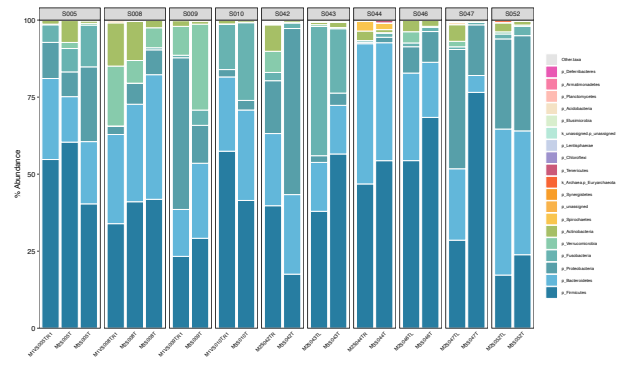
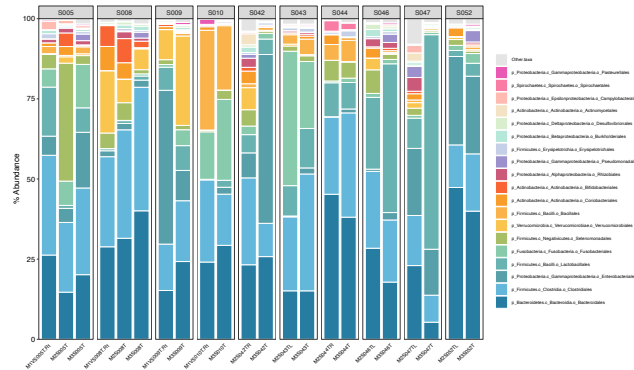


