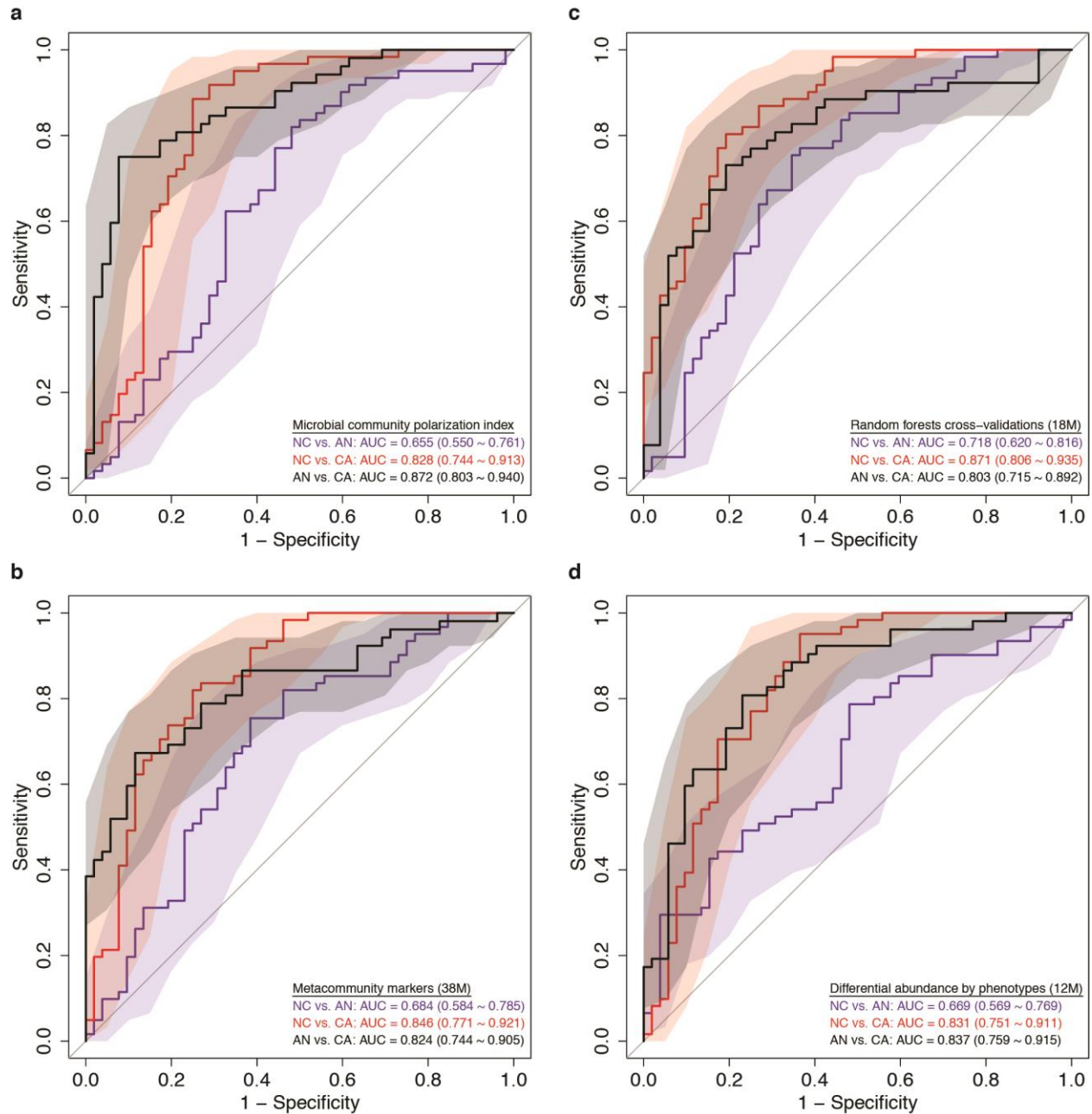
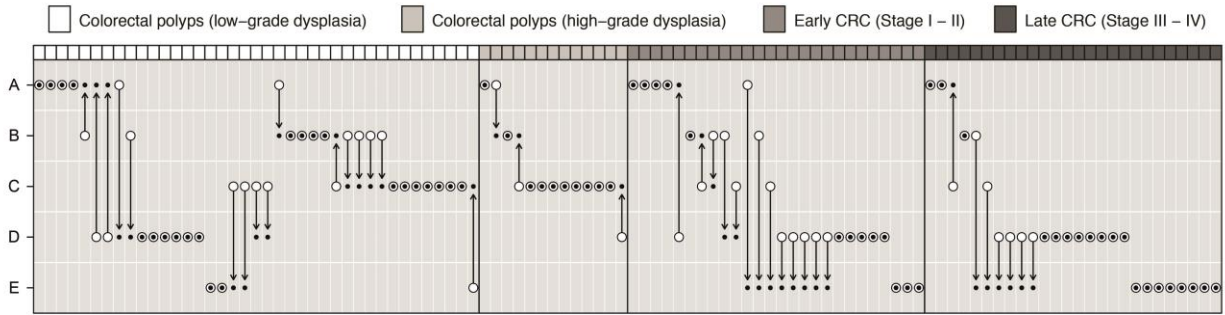


**Supplementary Figure 1 | Dimension reduction of metacommunities associated with colorectal carcinogenesis.** (a) Robustness of two widely used partitioning approaches in response to varying rarity thresholds. (b) Comparison of DMM- (top) and PAM-based (bottom) detection of microbial community clusters using non-metric multidimensional scaling (NMDS) of Jensen-Shannon Divergence distance matrices. (c) NMDS ordination of representative samples from DMM-based approach using relative abundance profiles of metacommunity markers. Metacommunities are represented by 80% confidence ellipses.

