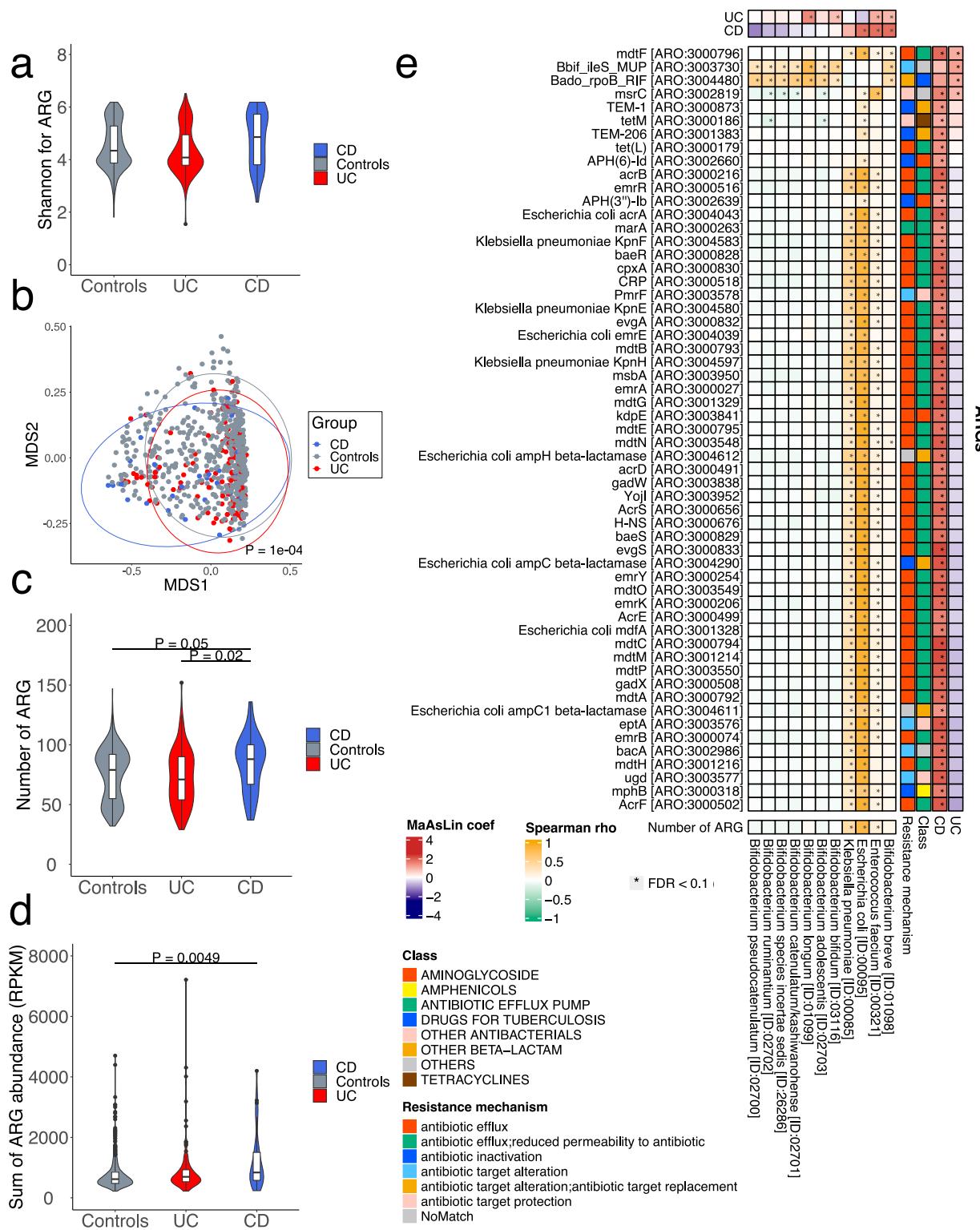
addition, KOs for peptidoglycan biosynthesis were significantly enriched in both UC and CD, and these KOs were positively correlated with *Streptococcus salivarius* and *Streptococcus parasanguinis*, which were also significantly enriched in UC and CD (Fig. 3g and Supplementary Fig. 8a). As for AIEC virulent factors, we examined 30 KOs previously reported to be involved in AIEC pathogenesis<sup>8</sup> (Supplementary Methods), and found that 12 of them were significantly (FDR  $< 0.1$ ) enriched in CD. Their functions predominantly included adhesion to and invasion of intestinal epithelial cells (e.g., type 1 fimbriae) (Fig. 3h and Supplementary Fig. 8b). Bacterial species positively correlated (Spearman's rho = 0.27–0.82) with these genes included *Escherichia coli* and *Klebsiella pneumoniae* (Supplementary Fig. 8b). This finding suggests that *Escherichia coli* has virulent genes for AIEC in CD, consistent with previous studies<sup>8</sup>.

#### IBD patients have altered gut-virome-encoding pathogenic proteins

We next investigated viral species in the metagenomes by mapping the sequences to the viral genome catalog (Methods). Alpha-diversity for the virome was significantly lower in patients with UC and CD compared with controls, and the Shannon index of UC was significantly decreased compared with CD in the Japanese 4D cohort (Fig. 4a and Supplementary Fig. 1g–i). MDS analysis showed significant differences in β-diversity for the virome among the three groups (Fig. 4b). The total abundance of lysogenic phages was not significantly different among the three groups while the total abundance of phages predicted to infect only one species (specialist phages) was significantly higher in both UC and CD compared with controls (Fig. 4c, d).

A comparison of viral operational taxonomic units (vOTUs) identified 65, 37, and 16 vOTUs with significant changes in patients with IBD, UC, and CD compared with controls, respectively, in the Japanese 4D cohort (FDR  $< 0.1$ , Supplementary Data 12). The predicted hosts of the vOTUs with the strongest associations were *Lactobacillus* (vOTU1252) in UC and *Streptococcus* (vOTU689) in CD (Fig. 4e). Meanwhile, vOTU77, a phage whose host was predictive of *Escherichia*, was specifically enriched in CD (FDR  $< 0.1$ ). Among vOTUs significantly depleted in UC and CD, the most common predicted host was *Ruminococcus*, followed by *Blaautia*, and *Bacteroides*, which are SCFA-producing bacteria depleted in UC and CD (Fig. 4e, Supplementary Fig. 9 and Supplementary Data 12, 13). The consistent alterations of the bacteriome and virome imply that the reduced richness and diversity of the phageome may be secondary to changes in the bacteriome.

Comparison of the virome signatures (i.e., set of coefficients from the regression model for each vOTU) between the Japanese 4D and other cohorts showed significant positive correlations for IBD, UC, and CD ( $P < 0.05$ , Spearman's rho = 0.25 to 0.45, Supplementary Fig. 10), indicating the robustness of the results. In particular, phages whose predicted hosts included SCFA-producing bacteria were consistently



depleted in all the cohorts, and vOTU87, 89, 90, and 77 were commonly enriched in CD patients of the US, ES, and CN cohorts (Fig. 4e).

Notably, functional analysis revealed that three of the seven enriched phages in UC or CD encoded putative pathogenic genes, vOTU89 and vOTU90, which were significantly enriched in CD, encoded *pagC* gene (K07804), an outer membrane protein that is highly immunogenic<sup>29</sup>. These phages had genome sizes of approximately 45.9 kb and 23.9 kb, respectively, and showed no sequence similarity to reference phage genomes in RefSeq. They were predicted to be

template phages infecting *Escherichia coli*, as they encoded integrases and some genes had similarities to the *Escherichia coli* genome. In addition, we found that vOTU689 encoded *rfb* gene clusters, such as *rfbB* and *rfbC* (K01710 and K01790, respectively), which could contribute to increased O antigen diversity (Fig. 4f)<sup>30</sup>. The genome size of this phage was approximately 46.1 kb, and alignments of CRISPR spacers predicted the host to be *Streptococcus* spp. The specific enrichments of vOTU89 and vOTU90 in CD were consistent across the US, ES, and CN cohorts (Fig. 4e). These results suggest that the

**Fig. 2 | Resistome signatures of UC and CD and their correlations with species possessing antibiotic resistance genes.** **a** Violin plots showing the Shannon index of the gut resistome in the controls ( $n = 540$  individuals), ulcerative colitis (UC) ( $n = 111$  individuals), and Crohn's disease (CD) ( $n = 31$  individuals).