Figure S1

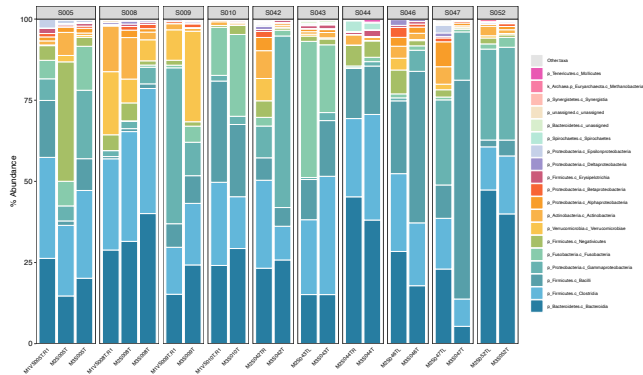
A Phylum Level Reps



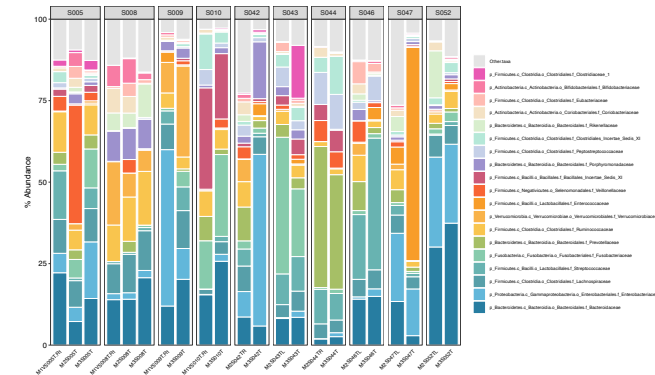
B Order Level Reps



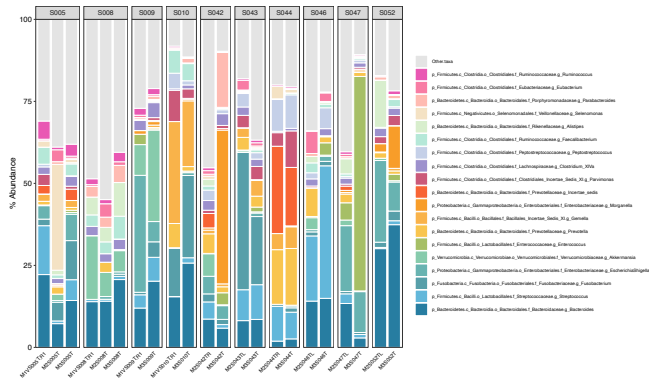
C Class Level Reps