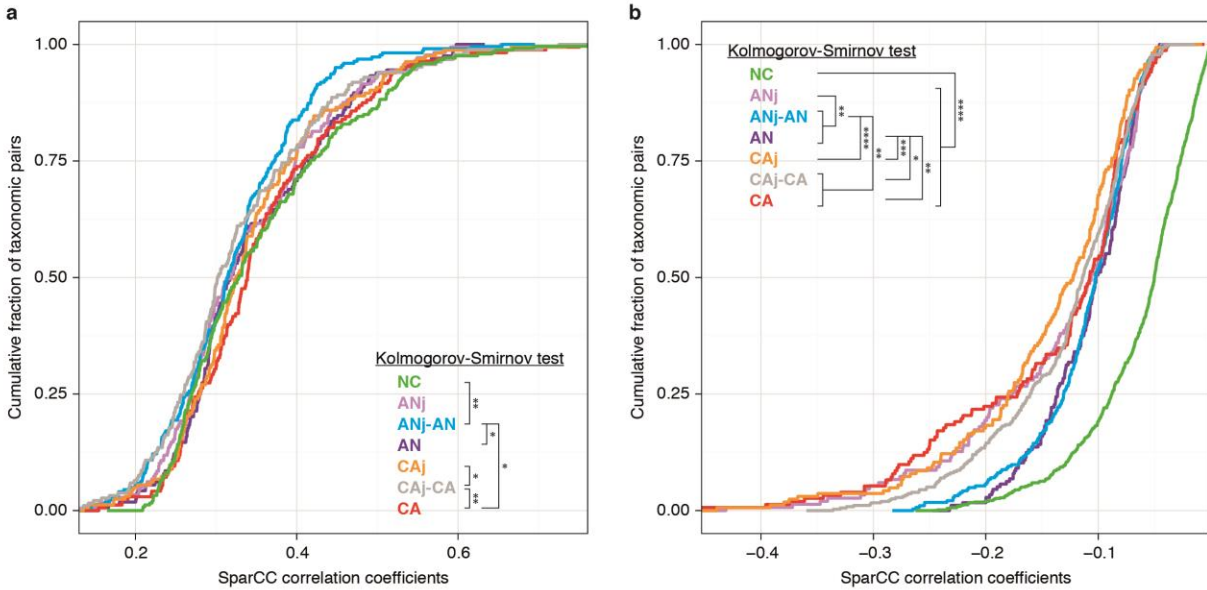


**Supplementary Figure 2 | Microbiome-based classification of colorectal tumour statuses.**

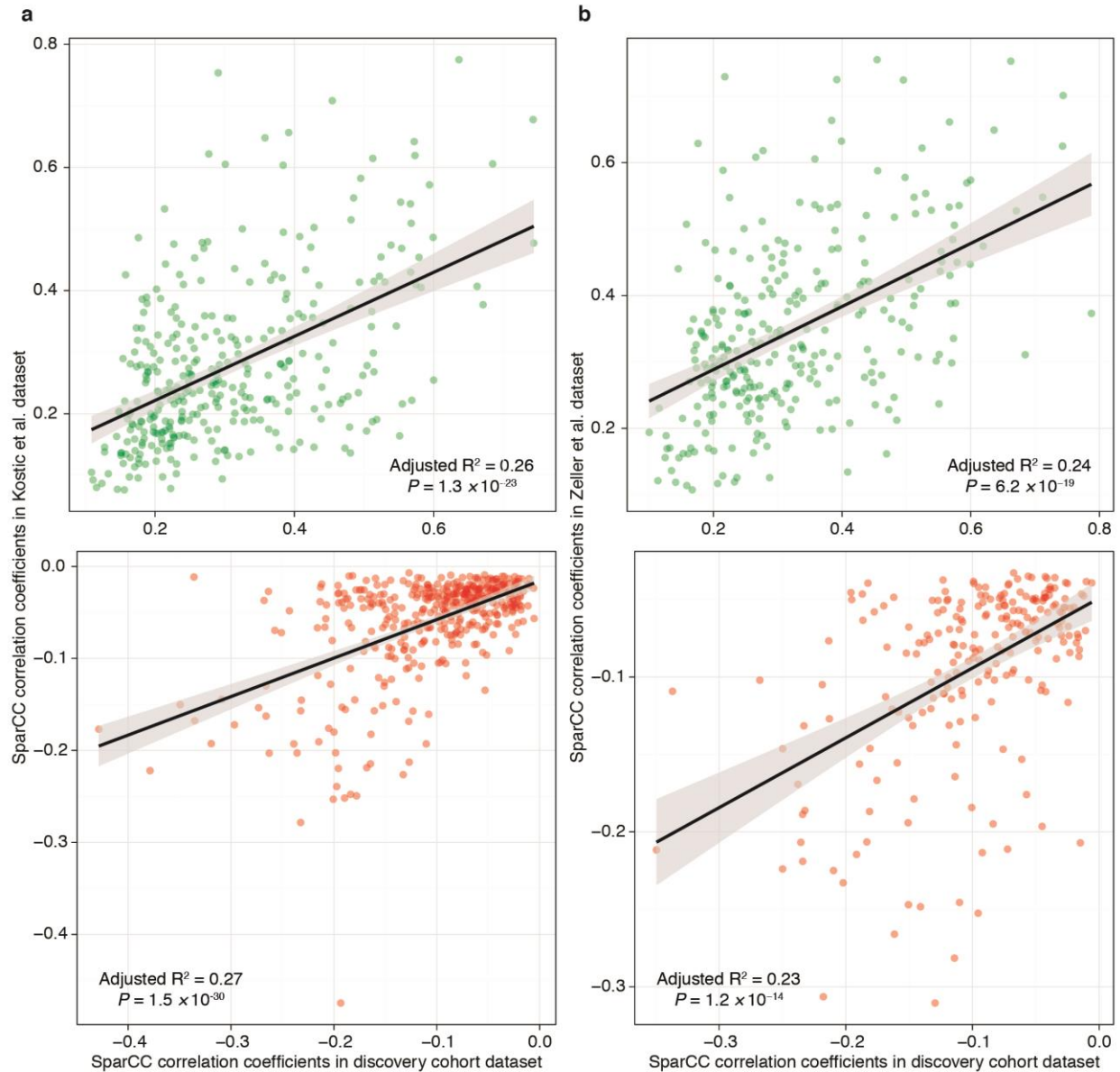
Receiver operating characteristic (ROC) analyses of (a) Microbial Community Polarization index (MCPI) and (b-d) LASSO classifier performance based on bacterial phylotypes that were identified by (b) DMM community typing, (c) 100 iterations of the ten-fold Random Forests cross-validations, and (d) differential abundance analysis using the LEfSe algorithm. AUC values are shown with 95% confidence intervals (shaded-area). NC, normal control; AN, adenoma; CA, carcinoma.



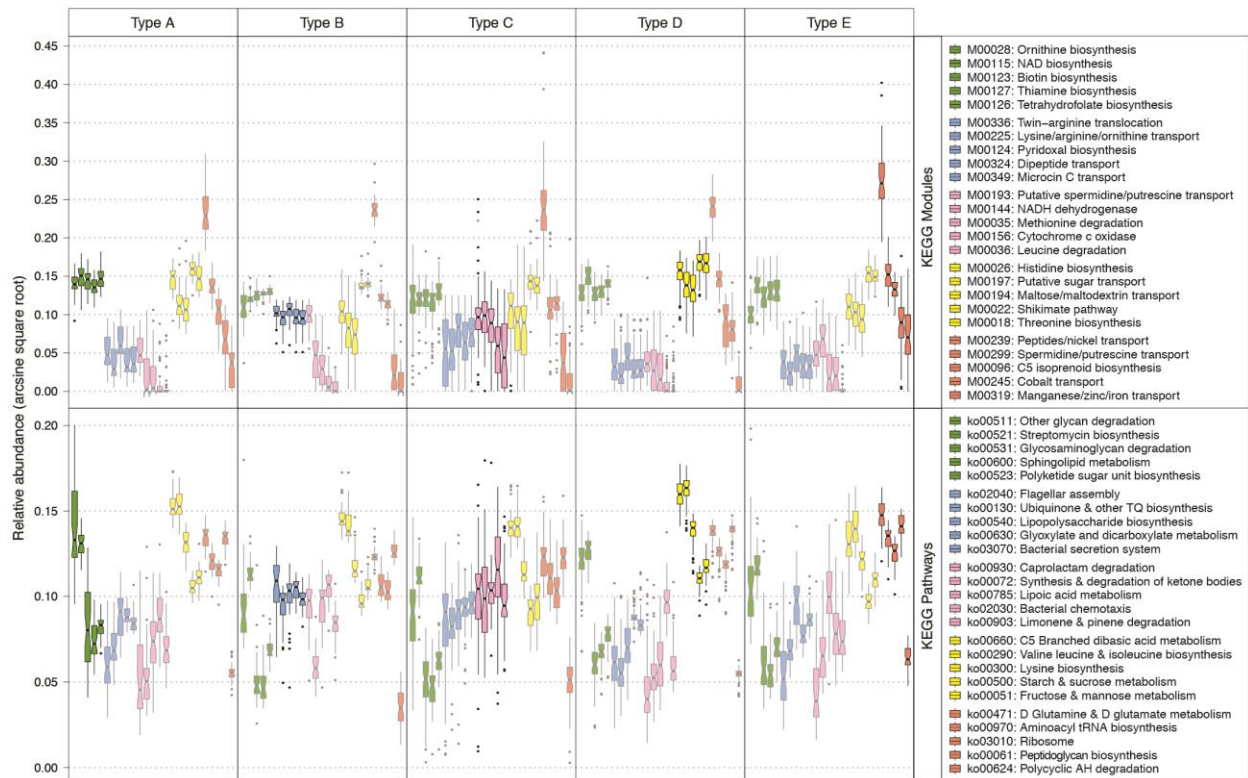
**Supplementary Figure 3 | Alterations of mucosal community types at lesions relative to adjacent normal mucosae along the adenoma-carcinoma sequence.** Open and closed circles represent lesion-adjacent mucosae and lesions, respectively.