**b** Multidimensional scaling (MDS) plots showing the similarity of samples. Gray, red, and blue circles represent controls, UC, and CD, respectively.  $P$ -values were calculated by permutational analysis of variance. **c** Box-and-whisker plots show the number of antibiotic resistance genes (ARGs) identified in the controls ( $n = 540$  individuals), UC ( $n = 111$  individuals), and CD ( $n = 31$  individuals). **d** Box-and-whisker plots show the total abundance values of ARGs in the controls ( $n = 540$  individuals), UC ( $n = 111$  individuals), and CD ( $n = 31$  individuals). RPKM, reads per kilobase million. **e** The heat map depicts correlations between bacterial species and ARGs in UC and CD in comparison to the healthy controls (HC). The heat map includes only

bacterial species showing absolute values of Spearman's rank correlation coefficient  $> 0.4$ . Row annotations depict coefficient values (MaAsLin2) of ARGs for UC and CD in comparison to HC. The row annotations include only ARGs showing coefficient  $> 1.0$  and false discovery rates (FDR)  $< 0.1$  in UC or CD compared with HC. The profiles of the ARG class and resistance mechanisms are shown in the row annotations. Column annotations depict coefficient values (MaAsLin2) of bacterial species with UC and CD in comparison to HC. Asterisks represent FDR  $< 0.1$ . Violin plots and box-and-whisker plots in **(a, c, and d)** were depicted using the ggplot2 package in R. In boxplots, boxes represent the interquartile range (IQR), and the lines inside show the median. Whiskers denote the lowest and highest values within 1.5 times the IQR. Values were compared using the Kruskal–Wallis test with a post hoc analysis of multiple comparisons using Dunn's test with the Holm correction (two-sided). Source data are provided as a Source Data file.

enrichment of phages with such specific genes contributes to the pathogenicity of the bacteriome in CD patients.

### Distinct mycobiome signature between UC and CD

We next investigated eukaryotic species such as fungi and protozoa in the metagenomes (Methods). In the Japanese 4D cohort, only 2, 3, and 1 fungi with FDR  $< 0.1$  were identified in patients with IBD, UC, and CD, respectively, compared to controls (Supplementary Data 14). Due to the low detection rate of eukaryotic species with FDR  $< 0.1$ , a  $P$ -value was used for exploratory analysis.

A comparison of the detected eukaryotic species identified 6, 3, and 2 fungi with significant changes ( $P < 0.05$ ) in patients with IBD, UC, and CD compared with controls, respectively in the Japanese 4D cohort (Supplementary Data 14). Fungi significantly enriched in CD included *Saccharomyces cerevisiae* ( $P = 0.018$ , FDR = 0.43) and *Debayomyces hansenii* ( $P < 0.001$ , FDR  $< 0.001$ ), whereas those enriched in UC included *Saccharomyces paradoxus* ( $P = 0.001$ , FDR = 0.053) and *Saccharomyces kudriavzevii* ( $P = 0.007$ , FDR = 0.077). A significant depletion of *Exophiala mesophila* was noted only in UC ( $P = 0.004$ , FDR = 0.053, Fig. 5a). No species classified into protists was significantly associated with IBD (Fig. 5a).

To validate the results in the Japanese 4D cohort, we compared eukaryotic species in the external datasets and found that *Saccharomyces cerevisiae* was also enriched in patients with CD in the US ( $P = 0.19$ , FDR = 0.48, Supplementary Data 14) and the ES cohorts ( $P = 0.064$ , FDR 0.54) but significantly depleted in CD in the CN cohort ( $P = 0.021$ , FDR = 0.21). *Saccharomyces paradoxus*, which was enriched in UC in the Japanese 4D cohort, was also significantly increased in UC in the US cohort ( $P = 0.040$ , FDR = 0.51). In the ES cohort, *Saccharomyces paradoxus* and *Saccharomyces kudriavzevii* were significantly enriched in CD ( $P = 0.022$  and FDR = 0.45 for *Saccharomyces paradoxus* and  $P = 0.012$  and FDR = 0.44 for *Saccharomyces kudriavzevii*). *Debayomyces hansenii* was significantly increased in UC in the ES cohort ( $P = 0.005$ , FDR = 0.48, Fig. 5a and Supplementary Data 14).

### Distinct intra- and trans-kingdom interactions between UC and CD

To investigate the interaction networks between disease-associated microbes, we constructed correlation networks among the significant bacteria, virus, and fungal species in the Japanese 4D cohort. To remove the confounding effects of disease on the correlations, we evaluated the correlations in the healthy controls. Associations significantly different from zero (FDR  $< 0.1$  and Spearman's rho  $> 0.2$ ) and important interactions (e.g., linkages between expected hosts and phages, fungi and bacteria, and pathogens and bacteria) are described in Fig. 5b, c.

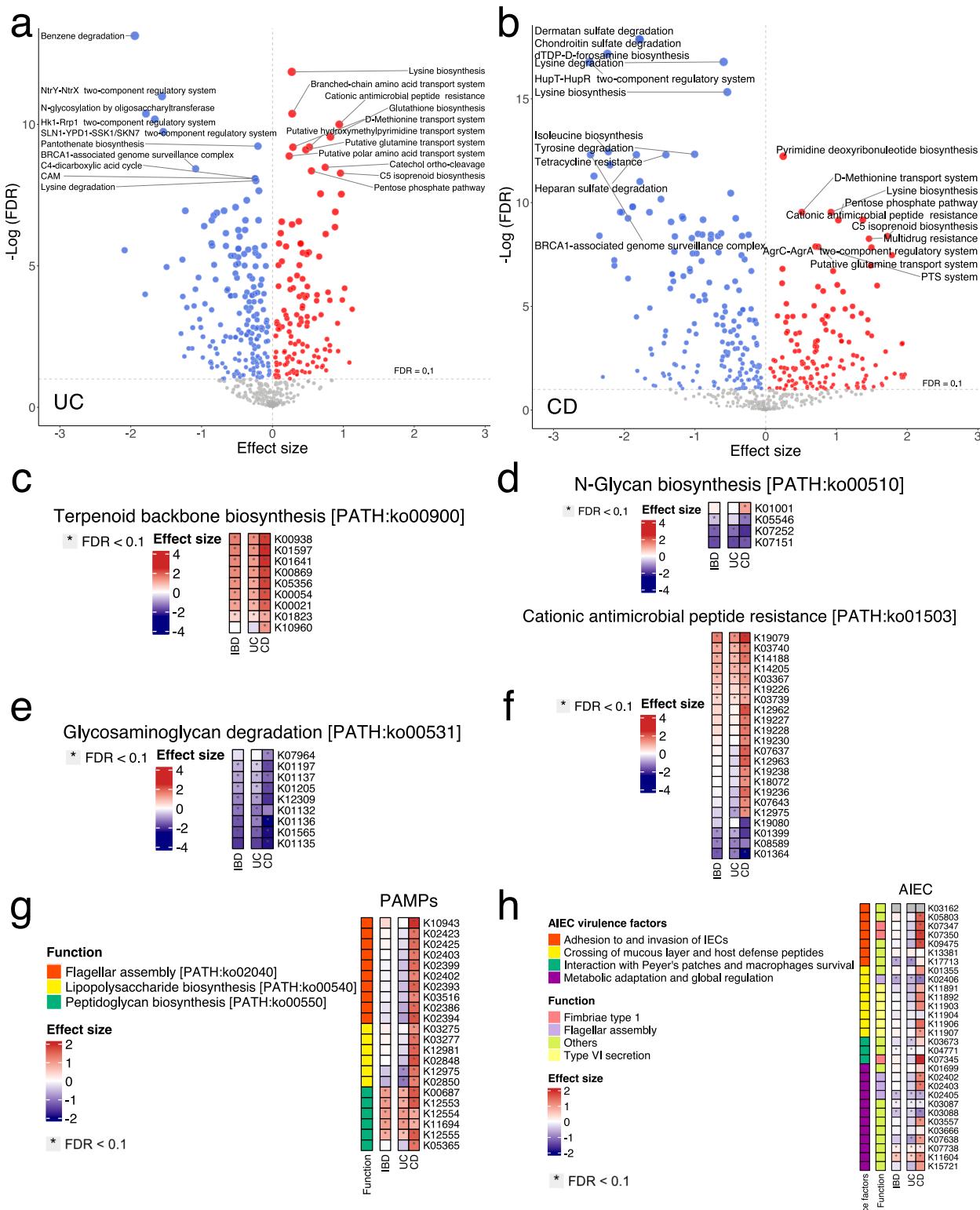
In the correlation network of UC-related species, SCFA-producing bacteria such as *Blautia massiliensis*, *Dorea longicatena*, and unclassified *Ruminococcus*, which were reduced in UC, formed clusters (Fig. 5b). Additionally, we observed several clusters of phages and their