D Family Level Reps

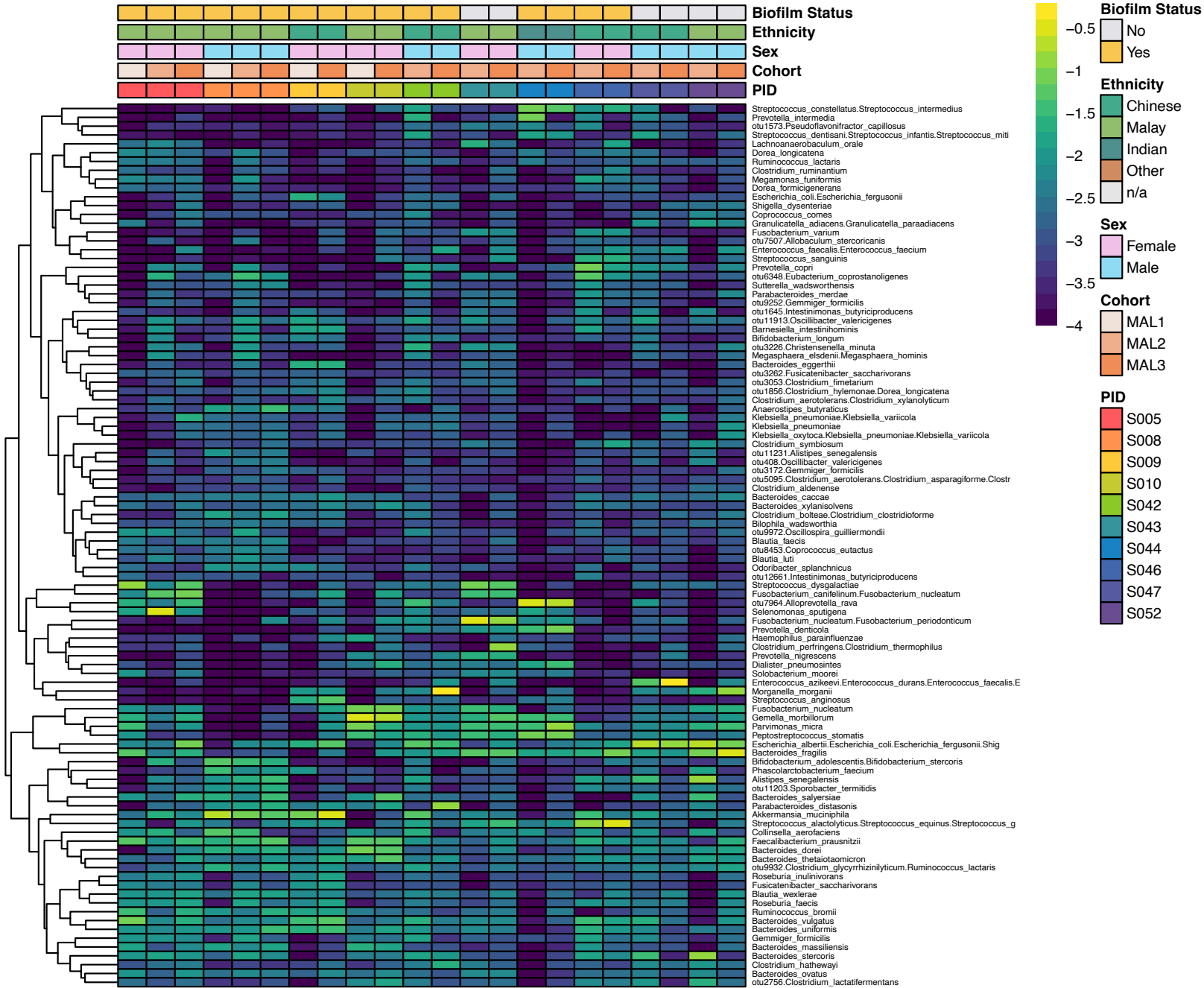


E Genus Level Reps



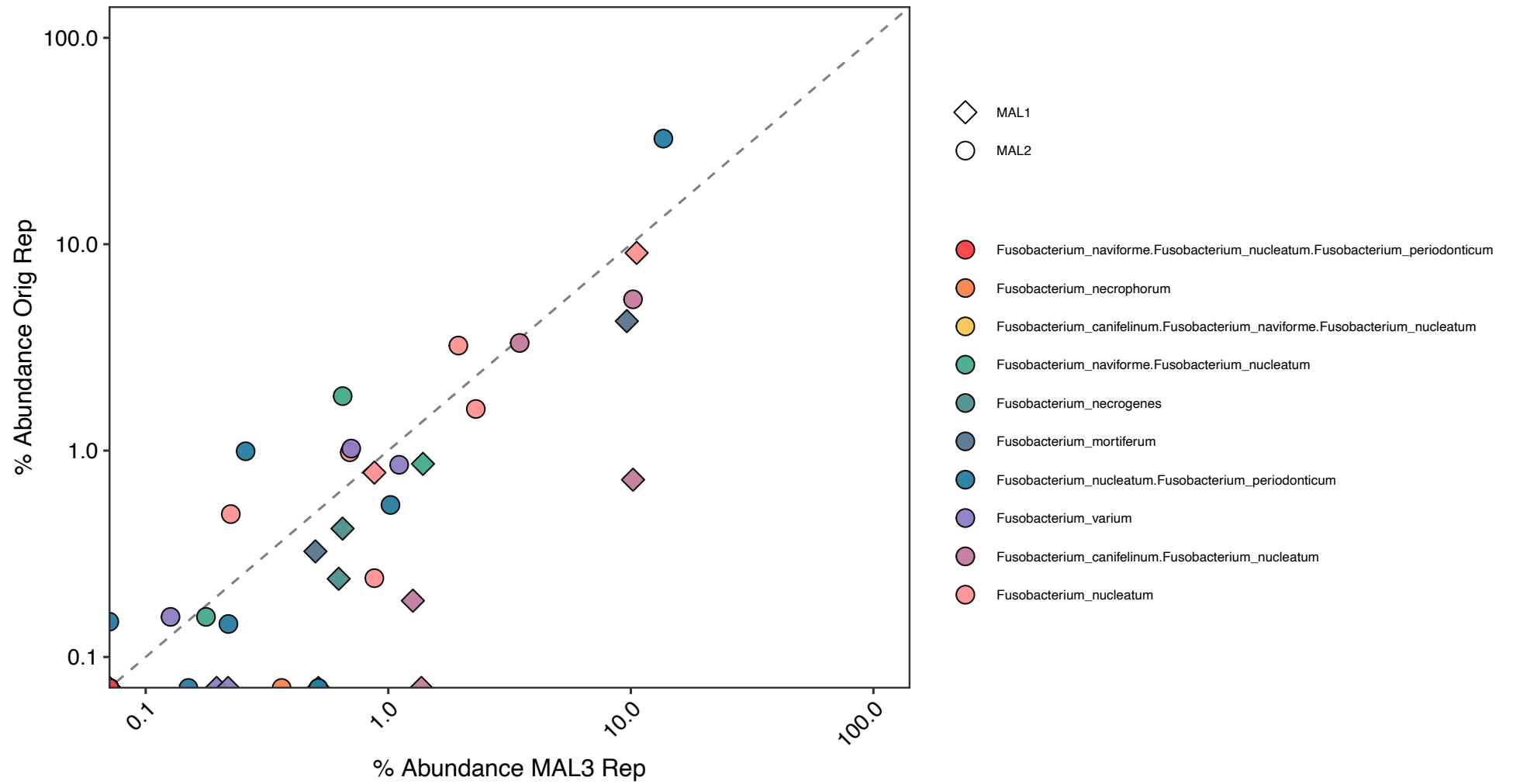
1 **Figure S1. Relative abundance of sequencing cohort replicates.** Relative abundance by 16S rRNA amplicon sequencing of 10 tumors
2 sequenced in replicate between cohorts MAL1, MAL2, and/or MAL3 at the A) phylum, B) order, C) Class, D) family, and E) genus
3 level.