**Supplementary Figure 4 | Cumulative distribution functions comparing the differences in taxonomic occurrences between disease-states.** Correlation coefficients with statistical significances (FDR < 0.1) were selected for visualization. Distances between two-sample distributions were assessed separately for each group of (a) positive and (b) negative correlations by Kolmogorov-Smirnov tests. *P*-values were adjusted by BH step-up procedure; \*  $q < 0.05$ ; \*\*  $q < 0.01$ ; \*\*\*  $q < 0.001$ ; \*\*\*\*  $q < 0.0001$ . NC, normal control; ANj, adenoma-adjacent; AN, adenoma; CAj, carcinoma-adjacent; CA, carcinoma.



**Supplementary Figure 5 | Reproducibility of ecological interactions within cancer-niches.** Re-analysis of taxonomic relationships revealed analogous patterns of statistically significant interactions in: **(a)** Kostic et al. dataset, and **(b)** Zeller et al. dataset. Shown are correlation coefficients (green for positive; red for negative) with concordant directions and false discovery rates of 0.25 or less between two studies. Adjusted R<sup>2</sup> and *p*-values are from multiple linear regression analyses.



**Supplementary Figure 6 | Representative functional modules and pathways that are enriched in metacommunities.** Modules or pathways are listed in the descending order of LDA scores from top to bottom in the legend keys and from left to right in the boxplot panels. Gene families are represented by the metacommunity in which they are overrepresented; underrepresented families are greyed out to provide contrasts for visualization. The heights of boxes show the interquartile range (IQR) between the first and third quartiles in which medians are bolded; minimum and maximum values are denoted by whiskers; closed-circles are outliers.

## Nucleotide & amino acid metabolism

- A: Arginine & proline metabolism**
  - M00028: Ornithine biosynthesis
- B: Alkaloid & other secondary metabolite biosynthesis**
  - M00033: Ectoine biosynthesis
- C: Lysine metabolism**
  - M00016: Lysine biosynthesis
- D: Serine & threonine metabolism**
  - M00018: Threonine biosynthesis
- E: Histidine metabolism**
  - M00026: Histidine biosynthesis
  - M00045: Histidine degradation
- F: Branched-chain amino acid metabolism**
  - M00019: Leucine biosynthesis
  - M00036: Leucine degradation
- G: Polyamine biosynthesis**
  - M00133: Polyamine biosynthesis
  - M00136: Prokaryotic GABA biosynthesis
- H: Aromatic amino acid metabolism**
  - M00025: Tyrosine biosynthesis
  - M00022: Shikimate pathway
- I: Pyrimidine metabolism**
  - M00051: Uridine monophosphate biosynthesis
  - M00053: Pyrimidine deoxyribonucleotide biosynthesis
  - M00046: beta-Alanine biosynthesis
- J: Purine metabolism**
  - M00050: Guanine nucleotide biosynthesis
  - M00048: Inosine monophosphate biosynthesis
  - M00049: Adenine nucleotide biosynthesis
- K: Cysteine & methionine metabolism**
  - M00017: Methionine biosynthesis
  - M00035: Methionine degradation
  - M00034: Methionine salvage pathway
- Co-factor & vitamin biosynthesis**
  - M00121: Heme biosynthesis
  - M00123: Biotin biosynthesis
  - M00122: Cobalamin biosynthesis
  - M00125: Riboflavin biosynthesis
  - M00124: Pyridoxal biosynthesis
  - M00127: Thiamine biosynthesis
  - M00128: Tetrahydrofolate biosynthesis
  - M00119: Pantothenate biosynthesis
  - M00115: NAD biosynthesis
  - M00116: Menaquione biosynthesis
  - M00117: Prokaryotic ubiquinone biosynthesis

## g: Carbohydrate & lipid metabolism (structure)

- L: Carbohydrate metabolism**
  - M00310: Pyruvate ferredoxin oxidoreductase
  - M00311: 2-oxoglutarate ferredoxin oxidoreductase

## i: Genetic information processing

- M: DNA polymerase**
  - M00260: Bacterial DNA polymerase III complex

## N: Proteasome

- M00342: Bacterial proteasome

## O: Ribosome

- M00178: Bacterial ribosome

## h: Energy metabolism (structure)

- P: Photosynthesis**
  - M00164: ATP synthase
- ATP synthesis**
  - M00155: Complex IV, cytochrome c ubiquinol oxidase
  - M00156: Complex IV, cytochrome c oxidase, cbb3-type
  - M00157: Bacterial F-type ATPase
  - M00150: Complex II, fumarate reductase
  - M00144: Complex I, NADH dehydrogenase I
  - M00149: Complex II, succinate dehydrogenase
  - M00159: Prokaryotic V-type ATPase

## Environmental information processing

- Q: Phosphotransferase system**
  - M00276: Sorbose-specific II component
  - M00287: Galactosamine-specific II component
  - M00277: N-acetyl-galactosamine-specific II component
  - M00278: Mannose-specific II component

## R: Peptide & nickel transport

- M00349: Microcin C transport
- M00348: Glutathione transport
- M00239: Peptides/nickel transport
- M00324: Dipeptide transport

## S: Metallic cation, iron-siderophore & vitamin B12 transport

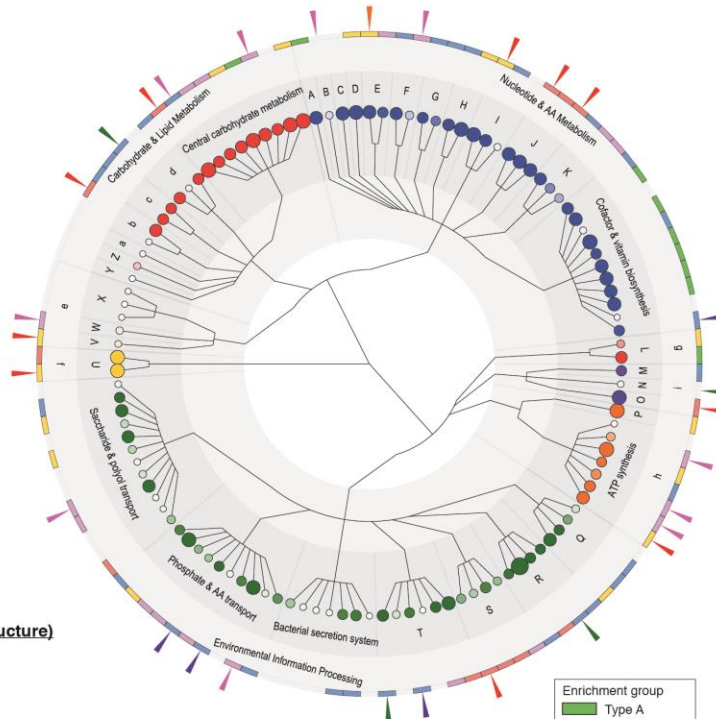
- M00317: Manganese/iron transport
- M00245: Cobalt transport
- M00246: Nickel transport
- M00319: Manganese/zinc/iron transport

## T: Mineral & organic ion transport

- M00299: Spermidine/putrescine transport
- M00193: Putative spermidine/putrescine transport
- M00301: Mannopine transport
- M00300: Putrescine transport
- M00302: 2-Aminoethylphosphonate transport
- M00185: Sulfate transport