predicted hosts (e.g., *Ruminococcus bromii* and *Ruminococcus* phages of vOTUs 160, 161, 162, and 164; *Fusicatenibacter saccharivorans* and *Fusicatenibacter* phages of vOTUs 289, 291, 293, and 294). In the correlation network of CD-related species, SCFA-producing bacteria such as *Eubacterium ventriosum*, *Blautia obeum* and unclassified *Clostridiales* formed a large cluster that also included several phages predicted to infect them. In contrast, *Escherichia coli* and *Saccharomyces cerevisiae*, which were significantly increased in CD, showed negative associations with these species. *Escherichia* phage (vOTU77) was significantly correlated with *Escherichia coli* (Fig. 5c). The finding that UC- and CD-related species correlated strongly with each other and formed multiple clusters highlights the importance of their interactions in the gut microbiome of patients.

To validate the correlation networks constructed in the Japanese 4D cohort, we investigated the network in the US, ES, NL, and CN cohorts. Our analysis of UC-related species showed that *Ruminococcus* phages (vOTU160, 161, 162, and 164) were also significantly correlated with *Ruminococcus bromii* in the US and ES cohorts, and this association was also observed in the NL cohort, although not at a statistically significant level (Fig. 5d). Meanwhile, our analysis of CD-related species showed that vOTU77, *Escherichia* phage, was significantly associated with *Escherichia coli* in the CN and ES cohorts. Significant positive associations between *Anaerostipes* phage (vOTU758) and *Anaerostipes hadrus* as well as *Streptococcus* phage (vOTU689) and *Streptococcus* sp. were also observed in the ES cohort. *Saccharomyces cerevisiae* formed negative correlations with *Eubacterium ventriosum* and *Clostridiales* sp. in the US cohort. Negative correlations between *Escherichia coli* and SCFA-producing bacteria (*Anaerostipes hadrus*, *Blautia obeum*, and *Clostridiales* sp.) were also observed in all cohorts (Fig. 5e). These results suggest the robustness of the correlation networks we discovered in the Japanese 4D cohort.

### Discussion

This study identifies distinct multi-biome dynamics and their intra- and trans-kingdom interactions in UC and CD, examining their bacteriomes including taxa, functions, resistomes, viromes, and mycobiomes. Multi-biomes and their interactions in our cohort were validated using large-scale external datasets from the US, Spain, Netherlands, and China. Specifically, first, our bacteriome analysis found that *Enterococcus faecium* and *Bifidobacterium* spp. were significantly enriched in both UC and CD, whereas *Escherichia coli* was specifically enriched in CD. Terpenoid backbone biosynthesis was commonly enriched in both UC and CD, and glycan metabolism was depleted in both diseases. Second, *Escherichia coli* and *Klebsiella pneumoniae*, specifically enriched in CD, showed a significant association with numerous antibiotic efflux pumps and AIEC virulence factors. Third, the virome reflected the bacteriome differences between UC and CD, and enriched phages carried specific genes encoding proteins that may contribute to pathogenicity. Finally, our multi-biome analysis revealed that *Saccharomyces cerevisiae* and



*Escherichia coli* formed negative correlations with SCFA-producing bacterial species in CD. These findings highlight the importance of cross-talk within the multi-biome for understanding the pathogenesis implicated in UC and CD.

Terpenoid backbone biosynthesis requires hydroxymethylglutaryl-CoA reductase<sup>31</sup>, an enzyme targeted by statins, which were enriched in UC and CD. A systematic review and meta-analysis revealed that statin

use was associated with a reduced risk of new-onset IBD<sup>32</sup>. All these findings support statins as potential candidates as preventive or therapeutic drugs for IBD. On the other hand, the depletion of glycan metabolism was also a common functional feature in both the UC and CD microbiomes. Gut bacteria utilize dietary fiber or mucin polysaccharides, including glycans and glycosaminoglycans, as their energy source. *Bacteroides* spp., which were depleted in UC and CD in our

**Fig. 3 | Functional signatures of UC and CD.** **a, b** The volcano plots show the associations between KEGG modules (MOs) and ulcerative colitis (UC) (**a**) and Crohn's disease (CD) (**b**). Effect size is the coefficient value obtained by MaAsLin2. Red and blue colors represent a significant enrichment and depletion of the MOs in the patients, respectively (absolute values of coefficient > 0 and false discovery rate (FDR) < 0.1 in UC or CD compared with controls based on MaAsLin2). **c–f** Heatmaps show coefficient values (MaAsLin2) of KEGG orthologies (KOs) with UC and CD in comparison to the healthy controls (HC). The plot includes only KOs showing absolute values of coefficient > 1.0 and FDR < 0.1 in UC or CD compared with HC. Heatmaps show the following pathways: (**c**) terpenoid backbone biosynthesis, (**d**) N-glycan biosynthesis, (**e**) glycosaminoglycan degradation, and (**f**) cationic

antimicrobial peptide (CAMP) resistance. Asterisks represent FDR < 0.1. **g** The heatmap depicts coefficient values (MaAsLin2) of KOs regarding pathogen-associated molecular pattern (PAMP) molecules in UC and CD in comparison to the healthy controls (HC). The plot includes only KOs showing coefficient values > 1.0 and FDR < 0.1 in UC or CD compared with HC. Row annotations show the hierarchy level C of flagellar assembly, lipopolysaccharide biosynthesis, and peptidoglycan biosynthesis. Asterisks represent FDR < 0.1. **h** The heatmap depicts coefficient values (MaAsLin2) of KOs regarding adherent-invasive *Escherichia coli* (AIEC) virulence factors in UC and CD in comparison to HC. Row annotations show the functions of these virulence factors. Asterisks represent FDR < 0.1. Source data are provided as a Source Data file.

study, are characterized as general glycan degraders, and some species (e.g., *Bacteroides thetaiotaomicron*) can switch from dietary glycan to host mucin glycan metabolism<sup>33</sup>. The intestinal mucus layer plays the role of a gut barrier against pathogens and is composed of mucin<sup>34</sup>. In patients with IBD, the intestinal mucin layer can be diminished due to goblet cell depletion<sup>35</sup> and the reduction of SCFA-producing bacteria that regulate intestinal homeostasis<sup>24</sup>. Our finding implies that the reduction of mucin glycan-degrading bacteria is secondary to intestinal mucin depletion in IBD.