Figure S2



4 **Figure S2. Species abundance of sequencing cohort replicates.** Heatmap depicting relative abundance by 16S rRNA amplicon
5 sequencing of 10 tumors sequenced in replicate between cohorts MAL1, MAL2, and/or MAL3 at the species level, with metadata color-
6 coded as depicted in the legend for biofilm status, ethnicity, sex, sequencing cohort, and patient identification number (PID).

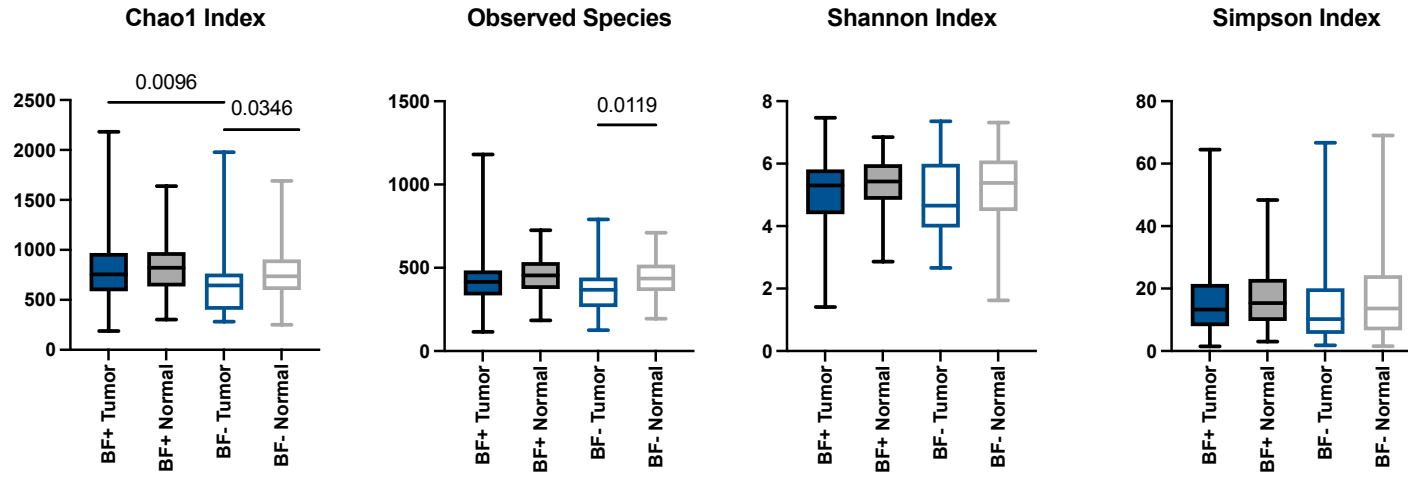
Figure S3



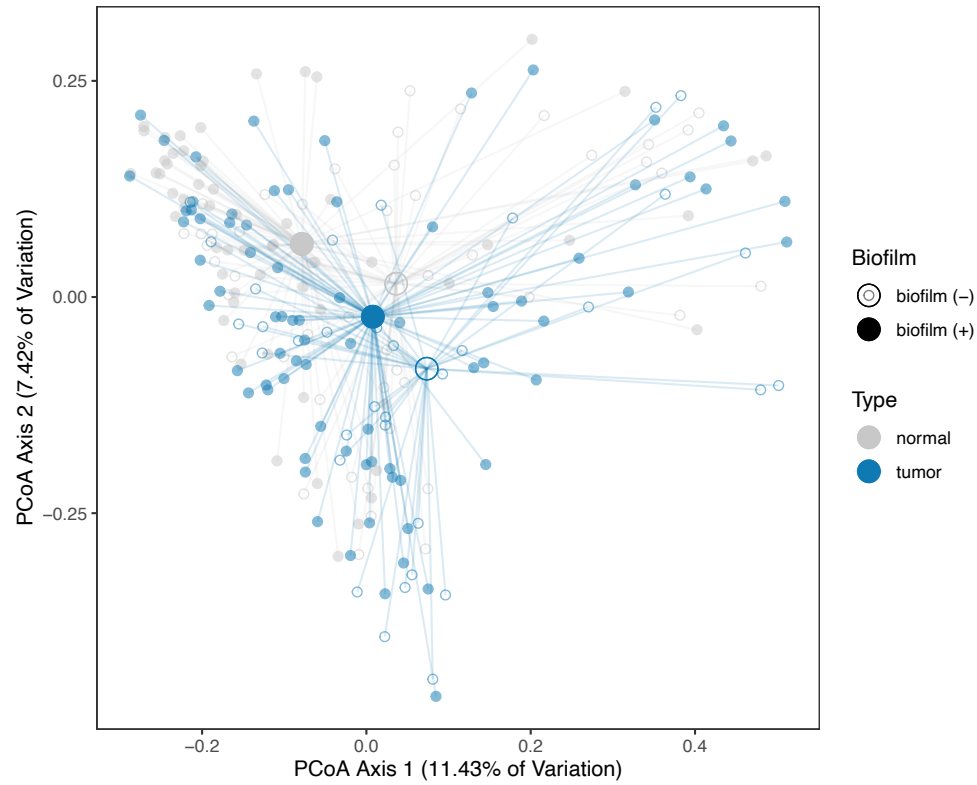
7 **Figure S3. Relative *Fusobacterium* species abundance of sequencing cohort replicates.** Relative abundance of *Fusobacterium*
8 species by 16S rRNA amplicon sequencing of 10 tumors sequenced in replicate between cohorts MAL1, MAL2, and/or MAL3 , depicted
9 as percent abundance of the original replicate in MAL1 or MAL2 graphed against percent abundance of the MAL3 replicate. Ambiguous
10 assignments depicted in the legend indicate inability to resolve two or more species (e.g.
11 “*Fusobacterium_nucleatum.Fusobacterium_peridonticum*” indicates that sequences could not distinguish between *F. nucleatum* and *F.*
12 *peridonticum*).