## Bacterial secretion system

- M00334: Type VI secretion
- M00335: Sec system
- M00336: Twin-arginine translocation
- M00326: RTX toxin transport
- M00332: Type III secretion
- M00333: Type IV secretion
- M00331: Type II general secretion



## Environmental information processing

- Phosphate & amino acid transport**
  - M00226: Histidine transport
  - M00235: Arginine/ornithine transport
  - M00237: Branched-chain amino acid transport
  - M00225: Lysine/arginine/ornithine transport
  - M00231: Octopine/nopaline transport
  - M00230: Glutamate/aspartate transport
  - M00233: Glutamate transport
  - M00232: General L-amino acid transport
  - M00222: Phosphate transport
  - M00229: Arginine transport
- Saccharide & polyol transport**
  - M00219: Al-2 transport
  - M00204: Trehalose/maltose transport
  - M00206: Cellobiose transport
  - M00207: Multiple sugar transport
  - M00200: Sorbitol/mannitol transport
  - M00201: alpha-Glucoside transport
  - M00202: Oligogalacturonide transport
  - M00194: Maltose/maltodextrin transport
  - M00220: Rhamnose transport
  - M00197: Putative sugar transport
  - M00198: sn-Glycerol 3-phosphate transport
  - M00199: Lactose/L-arabinose transport

## f: Metabolism

- U: Aminoacyl tRNA**
  - M00359: Eukaryotic Aminoacyl-tRNA biosynthesis
  - M00360: Prokaryotic Aminoacyl-tRNA biosynthesis

## e: Energy metabolism

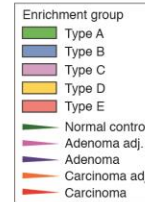
- V: Sulfur metabolism**
  - M00176: Sulfur reduction

- W: Methane metabolism**
  - M00174: Methane oxidation

- X: Carbon fixation**
  - M00166: Reductive pentose phosphate cycle
  - M00172: C4-dicarboxylic acid cycle
  - M00173: Reductive citric acid cycle

## Carbohydrate & lipid metabolism

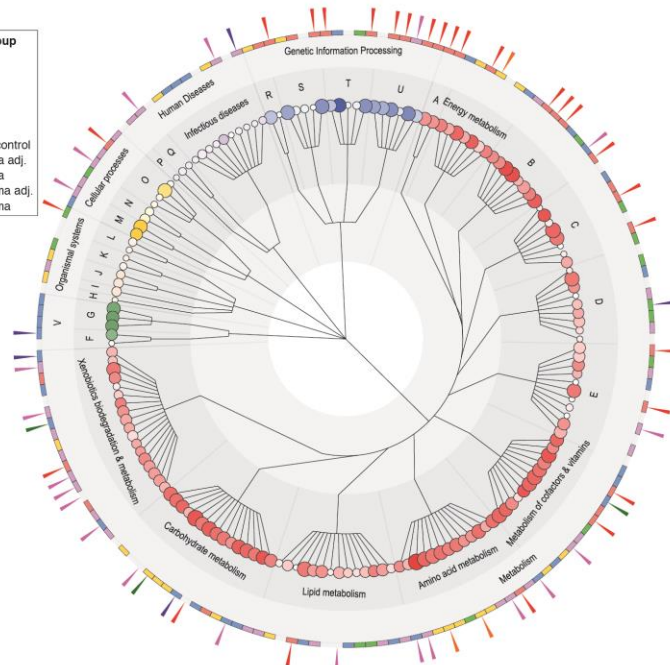
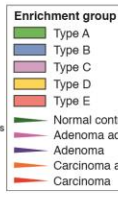
- Y: Lipid metabolism**
  - M00089: Triacylglycerol biosynthesis
- Z: Fatty acid metabolism**
  - M00088: Ketone body biosynthesis
- a: Glycosaminoglycan metabolism**
  - M00077: Chondroitin sulfate degradation
- b: Terpenoid backbone biosynthesis**
  - M00095: Mevalonate C5 isoprenoid biosynthesis
  - M00096: Non-mevalonate C5 isoprenoid biosynthesis
- c: Lipopolysaccharide metabolism**
  - M00090: Lipopolysaccharide biosynthesis
  - M00064: ADP-L-glycero-D-manno-heptose biosynthesis
- d: Other carbohydrate metabolism**
  - M00061: Uronic acid metabolism
  - M00373: Ethylmalonyl pathway
  - M00012: Glyoxylate cycle
- Central carbohydrate metabolism**
  - M00002: Glycolysis, 3C core module
  - M00008: Entner-Doudoroff pathway
  - M00009: TCA cycle
  - M00006: Oxidative pentose phosphate pathway
  - M00007: Non-oxidative pentose phosphate pathway
  - M00004: Pentose phosphate cycle
  - M00011: TCA cycle, second carbon oxidation
  - M00003: Gluconeogenesis
  - M00001: Embden-Meyerhof pathway



**Supplementary Figure 7 | Summary cladogram of differentially abundant KEGG modules imputed.** Node size and transparency represent the total relative abundance and prevalence of a module, respectively. Clades and nodes are annotated in a clockwise manner. Functional categories at level 2 of the BRITE module hierarchy are distinguished by respective node colors. Node color intensity is proportional to the coverage of a module. Inner and outer ring indicate differential enrichments of modules based on metacommunity and mucosal phenotypes, respectively.