Antimicrobial resistance (AMR) is a global public health threat<sup>36</sup>. Our resistome analysis found that *Escherichia coli* and *Klebsiella* sp. had a positive association with numerous antibiotic efflux pumps, whereas *Enterococcus faecium* was well correlated with *msrC*, an ABC-F subfamily protein conferring resistance to macrolide and streptogramin B antibiotics<sup>37</sup>. This finding suggests that patients with IBD were at high risk of infection with multiple-antibiotic-resistant-bacteria<sup>27</sup>. Previous investigations reported that *Escherichia coli* associated with ileal CD often possessed multidrug resistance genes in CD<sup>10,11</sup>, and AIEC strains were more likely to be resistant to antibiotics compared with non-AIEC strains<sup>38</sup>. Furthermore, a recent study showed that *Klebsiella pneumonia* strains featuring unique ARGs affected disease severity and worsening of IBD<sup>12</sup>. Therefore, further investigations are warranted to understand how AMR affects disease activity and prognosis in IBD.

We suggest possible countermeasures to control such IBD-related pathobionts. Our virome analysis found the enrichment of vOTU77 in CD, which was predicted to infect *Escherichia*. Our network analysis also showed that vOTU77 had significant positive interactions with *Escherichia coli*, and this association was also identified in the ES and CN cohorts. While previous investigations have demonstrated the effectiveness and safety of AIEC-targeted phages for a mouse colitis model<sup>39</sup>, our findings provide additional evidence for an approach targeting AIEC in patients with CD. Since other phages (e.g., vOTU89 and vOTU90) enriched in CD in our cohort as well as other cohorts carried genes encoding pathogenic proteins that can be associated with inflammation, phage selection for therapies based on detailed phage profiling seems likely to become a useful method.

Fungi and bacteria communicate with each other using fatty acids, lactic acid, and butyrate<sup>17</sup>. We found an enrichment of *Saccharomyces cerevisiae* in CD, and this fungus showed negative associations with SCFA-producing bacteria. These findings suggest that a unique trans-kingdom interaction exists between fungi and bacteria, and that *Saccharomyces cerevisiae* may increase to restore SCFA-producing bacteria in CD. Furthermore, we identified a negative correlation between *Escherichia coli* and SCFA-producing bacteria. AIEC uses the adhesin FimH, located at the top of type 1 fimbriae, to attach ileal epithelial cells via the interaction between FimH and mannose residues of glycoprotein CEACAM6<sup>40</sup>. An in vitro analysis has shown that colonic SCFA can inhibit *Escherichia coli* adhesion and invasion of intestinal epithelial cells<sup>41</sup>, suggesting that SCFA or SCFA-producing bacteria may be therapeutics to modulate AIEC in CD. While our result supports the idea that specific intra- or trans-kingdom interaction networks can be

fresh therapeutic targets in patients with IBD, future studies are important to clarify whether these strains interact in a mutually exclusive or mutualistic manner.

We acknowledge limitations in our study. Although our multi-biome analysis strongly suggests a mechanism of microbiome-mediated chronic inflammation in the host, this has not been demonstrated in animal or in-vitro studies. Although the majority of our results overlapped with those of the other cohorts, IBD-related microbial species identified in the Japanese 4D cohort were not necessarily associated with IBD in the validation cohorts. This may be attributable to differences in sample size, diet, or lifestyle characteristics in each country. For instance, our cohort had a larger sample size of healthy controls than the validation cohorts. However, the number of patients with CD in our cohort was modest, and this may reduce the power to detect a significant association between *Saccharomyces cerevisiae* and CD with FDR < 0.1. Furthermore, the CN cohort showed that *Saccharomyces cerevisiae* was significantly depleted in patients with CD, suggesting that the country in which the samples were collected may influence the mycobiome results. Therefore, further research with larger sample sizes of IBD, including systematic review and meta-analysis of shotgun sequencing data, is needed to determine whether these findings can be applied outside of the Japanese 4D cohort. Furthermore, the lack of high-quality fungal genomes limited the number of fungi that could be included in the analysis. Future research should incorporate the growing number of fungal genomes from various projects to provide comprehensive profiling of gut fungi.

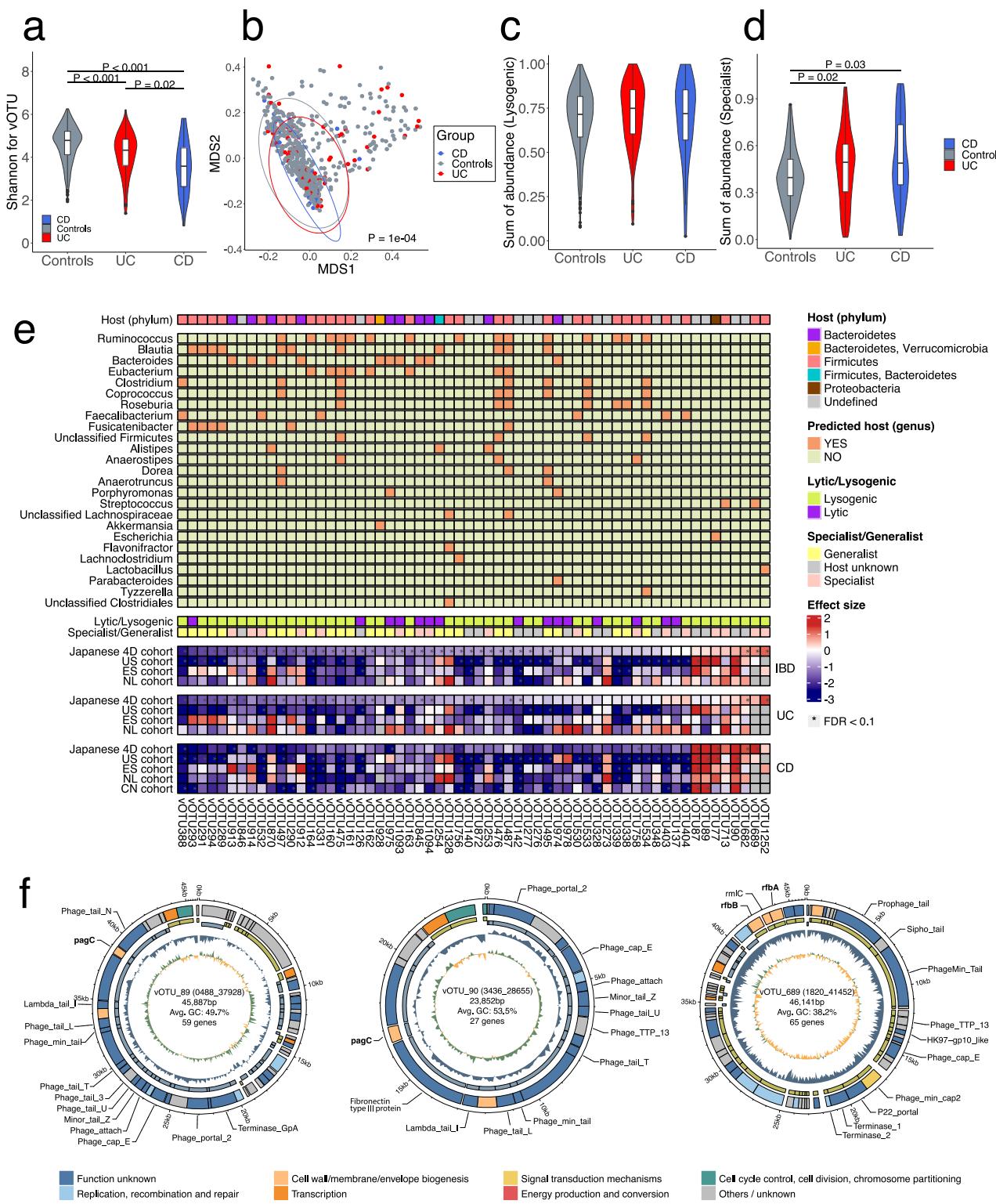
In conclusion, we performed an integrative multi-biome analysis and validated our findings using large-scale external datasets from multiple countries. This study demonstrated that UC and CD have distinct profiles of the bacteriome, virome, and mycobiome, underscoring the importance of altered trans-kingdom interactions in the pathogenesis of IBD. Our data revealed an increased risk of AMR pathogens, particularly in CD. Since the multi-biome forms unique intra- and trans-kingdom networks with CD-related pathogens, an approach targeting the interaction network rather than individual bacterial species may hold therapeutic promise for controlling IBD pathobionts.