Figure S4

A

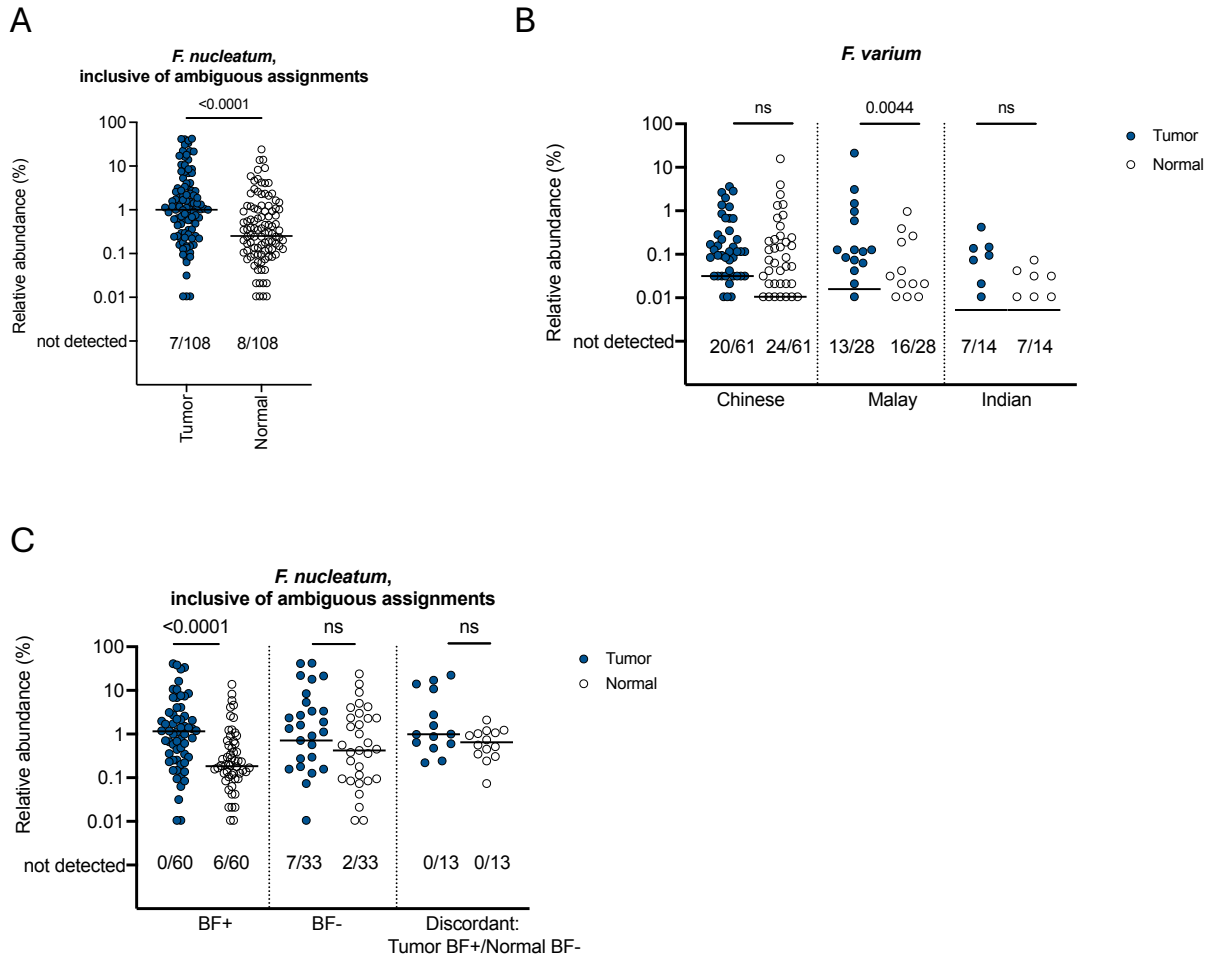


B



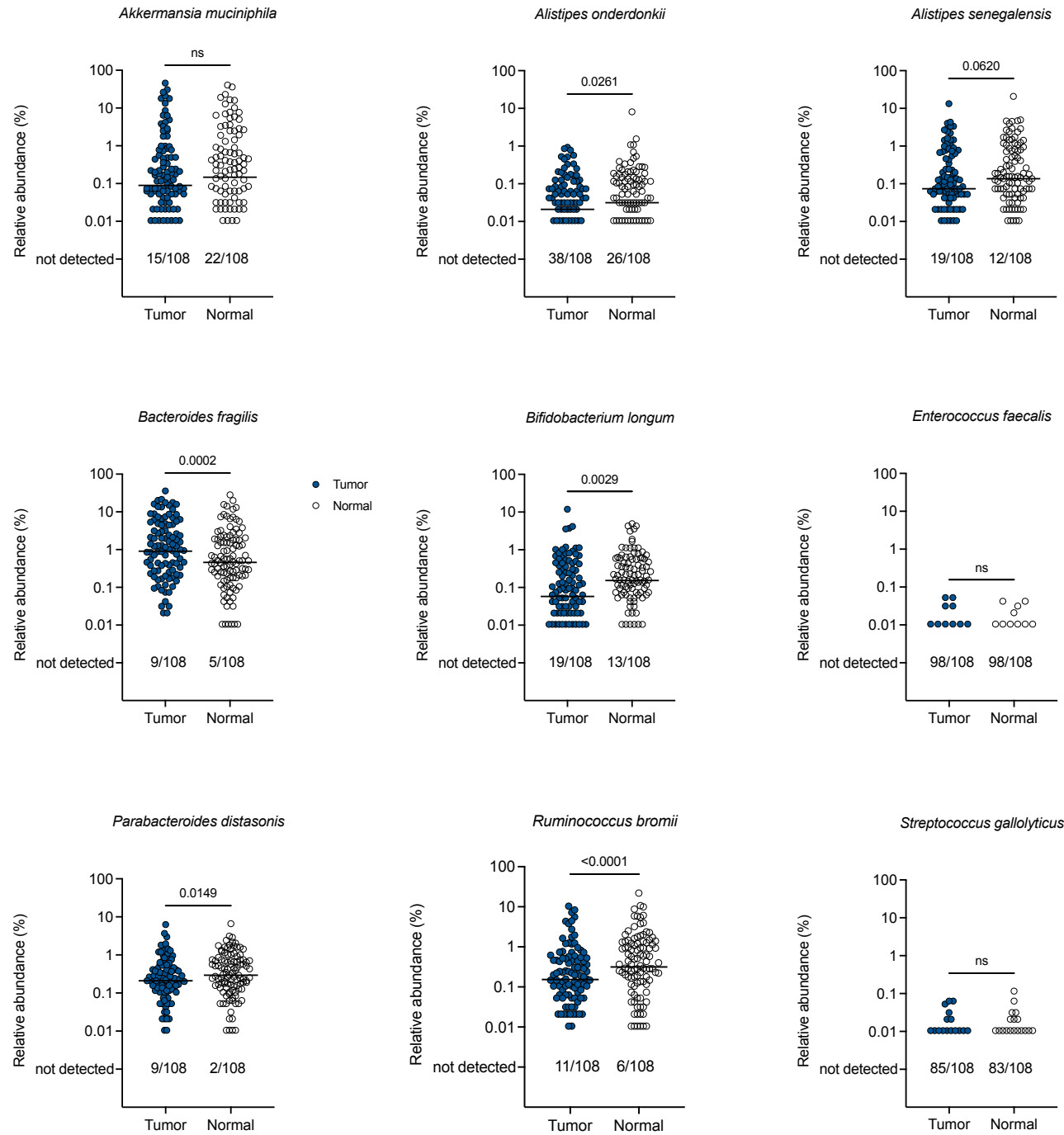
13 **Figure S4. Alpha and beta diversity of tumor and normal tissue by biofilm status.** A) Measures of alpha diversity for tumor and
14 normal samples with or without biofilms. BF+ indicates biofilm-positive and BF- indicates biofilm-negative. Box and whisker plots
15 depict min to max. Analyzed by two-tailed Mann Whitney test, with $p < 0.05$ considered significant. B) Principal coordinate analysis
16 plots (PCoA) of Bray Curtis dissimilarity depicted as centroids. Analyzed by PERMANOVA, with $p < 0.05$ considered significant.