## Metabolism

- A: Nucleotide metabolism**
  - ko0230: Purine metabolism
  - ko0240: Pyrimidine metabolism
- Energy metabolism**
  - ko0920: Sulfur metabolism
  - ko0190: Oxidative phosphorylation
  - ko00710: Photosynthetic carbon fixation
  - ko00800: Methane metabolism
  - ko00720: Prokaryotic carbon fixation pathways
  - ko00195: Photosynthesis
  - ko00910: Nitrogen metabolism
- B: Metabolism of other amino acids**
  - ko0480: Glutathione metabolism
  - ko0430: Taurine & hypotaurine metabolism
  - ko0473: D-alanine metabolism
  - ko0471: D-glutamine & D-glutamate metabolism
  - ko0440: Phosphate & phosphite metabolism
  - ko0460: Cyanoamino acid metabolism
  - ko0410: beta Alanine metabolism
  - ko0450: Selenocompound metabolism
- C: Glycan biosynthesis & metabolism**
  - ko0601: Glycosphingolipid biosynthesis lacto & neolacto series
  - ko0540: Lipopolysaccharide biosynthesis
  - ko0510: N-glycan biosynthesis
  - ko0550: Peptidoglycan biosynthesis
  - ko0511: Other glycan degradation
  - ko0563: Glycosylphosphatidylinositol anchor biosynthesis
  - ko0513: Various types of N-glycan biosynthesis
  - ko0531: Glycosaminoglycan degradation
- D: Biosynthesis of other secondary metabolites**
  - ko0901: Indole alkaloid biosynthesis
  - ko0521: Streptomycin biosynthesis
  - ko0090: Tropane piperidine & pyridine alkaloid biosynthesis
  - ko00331: Clavulanic acid biosynthesis
  - ko00311: Penicillin & cephalosporin biosynthesis
  - ko00312: beta Lactam resistance
  - ko00940: Phenylpropanoid biosynthesis
  - ko00941: Flavonoid biosynthesis
  - ko00943: Isoflavonoid biosynthesis
- E: Metabolism of terpenoids & polyketides**
  - ko00608: Zeatin biosynthesis
  - ko00293: Polyketide sugar unit biosynthesis
  - ko00281: Geraniol degradation
  - ko1053: Biosynthesis of siderophore group nonribosomal peptides
  - ko00909: Sesquiterpenoid biosynthesis
  - ko00900: Terpenoid backbone biosynthesis
  - ko1056: Biosynthesis of type II polyketide backbone
  - ko00906: Carotenoid biosynthesis
  - ko00522: Biosynthesis of 12-, 14-, & 16-membered macrolides
  - ko00903: Limonene & pinene degradation
- Metabolism of cofactors & vitamins**
  - ko00253: Tetracycline biosynthesis
  - ko00750: Vitamin B6 metabolism
  - ko00130: Ubiquinone & other terpenoid quinone biosynthesis
  - ko00730: Thiamine metabolism
  - ko00740: Riboflavin metabolism
  - ko00790: Folate biosynthesis
  - ko00870: One carbon pool by folate
  - ko00780: Biotin metabolism
  - ko00785: Lipic acid metabolism
  - ko00830: Retinol metabolism
  - ko00960: Porphyrin & chlorophyll metabolism
  - ko00760: Nicotinate & nicotinamide metabolism
  - ko00770: Pantothenate & CoA biosynthesis
- Amino acid metabolism**
  - ko00290: Glycine serine & threonine metabolism
  - ko00310: Lysine degradation
  - ko00270: Cysteine & methionine metabolism
  - ko00350: Tyrosine metabolism
  - ko00280: Valine leucine & isoleucine degradation
  - ko00400: Phenylalanine tyrosine & tryptophan biosynthesis
  - ko00330: Arginine & proline metabolism
  - ko00340: Histidine metabolism
  - ko00300: Lysine biosynthesis
  - ko00250: Alanine aspartate & glutamate metabolism
  - ko00290: Valine leucine & isoleucine biosynthesis
  - ko00380: Tryptophan metabolism
  - ko00360: Phenylalanine metabolism
- Lipid Metabolism**
  - ko00590: Arachidonic acid metabolism
  - ko00564: Glycerophospholipid metabolism
  - ko00072: Synthesis & degradation of ketone bodies
  - ko00900: Sphingolipid metabolism
  - ko00140: Steroid hormone biosynthesis
  - ko00592: Alpha linolenic acid metabolism
  - ko01040: Biosynthesis of unsaturated fatty acids
  - ko00100: Steroid biosynthesis
  - ko00071: Fatty acid metabolism
  - ko00561: Glycerolipid metabolism
  - ko00061: Fatty acid biosynthesis
  - ko00062: Mitochondrial fatty acid elongation
  - ko00591: Linoleic acid metabolism
  - ko00565: Ether lipid metabolism
- Carbohydrate metabolism**
  - ko00940: Propanoate metabolism
  - ko00020: TCA cycle
  - ko00630: Glyoxylate & dicarboxylate metabolism
  - ko00051: Fructose & mannose metabolism
  - ko00920: Pyruvate metabolism
  - ko00053: Ascorbate & aldarate metabolism
  - ko00520: Amino sugar & nucleotide sugar metabolism
  - ko00040: Pentose & glucuronate interconversions
  - ko00030: Pentose phosphate pathway
  - ko00500: Starch & sucrose metabolism
  - ko00660: C5-branched dibasic acid metabolism
  - ko00562: Inositol phosphate metabolism
  - ko00650: Butanoate metabolism
  - ko00052: Galactose metabolism
  - ko00010: Glycolysis & gluconeogenesis



## Metabolism

- Xenobiotics biodegradation & metabolism**
  - ko00362: Benzoate degradation
  - ko00623: Toluene degradation
  - ko00626: Naphthalene degradation
  - ko00642: Ethylbenzene degradation
  - ko00625: Chloroalkane & chloroalkene degradation
  - ko00643: Styrene degradation
  - ko00791: Atrazine degradation
  - ko00624: Polycyclic aromatic hydrocarbon degradation
  - ko00627: Aminobenzoate degradation
  - ko00363: Bisphenol degradation
  - ko00630: Caprolactam degradation
  - ko00633: Nitrotoluene degradation
  - ko00364: Fluorobenzoate degradation
  - ko00351: 1, 1, 1-Trichloro-2,2-bis-4-chlorophenyl ethane DDT degradation
  - ko00621: Dioxin degradation
  - ko00683: Drug metabolism other enzymes
  - ko00361: Chlorocyclohexane & chlorobenzene degradation
  - ko00622: Xylene degradation

## V: Environmental information processing

- F: Signal transduction**
  - ko04310: Wnt signaling pathway
  - ko02320: Two component system
- G: Membrane transport**
  - ko03070: Bacterial secretion system
  - ko02060: Phosphotransferase system
  - ko02010: ABC transporters
- Organismal system**
  - H: Excretory system**
    - ko04962: Vasopressin regulated water reabsorption
  - I: Environmental adaptation**
    - ko04626: Plant pathogen interaction
  - J: Endocrine system**
    - ko03320: PPAR signaling pathway
    - ko04910: Insulin signaling pathway
  - K: Digestive system**
    - ko04974: Protein digestion & absorption
    - ko04976: Bile secretion
  - L: Immune system**
    - ko04622: RIG-I-like receptor signaling pathway
    - ko04621: NOD-like receptor signaling pathway

## Cellular processes

- M: Cell motility**
  - ko02040: Flagellar assembly
  - ko02030: Bacterial chemotaxis
- N: Transport & catabolism**
  - ko04146: Peroxisome
  - ko04142: Lysosome
- O: Cell growth & death**
  - ko04115: p53 signaling pathway
  - ko04210: Apoptosis
  - ko04112: Cell cycle Caulobacter
  - ko04113: Meiosis yeast

## Human diseases

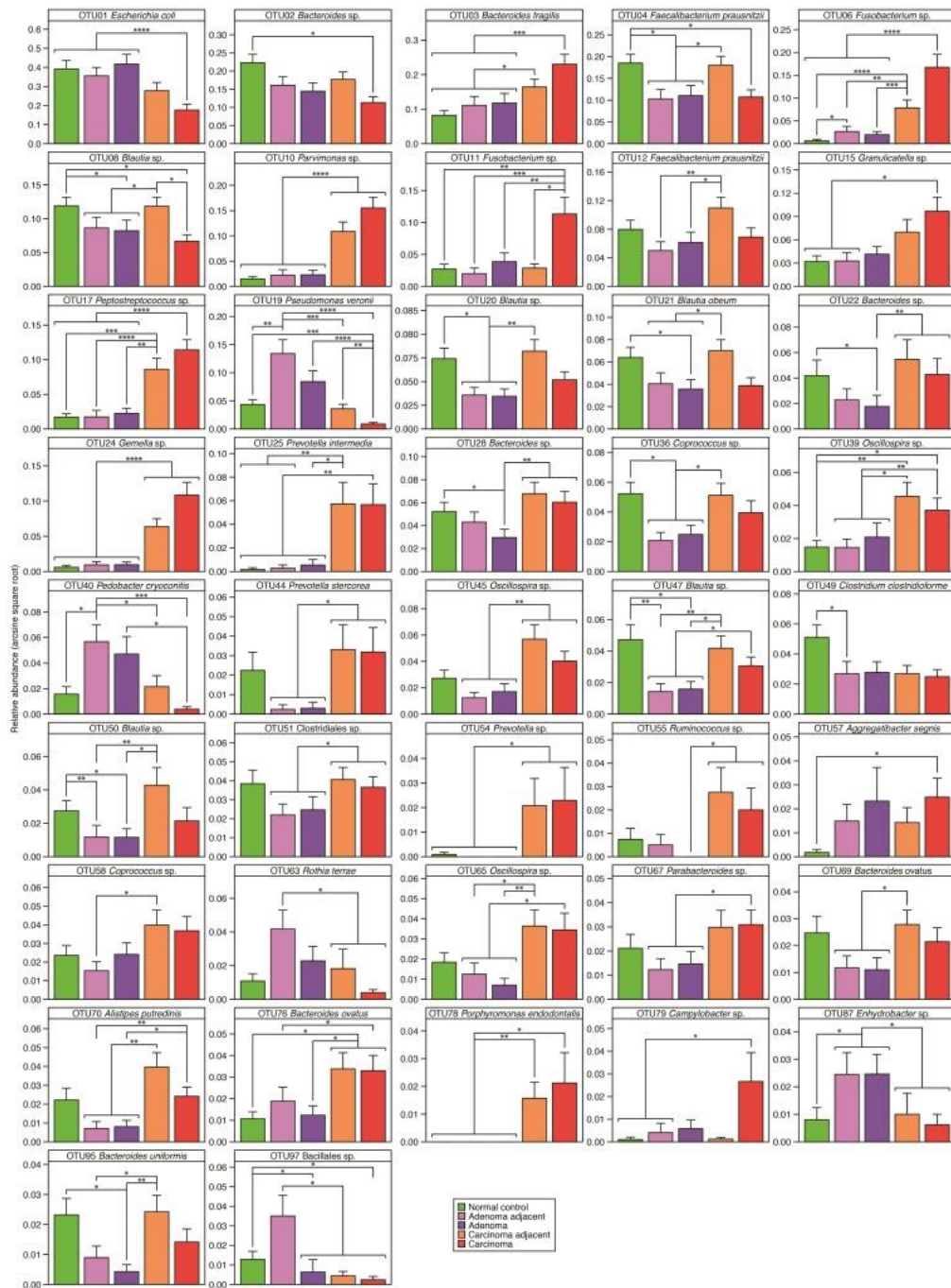
- P: Immune system diseases**
  - ko05322: Systemic lupus erythematosus
- Q: Neurodegenerative diseases**
  - ko05010: Alzheimer's disease
  - ko05012: Parkinson's disease
- Infectious diseases**
  - ko05142: Chagas disease American trypanosomiasis
  - ko05120: Epithelial cell signaling in Helicobacter pylori infection
  - ko05143: African trypanosomiasis
  - ko05100: Bacterial invasion of epithelial cells
  - ko05111: Vibrio cholerae pathogenic cycle
  - ko05131: Shigellosis
  - ko05146: Amoebiasis
  - ko05145: Toxoplasmosis
  - ko05110: Vibrio cholerae infection
  - ko05150: Staphylococcus aureus infection