## Methods

### Study design, sample collection, and metadata

The Japanese 4D (Disease, Drug, Diet, Daily life) microbiome project is a prospective cohort study that commenced in January 2015 and is ongoing (Supplementary Methods)<sup>42–44</sup>. The protocol was approved by the medical ethics committees of Tokyo Medical University (Approval No. E2023-0026), the National Center for Global Health and Medicine (Approval No. NCGM-S-004586), the RIKEN Center for Integrative Medical Sciences (Approval No.H30-7), and the University of Tsukuba (Approval No. R05-022). Written informed consent was obtained from all subjects.

From the Japanese 4D cohort, we selected all patients diagnosed with UC ( $n = 111$ ) and CD ( $n = 31$ ) during the study period between 2015 and 2019 at the National Center for Global Health and Medicine. All IBD



patients and healthy individuals underwent colonoscopy, and diagnosis of IBD was based primarily on endoscopic findings and biopsy pathology, with infectious enterocolitis ruled out by negative culture tests and clinical courses<sup>45</sup>. Well-trained endoscopists performed total colonoscopy with an electronic high-resolution video endoscope (Olympus Optical, Tokyo, Japan). Healthy controls ( $n=540$ ) were defined as individuals who did not have any diseases or take any medications. We prospectively collected fecal samples and metadata from both healthy and diseased participants as previously described<sup>42</sup>. The fecal samples were collected into tubes containing Cary-Blair

medium, which is relatively stable at 4 °C and 25 °C<sup>46</sup>. Since bowel preparation agents can alter the gut microbiome<sup>47</sup>, we avoided collecting stool samples within one month after bowel preparation when participants had colonoscopy. Stool samples were stored at -80 °C until DNA extraction.

To validate associations found between IBD and the gut bacteriome in the internal dataset, we used propensity score matching analysis to reduce the effects of selection bias and confounders<sup>48</sup>. To estimate the propensity score, we selected a logistic regression model with important confounders such as age, sex, body mass index (BMI),

**Fig. 4 | Virome signatures of UC and CD and their correlations with the predicted host.** **a** Violin plots showing the Shannon index of the gut virome in controls ( $n = 540$  individuals), ulcerative colitis (UC) ( $n = 111$  individuals), and Crohn's disease (CD) ( $n = 31$  individuals).  $P$ -value (controls vs. UC): 4.23E-6,  $P$ -value (controls vs. CD): 7.23E-7. **b** Multidimensional scaling (MDS) plots showing the similarity of samples. Gray, red, and blue circles represent controls, UC, and CD, respectively.  $P$ -values were calculated by permutational analysis of variance. **c** Box-and-whisker plots show the total abundance values of lysogenic phages in the controls ( $n = 540$  individuals), UC ( $n = 111$  individuals), and CD ( $n = 31$  individuals). **d** Box-and-whisker plots show the total abundance values of specialist phages in the controls ( $n = 540$  individuals), UC ( $n = 111$  individuals), and CD ( $n = 31$  individuals). **e** The heatmap shows coefficient values of vOTUs with inflammatory bowel disease (IBD), UC, and CD in comparison to the healthy controls in the Japanese 4D cohort, United States (US), Spanish (ES), Netherlands (NL), and Chinese (CN) cohorts. The plot includes only vOTUs showing absolute values of coefficient  $> 1.0$  and false discovery rates (FDR)  $< 0.25$  in UC or CD (Japanese 4D cohort) compared with

controls based on MaAsLin2. Predicted hosts (genus) are shown in the row annotation and listed from left to right in order of frequency. Row annotations also show the host phylum, lytic/lysogenic phage, and specialist/generalist phage. Effect size is the coefficient value obtained by MaAsLin2. Asterisks represent FDR  $< 0.1$ .

**f** Genome structures of the three representative phages with potential pathogenic genes (i.e., vOTU 89, 90, and 689). From inner to outside, GC skew, GC content, the strand of encoded genes, and functional annotation of the genes are shown. Potential pathogenic genes and phage hallmark genes annotated with the Pfam database are shown outside of the circle. Violin plots and box-and-whisker plots in (a, c, and d) were depicted using the ggplot2 package in R. In boxplots, boxes represent the interquartile range (IQR), and the lines inside show the median. Whiskers denote the lowest and highest values within 1.5 times the IQR. Values were compared using the Kruskal-Wallis test with a post hoc analysis of multiple comparisons using Dunn's test with the Holm correction (two-sided). Source data are provided as a Source Data file.

alcohol consumption, current smoking status, diet, exercise, comorbidities, Bristol Stool Scale<sup>23</sup>, proton pump inhibitors, and antibiotics. We performed one-to-four propensity score matching between the two groups using the nearest neighbor method within a caliper width of 0.2 of the standard deviation of the logit of propensity score. Before matching, the areas under the receiver operating characteristic curve for the logit of propensity scores for IBD, UC, and CD were 0.80 (95% confidence interval [CI], 0.76–0.84), 0.80 (95% CI 0.76–0.84), and 0.90 (95% CI 0.84–0.95), respectively.

We compared the findings of the Japanese 4D cohort to an external shotgun metagenomic dataset from a US cohort including 119 patients with IBD (51 UC and 68 CD) and 34 non-IBD controls<sup>2</sup>. We also used the validation cohort from the NL, which includes 43 patients with IBD (23 UC and 20 CD) and 22 controls, for comparison. This cohort was used as a validation cohort in the US study<sup>2</sup>. An ES cohort including 76 patients with IBD (31 UC and 45 CD) and 62 controls<sup>6</sup>, and a CN cohort including 49 patients with CD and 60 controls<sup>7</sup> were also used to validate our findings.

#### Metagenomic analysis of stool samples and bioinformatic analysis

Shotgun metagenomic sequencing of fecal samples and data analysis for the bacteriome and virome were performed as described previously<sup>43,49</sup>. Briefly, microbial DNA was prepared using the enzyme lysis method, and libraries for sequencing were prepared and sequenced using the Illumina HiSeq X. Quality filtering and adaptor trimming of the sequenced reads were performed using prinseq<sup>50</sup> and cutadapt<sup>51</sup>, respectively. The high-quality metagenomic reads were mapped to the human genome (HG38) using bowtie2<sup>52</sup>, and mapped reads were excluded from the downstream analysis. The taxonomic profiles of bacterial species were estimated using mOTUs (v3.0.1)<sup>53</sup>, and those of viral profiles were obtained using the custom phage database<sup>44</sup>. Functional profiles of the bacteriome based on Kyoto Encyclopedia of Genes and Genomes (KEGG)<sup>54</sup> and CARD<sup>55</sup> annotations were obtained by mapping metagenomic reads to a previously constructed non-redundant gene catalog (Supplementary Methods)<sup>43</sup>.

To profile eukaryotic microbial species in metagenomes, a custom genome database was constructed by combining fungal ( $n = 550$ ) and protozoan ( $n = 79$ ) genomes downloaded from NCBI RefSeq and Genbank, respectively (August 2023). Also, six genomes of *Blastocystis* species, which is abundant and prevalent in the human gut<sup>56</sup>, were additionally downloaded from NCBI GenBank and added to the database (*Blastocystis* subtype 1, 2, 3, 6, 8, and 9) (Supplementary Data 15). Since these fungal and protozoan genomes have possible contaminations from bacterial genomes<sup>56</sup>, we aligned complete bacterial genomes downloaded from NCBI RefSeq ( $n = 4960$ ) to the fungal and protozoan genomes using minimap2<sup>57</sup> and excluded sequences having similarity to the bacterial genomes (>90% identity over 1000 bp).

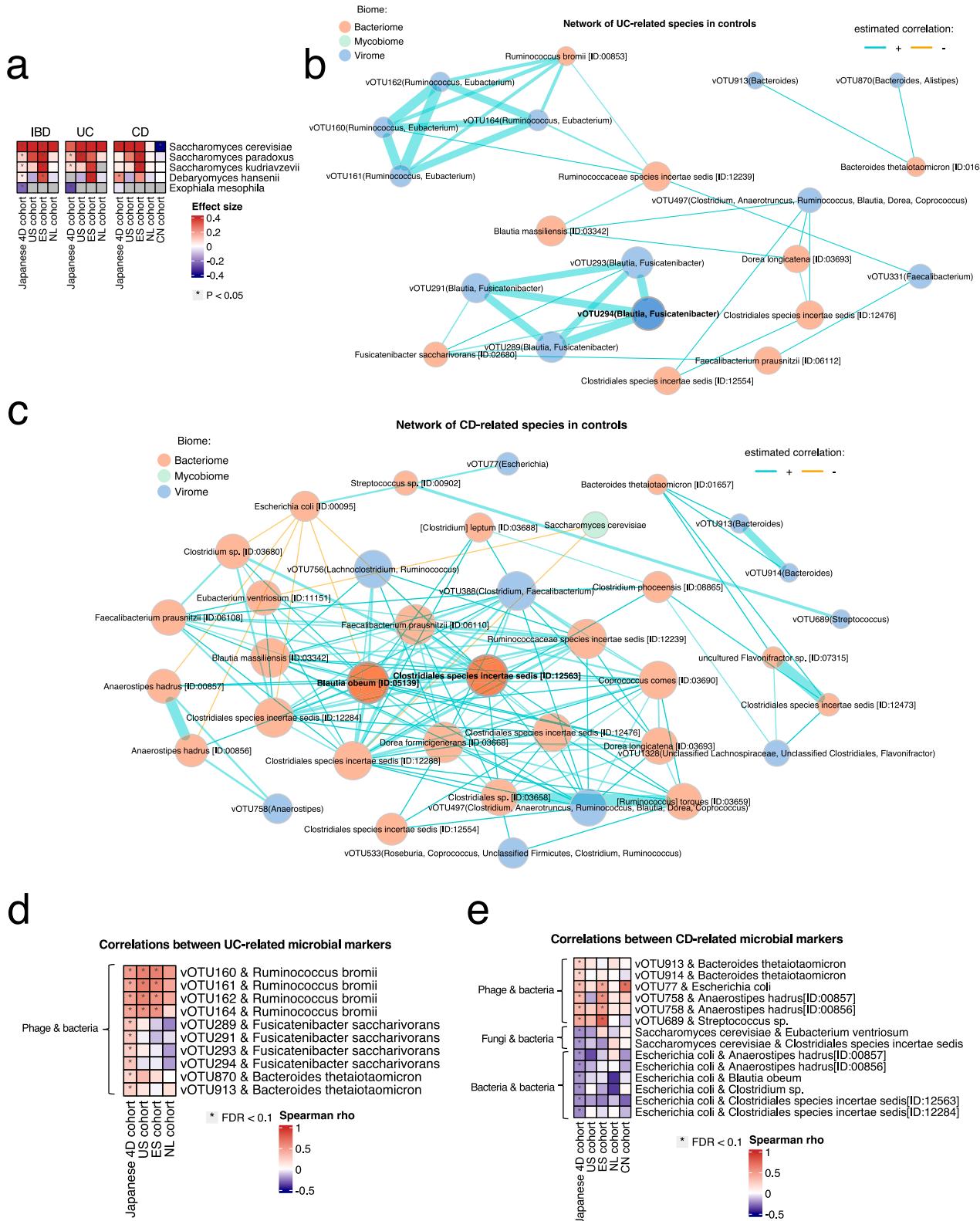
Metagenomic reads were mapped to the custom eukaryotic genome database using bowtie2 with a 95% identity cutoff, and reads per kilobase of exon per million mapped reads for each genome were calculated. To exclude possible false positive detection, we counted the number of reads that mapped to different genomic regions for each eukaryotic species, where the regions were defined by dividing the whole genomes into respective 1000 bp sequences. If more than 10 regions were detected from a species, the species was defined as present in the metagenome (Supplementary Methods).

#### Statistical analysis for the gut microbiome

All statistical analyzes were performed using R (version 4.2.1). The Shannon index was calculated using the diversity function of the vegan package, and MDS analysis based on the Bray-Curtis distances was performed using the ordinate function of the phyloseq package. Permutational analysis of variance was conducted using the adonis2 function of the phyloseq package. Violin plots and box-and-whisker plots were depicted using the ggplot2 package. Values were compared using the Kruskal-Wallis test with a post hoc analysis of multiple comparisons using Dunn's test with the Holm correction.

A Microbiome Multivariable Association with Linear Models version 2.0 package (MaAsLin2) was used for the association analyzes between each multi-biome and IBD (UC and CD) in comparison to healthy controls. For the MaAsLin2 analysis, we included bacterial genera, species, ARGs, KOs, MOs, phages, and fungi/protozoa with mean relative abundances above the thresholds (1E-4, 1E-4, 1E-4, 1E-8, 1E-6, 1E-4, and 1E-10, respectively) and minimum prevalences above the thresholds of 10%, except for eukaryotes (0.5%). In this analysis, FDR less than 0.1 was considered statistically significant except for eukaryotes. Due to the low detection rate of eukaryotic species, a  $P$ -value  $< 0.05$  was considered significant for them. Spearman's rank correlation coefficient was used to assess correlations between features of the gut microbiome, such as bacterial species and KEGG orthologies (KO), and ARGs. Heatmaps were used to show coefficient values in MaAsLin2 or Spearman correlation analysis and displayed using the ComplexHeatmap package (version 2.13.1). Scatterplots were created to compare the coefficient values for each bacterial species between the JP and US cohorts using the ggplot2 and ggpublish packages.

To create a correlation network between the bacteriome, virome, and mycobiome, we used the R package NetCoMi (Network Construction and comparison for Microbiome data)<sup>58</sup>, and Spearman correlation analysis was used to calculate the correlation coefficients. Species included in the network were selected based on the results of MaAsLin2. As for the bacteriome and virome, bacterial species and vOTUs showing absolute values of coefficient  $> 1.0$  and FDR  $< 0.1$  in UC or CD compared with controls were included. For eukaryotes,



those with  $P$ -value  $< 0.05$  were selected. Associations significantly different from zero (Spearman's  $\rho > 0.2$  and  $FDR < 0.1$ ) were described in the network graphs.

#### Statistics & reproducibility

No statistical method was used to predetermine the sample size. Data from patients who met the exclusion criteria (Supplementary Methods) were excluded. The experiments were not randomized, and the

investigators were not blinded to allocation during experiments and outcome assessment. To confirm the reproducibility of our findings, external metagenomic datasets from the US, ES, NL, and CN were analyzed as described above.

#### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

**Fig. 5 | Mycobiome signatures and multi-biome interactions in UC and CD.** **a** The heat map depicts coefficient values of the mycobiome with inflammatory bowel disease (IBD), ulcerative colitis (UC), and Crohn's disease (CD) in comparison to healthy controls in the Japanese 4D cohort, United States (US), Spanish (ES), Netherlands (NL), and Chinese (CN) cohorts. The plot includes only fungi showing  $P$ -value  $< 0.05$  in UC or CD (Japanese 4D cohort) compared with controls based on MaAsLin2 (two-sided). Effect size is the coefficient value obtained by MaAsLin2. Asterisks represent  $P$ -value  $< 0.05$  (MaAsLin2, two-sided). All species are fungi.