## Genetic information processing

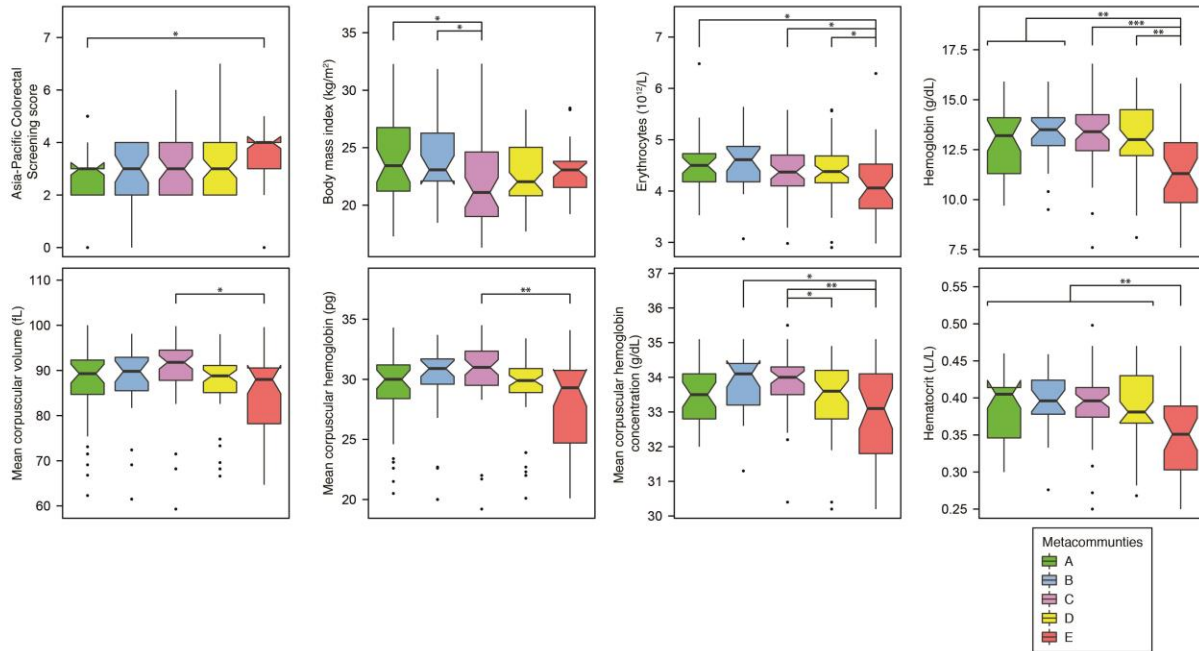
- R: Transcription**
  - ko03020: RNA polymerase
  - ko03022: Basal transcription factors
- S: Translation**
  - ko03010: Ribosome
  - ko03013: RNA transport
  - ko03008: Ribosome biogenesis in eukaryotes
  - ko03015: mRNA surveillance pathway
  - ko00970: Aminoacyl tRNA biosynthesis
- T: Folding, sorting, & degradation**
  - ko03018: RNA degradation
  - ko04122: Sulfur relay system
  - ko03050: Proteasome
  - ko04141: Protein processing in endoplasmic reticulum
  - ko03060: Protein export
- U: Replication & repair**
  - ko03410: Base excision repair
  - ko03030: DNA replication
  - ko03440: Homologous recombination
  - ko03450: Non homologous end joining
  - ko00430: Mismatch repair
  - ko03420: Nucleotide excision repair

**Supplementary Figure 8 | Summary cladogram of differentially abundant KEGG pathways imputed.** Node size and transparency represent the total relative abundance and prevalence of a pathway, respectively. Clades and nodes are annotated in a clockwise manner. Functional categories at level 1 of the BRITE pathway hierarchy are distinguished by respective node colors. Inner and outer ring indicate differential enrichment of modules based on metacommunity and mucosal phenotypes, respectively.