**b** The network graph shows intra- and trans-kingdom interactions among the UC-related bacteriome, virome, and mycobiome in healthy controls. **c** The network graph shows intra- and trans-kingdom interactions among the CD-related bacteriome, virome, and mycobiome in healthy controls. The UC- or CD-related bacterial, fungal, and viral species were selected based on the results of MaAsLin2. This analysis included bacterial or viral species showing absolute values of coefficient  $> 1.0$  and false discovery rates (FDR)  $< 0.1$  in **(b)** UC compared with controls and **(c)**

CD compared with controls. Fungal species with  $P$ -value  $< 0.05$  (MaAsLin2, two-sided) in **(b)** UC compared with controls and **(c)** CD compared with controls were selected. The bacteria (genus) in parentheses next to the phage names are expected hosts. If no bacteria are listed in the parentheses, the host is undefined. Associations significantly different from zero (Spearman's rho  $> 0.2$  and FDR  $< 0.1$ ) and important interactions (e.g., linkages between expected hosts and phages, fungi and bacteria, and pathobionts and bacteria) are described in the network graphs. The *turquoise* and *orange* colors of the lines represent positive and negative associations, respectively. The thicker the lines, the higher the correlation. **d–e** This heatmap shows coefficient values assessing the associations among **(d)** UC-related species and **(e)** CD-related species in healthy controls from the Japanese 4D cohort ( $n = 540$ ). This plot includes coefficient values of these interactions obtained from controls in the US ( $n = 34$ ), ES ( $n = 62$ ), NL ( $n = 22$ ), and CN cohorts ( $n = 60$ ). Effect size is the coefficient value obtained from Spearman's correlation analysis. Asterisks represent FDR  $< 0.1$ . Source data are provided as a Source Data file.

## Data availability

All human-removed metagenomic sequencing data ( $n = 682$ ) used in this study have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under the accession numbers PRJNA1169537 ( $n = 654$ ) and PRJNA832909 ( $n = 28$ ). Metagenomic samples used in this study are described in Supplementary Data 2. Based on the Article 27 of the Act on the Protection of Personal Information (Japanese law), metadata cannot be released to public due to privacy reasons. However, if you wish to access the metadata, it is possible to share them as original research after obtaining ethical approval from both our research institute and the requesting institutions. For questions regarding metagenomic data and metadata, please contact the corresponding authors: Suguru Nishijima (nishijima.suguru@gmail.com) for metagenomic data and Naoyoshi Nagata (nnagata\_ncgm@yahoo.co.jp) for metadata. They are expected to respond to requests within four weeks. Once an agreement has been reached between the institutions, the data will be available for any type of research covered by the ethical approval under which the original study was carried out. Access will be granted for as long as necessary for analysis. We have provided all the information we obtained by the analysis to the extent possible, such as all the results for the multivariate analysis in Supplementary Data 1–15. Source data are provided with this paper.

## Code availability

R scripts used for the analyzes and figure generation of this paper are available at <https://github.com/ShintaroAkiyama/IBD-multi-biome>, or with the Zenodo DOI14021369<sup>59</sup>.

## References

- Morgan, X. C. et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. *Genome Biol.* **13**, R79 (2012).
- Franzosa, E. A. et al. Gut microbiome structure and metabolic activity in inflammatory bowel disease. *Nat. Microbiol.* **4**, 293–305 (2019).
- Lewis, J. D. et al. Inflammation, antibiotics, and diet as environmental stressors of the gut microbiome in pediatric crohn's disease. *Cell Host Microbe* **18**, 489–500 (2015).
- Gevers, D. et al. The treatment-naïve microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15**, 382–392 (2014).
- Lloyd-Price, J. et al. Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases. *Nature* **569**, 655–662 (2019).
- Nielsen, H. B. et al. Identification and assembly of genomes and genetic elements in complex metagenomic samples without using reference genomes. *Nat. Biotechnol.* **32**, 822–828 (2014).
- He, Q. et al. Two distinct metacommunities characterize the gut microbiota in Crohn's disease patients. *Gigascience* **6**, 1–11 (2017).
- Palmela, C. et al. Adherent-invasive Escherichia coli in inflammatory bowel disease. *Gut* **67**, 574–587 (2018).
- Seishima, J. et al. Gut-derived Enterococcus faecium from ulcerative colitis patients promotes colitis in a genetically susceptible mouse host. *Genome Biol.* **20**, 252 (2019).
- Dogan, B. et al. Multidrug resistance is common in Escherichia coli associated with ileal Crohn's disease. *Inflamm. Bowel Dis.* **19**, 141–150 (2013).
- Xia, Y. et al. Combined analysis of metagenomic data revealed consistent changes of gut microbiome structure and function in inflammatory bowel disease. *J. Appl. Microbiol.* **131**, 3018–3031 (2021).
- Federici, S. et al. Targeted suppression of human IBD-associated gut microbiota commensals by phage consortia for treatment of intestinal inflammation. *Cell* **185**, 2879–2898 e2824 (2022).
- Balcazar, J. L. Implications of bacteriophages on the acquisition and spread of antibiotic resistance in the environment. *Int. Microbiol.* **23**, 475–479 (2020).
- Manrique, P. et al. Healthy human gut phageome. *Proc. Natl. Acad. Sci. USA* **113**, 10400–10405 (2016).
- Clooney, A. G. et al. Whole-virome analysis sheds light on viral dark matter in inflammatory bowel disease. *Cell Host Microbe* **26**, 764–778.e765 (2019).
- Norman, J. M. et al. Disease-specific alterations in the enteric virome in inflammatory bowel disease. *Cell* **160**, 447–460 (2015).
- Zhang, F., Aschenbrenner, D., Yoo, J. Y. & Zuo, T. The gut mycobiome in health, disease, and clinical applications in association with the gut bacterial microbiome assembly. *Lancet Microbe* **3**, e969–e983 (2022).
- Sokol, H. et al. Fungal microbiota dysbiosis in IBD. *Gut* **66**, 1039–1048 (2017).
- Quinton, J. F. et al. Anti-Saccharomyces cerevisiae mannan antibodies combined with antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease: prevalence and diagnostic role. *Gut* **42**, 788–791 (1998).
- Hoffmann, C. et al. Archaea and fungi of the human gut microbiome: correlations with diet and bacterial residents. *PLoS ONE* **8**, e66019 (2013).
- Liu, Q. et al. Multi-kingdom gut microbiota analyses define COVID-19 severity and post-acute COVID-19 syndrome. *Nat. Commun.* **13**, 6806 (2022).
- Mac Aogain, M. et al. Integrative microbiomics in bronchiectasis exacerbations. *Nat. Med.* **27**, 688–699 (2021).
- Mearin, F. et al. Bowel Disorders. *Gastroenterology* **150**, 1393–1407 (2016).
- Lee, M. & Chang, E. B. Inflammatory bowel diseases (IBD) and the microbiome—searching the crime scene for clues. *Gastroenterology* **160**, 524–537 (2021).