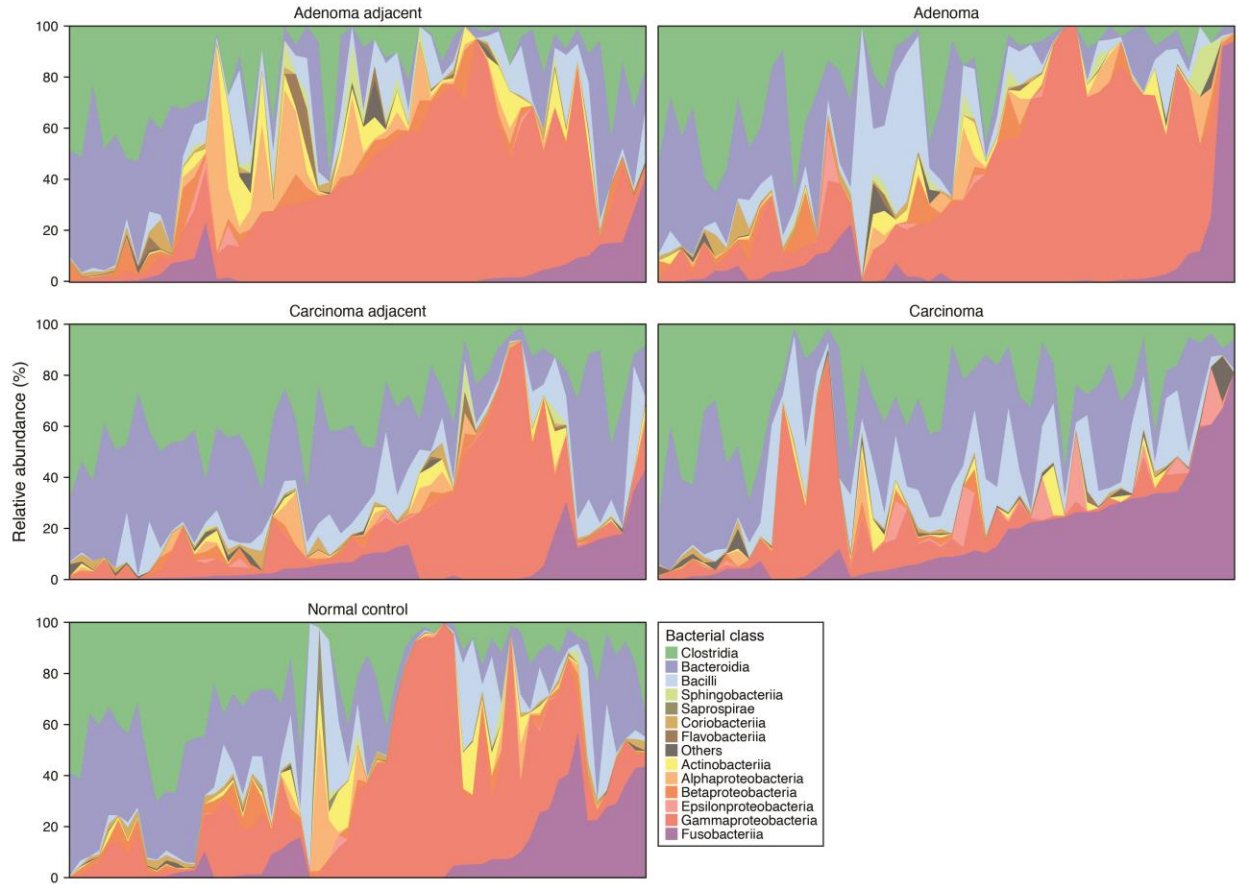




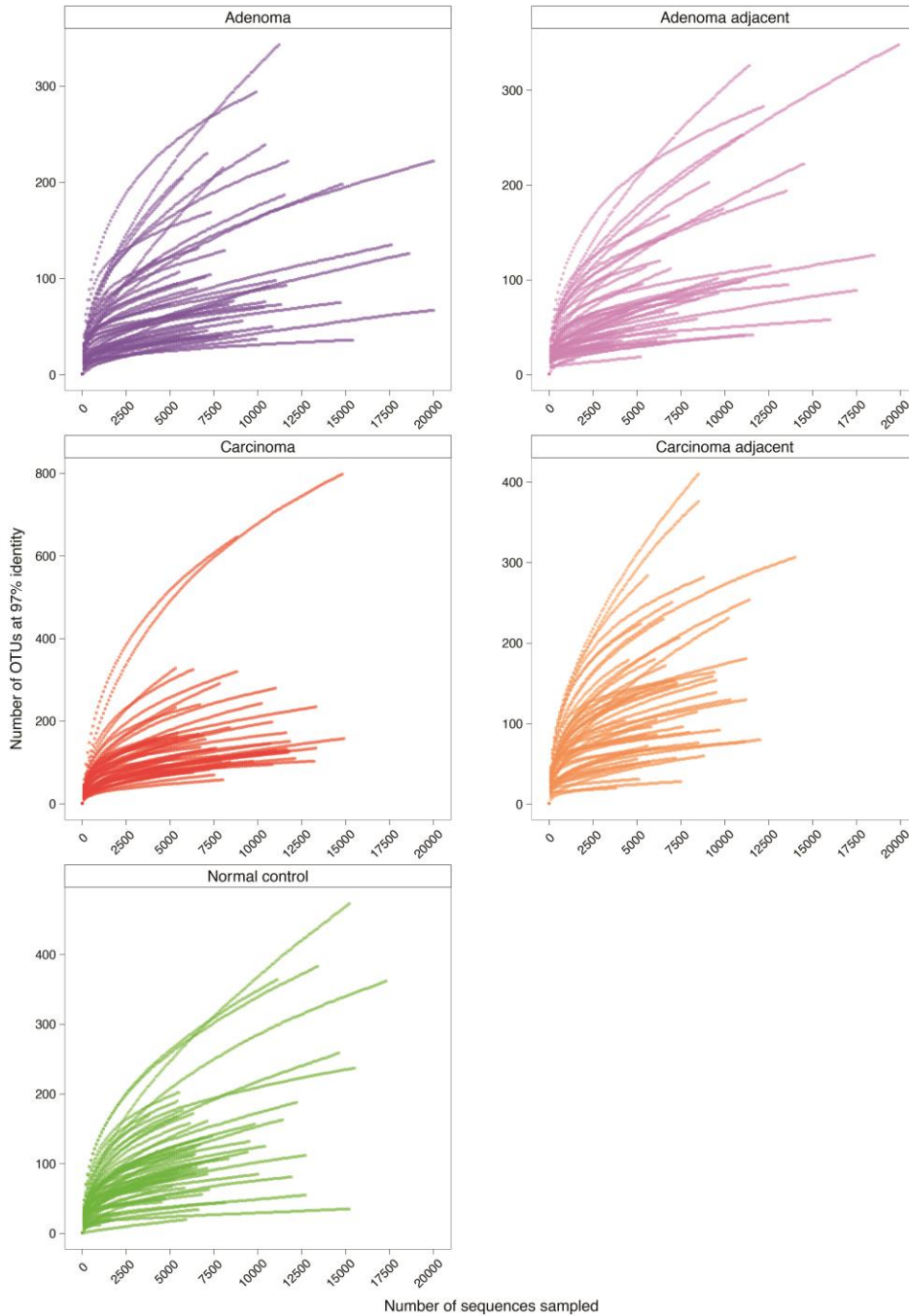
**Supplementary Figure 9 | Relative taxonomic abundance of the top 42 operational taxonomic units (OTUs) at 97% identity.** OTUs are ranked in the ascending order of the size of total relative abundance after rarefaction. Error bars represent standard errors of the means (SEMs). Mann-Whitney U test corrected by BH step-up procedure; \*  $q < 0.05$ ; \*\*  $q < 0.01$ ; \*\*\*  $q < 0.001$ ; \*\*\*\*  $q < 0.0001$ .



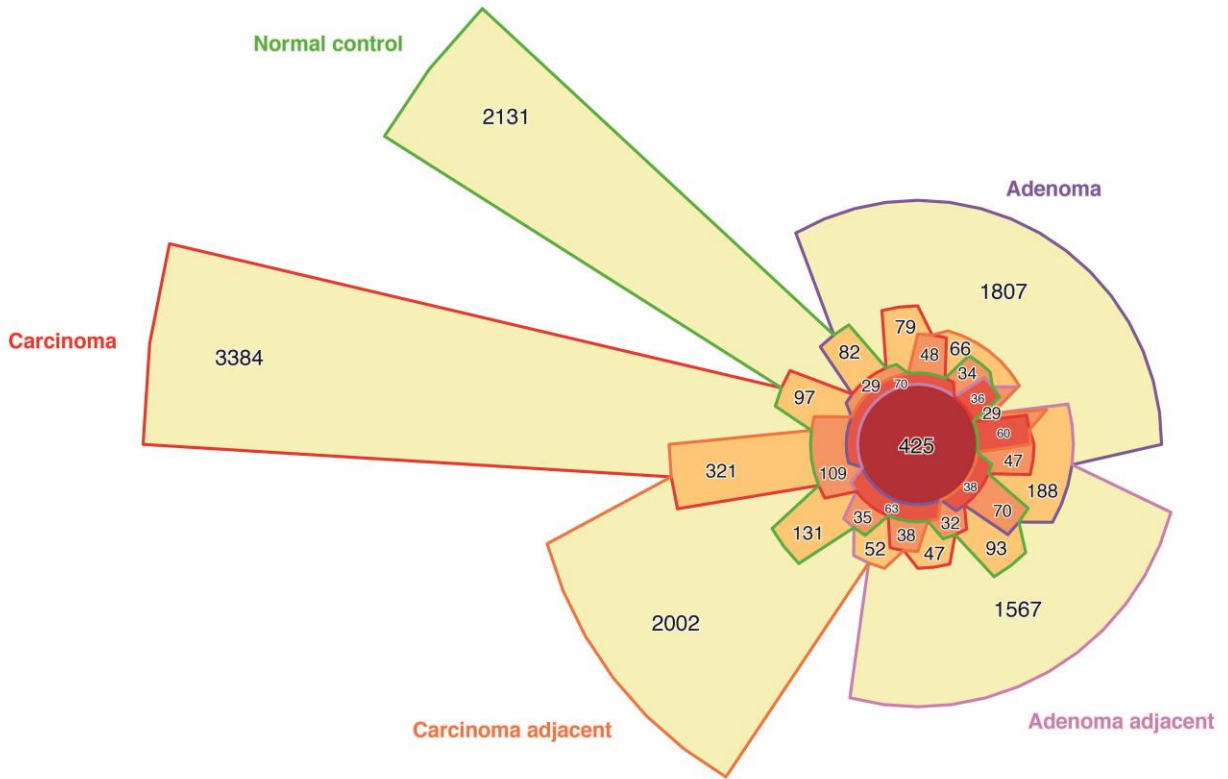
**Supplementary Figure 10 | Associations of metacommunities with clinical metadata.** The heights of boxes represent the IQR between the first and third quartiles in which medians are bolded; minimum and maximum values are denoted by whiskers; closed-circles are outliers. Mann-Whitney U tests corrected by BH step-up procedure; \*  $q < 0.05$ ; \*\*  $q < 0.01$ ; \*\*\*  $q < 0.001$ .



**Supplementary Figure 11 | Breakdown of sequencing reads at bacterial class-level.** Areas represent microbiome profiles of mucosal biopsies from normal ( $n = 61$ ), adenoma-affected ( $n = 47$ ), carcinoma-affected ( $n = 52$ ) colon.



**Supplementary Figure 12 | Comparisons of OTU sampling depths among biopsy phenotypes.** Rarefaction curves describe the number of 97% OTUs detected as sequencing effort increases.



**Supplementary Figure 13 | Five-way Venn diagram showing the distribution of 97% OTUs shared among colorectal mucosal phenotypes.** Number of shared sequence clusters are provided in the intersected areas.

<b>PATIENT CHARACTERISTICS</b>	<b>NORMAL COLON</b>	<b>COLORECTAL ADENOMA</b>	<b>COLORECTAL CARCINOMA</b>	<b>Adjusted <i>p</i>-value (Bonferroni)</b>
<b>Age, years (mean ± s.d.)</b>	60.13 ± 5.99	67.32 ± 8.80	67.85 ± 13.18	< 0.001
<b>Gender (<i>n</i>, percent)</b>				1
Male	25 (40.98)	21 (44.68)	31 (59.62)	
Female	36 (59.02)	26 (55.32)	21 (40.38)	
<b>BMI, kg/m<sup>2</sup> (mean ± s.d.)</b>	22.99 ± 2.75	24.02 ± 3.95	22.60 ± 2.79	1
<b>Chronic alcohol use (<i>n</i>, percent)</b>	3 (4.92)	5 (10.64)	10 (19.23)	1
<b>Smoking history (<i>n</i>, percent)</b>				1
Chronic smoker	2 (3.28)	4 (8.51)	12 (23.08)	
Current non-smoker	0 (0.00)	8 (17.02)	14 (26.92)	
<b>First degree family history of CRC (<i>n</i>, percent)</b>	2 (3.28)	6 (12.77)	1 (1.92)	1
<b>Diabetes mellitus (<i>n</i>, percent)</b>	2 (3.28)	2 (4.26)	4 (7.69)	1
<b>Anatomic origin (<i>n</i>, percent)</b>				
Proximal	27 (44.26)	20 (42.55)	13 (25.00)	1
Cecum	2 (3.28)	3 (6.38)	2 (3.85)	1
Ascending colon	25 (40.98)	11 (23.40)	10 (19.23)	1
Hepatic flexure	0 (0.00)	2 (4.26)	1 (1.92)	1
Transverse colon	0 (0.00)	6 (12.77)	1 (1.92)	0.121
Distal	34 (55.74)	27 (57.45)	39 (75.00)	1
Splenic flexure	0 (0.00)	1 (2.13)	0 (0.00)	1
Descending colon	10 (16.39)	7 (14.89)	3 (5.77)	1
Sigmoid colon	2 (3.28)	7 (14.89)	10 (19.23)	0.757
Rectosigmoid junction	0 (0.00)	2 (4.26)	6 (11.54)	0.279
Rectum	22 (36.07)	8 (17.02)	19 (36.54)	1
<b>TNM staging, AJCC 7th Edition (<i>n</i>, percent)</b>				

Zero	–	–	1 (1.92)	–
I	–	–	11 (21.15)	–
II	–	–	14 (26.92)	–
III	–	–	14 (26.92)	–
IV	–	–	12 (23.08)	–
<b>Histomorphologic type (n, percent)</b>				
Sessile-serrated	–	3 (6.38)	–	–
Tubular	–	28 (59.57)	–	–
Tubulovillous	–	15 (31.91)	–	–
Villous	–	1 (2.13)	–	–
<b>Grade of dysplasia (n, percent)</b>				
Low	–	35 (74.47)	–	–
High	–	12 (25.53)	–	–

**Supplementary Table 1 | Overview of 16S rRNA discovery cohort.** Multiple group comparisons for categorical and continuous variables were performed by Chi-squared/Fisher's exact tests and Kruskal-Wallis rank-sum tests, respectively.

PATIENT CHARACTERISTICS	NORMAL COLON	COLORECTAL ADENOMA	COLORECTAL CARCINOMA	Adjusted <i>p</i> -value (Bonferroni)
Age, years (mean ± s.d.)	41.28 ± 7.87	55.80 ± 11.36	61.34 ± 9.97	1
Gender ( <i>n</i> , percent)				0.086
Male	10 (40.00)	32 (78.05)	26 (52.00)	
Female	15 (60.00)	9 (21.95)	24 (48.00)	
Anatomic origin ( <i>n</i> , percent)				0.031
Proximal	0 (0.00)	11 (26.83)	24 (48.00)	
Distal	25 (100.00)	30 (73.17)	36 (72.00)	1
TNM staging, AJCC 7th Edition ( <i>n</i> , percent)				
Zero	–	–	0 (0.00)	–
I	–	–	6 (12.00)	–
II	–	–	17 (34.00)	–
III	–	–	25 (50.00)	–
IV	–	–	2 (4.00)	–
Histomorphologic type ( <i>n</i> , percent)				
Sessile-serrated	–	2 (4.88)	–	–
Tubular	–	27 (65.85)	–	–
Tubulovillous	–	11 (26.83)	–	–
Villous	–	0 (0.00)	–	–
Grade of dysplasia ( <i>n</i> , percent)				
Low	–	14 (28.00)	–	–
High	–	6 (12.00)	–	–
NA	–	21 (42.00)	–	–

**Supplementary Table 2 | Overview of real-time PCR validation cohort.** Multiple group comparisons for categorical and continuous variables were performed by Chi-squared/Fisher’s exact tests and Kruskal-Wallis rank-sum tests, respectively.