25. Pittayanan, R. et al. Differences in gut microbiota in patients with vs without inflammatory bowel diseases: a systematic review. *Gastroenterology* **158**, 930–946.e931 (2020).
26. De Oliveira, D. M. P. et al. Antimicrobial resistance in ESKAPE pathogens. *Clin. Microbiol. Rev.* **33**, e00181–19 (2020).
27. Puzari, M., Baishya, S., Choudhury, M. D. & Chetia, P. Gene network analysis of efflux pump proteins in *Shigella* spp. *Gene Reports* **21**, 100839 (2020).
28. Tang, D., Kang, R., Coyne, C. B., Zeh, H. J. & Lotze, M. T. PAMPs and DAMPs: signal Os that spur autophagy and immunity. *Immunol. Rev.* **249**, 158–175 (2012).
29. Yang, T. C. et al. Screening of the *Salmonella paratyphi A* CMCC 50973 strain outer membrane proteins for the identification of potential vaccine targets. *Mol. Med. Rep.* **5**, 78–83 (2012).
30. Fitzgerald, C., Sherwood, R., Gheesling, L. L., Brenner, F. W. & Fields, P. I. Molecular analysis of the rfb O antigen gene cluster of *Salmonella enterica* serogroup O:6,14 and development of a serogroup-specific PCR assay. *Appl. Environ. Microbiol.* **69**, 6099–6105 (2003).
31. Heuston, S., Begley, M., Gahan, C. G. M. & Hill, C. Isoprenoid biosynthesis in bacterial pathogens. *Microbiology* **158**, 1389–1401 (2012).
32. Bhagavathula, A. S., Clark, C. & Rahmani, J. Statin use and new-onset of inflammatory bowel disease: a systematic review and meta-analysis of over ten million participants. *Eur. J. Pharmacol.* **891**, 173750 (2021).
33. Sonnenburg, J. L. et al. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* **307**, 1955–1959 (2005).
34. Johansson, M. E., Sjovall, H. & Hansson, G. C. The gastrointestinal mucus system in health and disease. *Nat. Rev. Gastroenterol. Hepatol.* **10**, 352–361 (2013).
35. Alipour, M. et al. Mucosal barrier depletion and loss of bacterial diversity are primary abnormalities in paediatric ulcerative colitis. *J. Crohns Colitis* **10**, 462–471 (2016).
36. Antimicrobial Resistance Collaborators Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. *Lancet* **399**, 629–655 (2022).
37. Singh, K. V., Malathum, K. & Murray, B. E. Disruption of an *Enterococcus faecium* species-specific gene, a homologue of acquired macrolide resistance genes of staphylococci, is associated with an increase in macrolide susceptibility. *Antimicrob. Agents Chemother.* **45**, 263–266 (2001).
38. Martinez-Medina, M. et al. Antimicrobial resistance profiles of adherent invasive *Escherichia coli* show increased resistance to beta-lactams. *Antibiotics* **9**, 251 (2020).
39. Titecat, M. et al. Safety and efficacy of an AIEC-targeted bacteriophage cocktail in a mice colitis model. *J Crohns Colitis* **16**, 1617–1627 (2022).
40. Barnich, N. et al. CEACAM6 acts as a receptor for adherent-invasive *E. coli*, supporting ileal mucosa colonization in Crohn disease. *J. Clin. Invest.* **117**, 1566–1574 (2007).
41. Zhang, S. et al. Short chain fatty acids modulate the growth and virulence of pathosymbiont *Escherichia coli* and host response. *Antibiotics* **9**, 462 (2020).
42. Nagata, N. et al. Metagenomic identification of microbial signatures predicting pancreatic cancer from a multinational study. *Gastroenterology* **163**, 222–238 (2022).
43. Nagata, N. et al. Population-level metagenomics uncovers distinct effects of multiple medications on the human gut microbiome. *Gastroenterology* **163**, 1038–1052 (2022).
44. Nishijima, S. et al. Extensive gut virome variation and its associations with host and environmental factors in a population-level cohort. *Nat. Commun.* **13**, 5252 (2022).
45. Nagata, N. et al. Combined endoscopy, aspiration, and biopsy analysis for identifying infectious colitis in patients with ileocecal ulcers. *Clin. Gastroenterol. Hepatol.* **11**, 673–680.e672 (2013).
46. Nagata, N. et al. Effects of storage temperature, storage time, and Cary-Blair transport medium on the stability of the gut microbiota. *Drug Discov. Ther.* **13**, 256–260 (2019).
47. Nagata, N. et al. Effects of bowel preparation on the human gut microbiome and metabolome. *Sci. Rep.* **9**, 4042 (2019).
48. D'Agostino, R. B. Jr. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. *Stat. Med.* **17**, 2265–2281 (1998).
49. Nishijima, S. et al. The gut microbiome of healthy Japanese and its microbial and functional uniqueness. *DNA Res.* **23**, 125–133 (2016).
50. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic datasets. *Bioinformatics* **27**, 863–864 (2011).
51. Kechin, A., Boyarskikh, U., Kel, A. & Filipenko, M. cutPrimers: a new tool for accurate cutting of primers from reads of targeted next generation sequencing. *J Comput. Biol.* **24**, 1138–1143 (2017).
52. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
53. Ruscheweyh, H. J. et al. Cultivation-independent genomes greatly expand taxonomic-profiling capabilities of mOTUs across various environments. *Microbiome* **10**, 212 (2022).
54. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592 (2023).
55. Alcock, B. P. et al. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **48**, D517–D525 (2020).
56. Beghini, F. et al. Large-scale comparative metagenomics of Blastoocystis, a common member of the human gut microbiome. *ISME J.* **11**, 2848–2863 (2017).
57. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
58. Peschel, S., Muller, C. L., von Mutius, E., Boulesteix, A. L. & Depner, M. NetCoMi: network construction and comparison for microbiome data in R. *Brief. Bioinform.* **22**, bbaa290 (2021).
59. Akiyama, S. ShintaroAkiyama/IBD-multi-biome: IBD multi-biome analysis (v1.0.0). Zenodo. <https://doi.org/10.5281/zenodo.14021369> (2024).

## Acknowledgements

We thank Tohru Miyoshi-Akiyama, Takashi Hamabata, and Norihiko Takemoto (National Center for Global Health and Medicine) for DNA extraction and shotgun sequencing, Oleksandr Maistrenko and Chan Yeong Kim (EMBL) for discussion about the analysis of eukaryotic species, and Thane Doss for English writing assistance. This work was partially supported by Japan Agency for Medical Research and Development (AMED) (Research Program on HIV/AIDS: JP22fk0410051, and Research Program on Emerging and Re-emerging Infectious Diseases: JP22fk0108538), Grants-in-Aid for Research from the National Center for Global Health and Medicine (28-2401, 19A1011, 24A1010), JSPS KAKENHI Grant (JP17K09365 and 20K08366), Smoking Research Foundation, DANONE RESEARCH GRANT, research funding of Japan Dairy Association (J-Milk), Takeda Science Foundation, Tokyo Medical University Foundation, and COCKPIT funding. The funding bodies had no role in the study design, data collection or analysis, decision to publish, or preparation of the manuscript.

## Author contributions

N.N. and S.N. contributed equally to this work. N.N. is a principal investigator of the Japanese 4D cohort database. S.A., N.N., S.N., and M.H. designed the study. S.A., N.N., and S.N. wrote the manuscript. N.N. constructed metadata and integrated the drug information. S.N. performed bioinformatic analysis. S.A., S.N., and N.N. conducted statistical analyzes. Y.K., M.K., K.U., M.O., M.M., M.H., K.T., N.U., and T.K. participated in the study's conception and edited the manuscript. M.H. supervised the study. P.B. provided external datasets in previous

studies. All authors have read and approved the submitted version of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at  
<https://doi.org/10.1038/s41467-024-54797-8>.

**Correspondence** and requests for materials should be addressed to Suguru Nishijima or Naoyoshi Nagata.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at  
<http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2024

<sup>1</sup>Department of Gastroenterology, Institute of Medicine, University of Tsukuba, Ibaraki, Japan. <sup>2</sup>Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. <sup>3</sup>Department of Gastroenterology and Hepatology, National Center for Global Health and Medicine, Tokyo, Japan. <sup>4</sup>Department of Clinical Research Strategic Planning Center for Clinical Sciences, National Center for Global Health and Medicine, Tokyo, Japan.

<sup>5</sup>Department of Diabetes, Endocrinology and Metabolism, Center Hospital, National Center for Global Health and Medicine, Tokyo, Japan. <sup>6</sup>Diabetes Research Center, Research Institute, National Center for Global Health and Medicine, Chiba, Japan. <sup>7</sup>Genome Medical Sciences Project, Research Institute, National Center for Global Health and Medicine, Chiba, Japan. <sup>8</sup>Laboratory for Microbiome Sciences, RIKEN Center for Integrative Medical Sciences, Yokohama, Japan. <sup>9</sup>Department of Gastroenterology and Hepatology, National Center for Global Health and Medicine, Kohnodai Hospital, Chiba, Japan. <sup>10</sup>Department of Gastroenterological Endoscopy, Tokyo Medical University, Tokyo, Japan. <sup>11</sup>These authors contributed equally: Suguru Nishijima, Naoyoshi Nagata.

✉ e-mail: [nishijima.suguru@gmail.com](mailto:nishijima.suguru@gmail.com); [nnagata\\_ncgm@yahoo.co.jp](mailto:nnagata_ncgm@yahoo.co.jp)