Figure S5



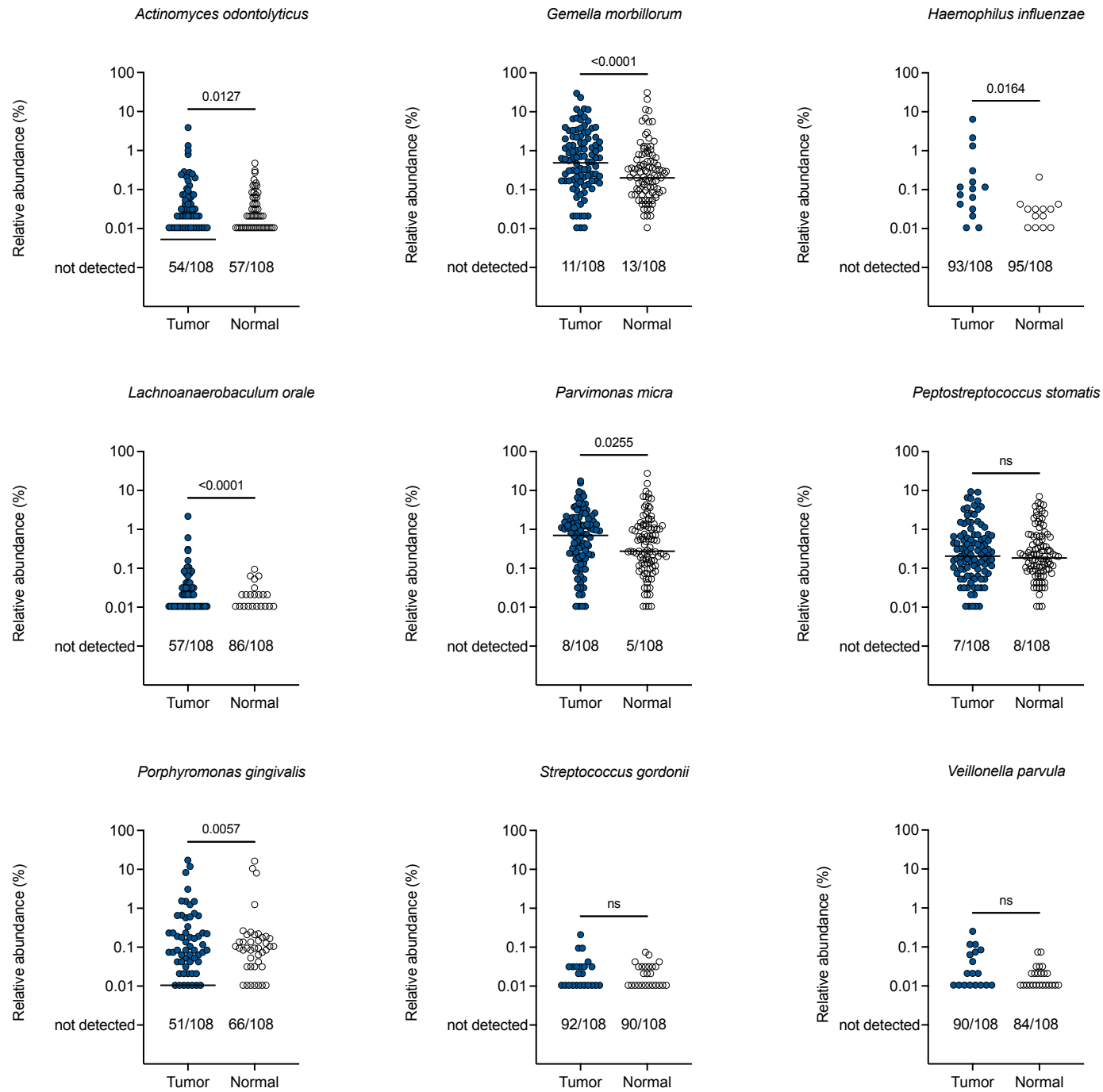
17 **Figure S5. Additional details on *Fusobacterium nucleatum* and *Fusobacterium varium* abundance.** A, C) Relative abundance by
18 16S rRNA amplicon sequencing of each depicted *Fusobacterium nucleatum* as a sum of confident and ambiguous species assignments
19 in pairs of tumor (blue circle) and normal (clear circle) tissues, stratified by biofilm (BF) status in panel C. B) Relative abundance of
20 *Fusobacterium varium* abundance stratified by ethnicity in pairs of tumor (blue circle) and normal (clear circle) tissues. Bars indicate
21 median. Numbers below each graph depict the number of samples out of the total in which each species was not detected by sequencing.
22 Analyzed by two-tailed Wilcoxon matched-pairs signed rank test, with $p < 0.05$ considered significant.

Figure S6



23 **Figure S6. Abundance of other CRC-associated species.** Relative abundance by 16S rRNA amplicon sequencing of each depicted
24 species in pairs of tumor (blue circle) and normal (clear circle) tissues. Bars indicate median. Numbers below each graph depict the
25 number of samples out of the total in which each species was not detected by sequencing. Analyzed by two-tailed Wilcoxon matched-
26 pairs signed rank test, with $p < 0.05$ considered significant.

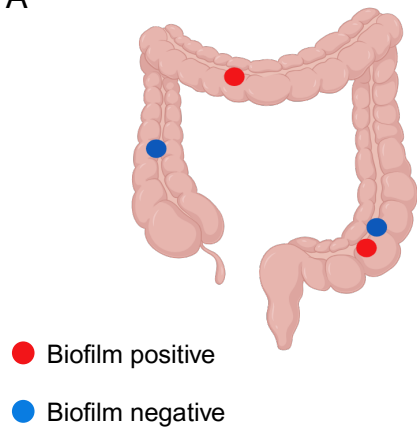
Figure S7



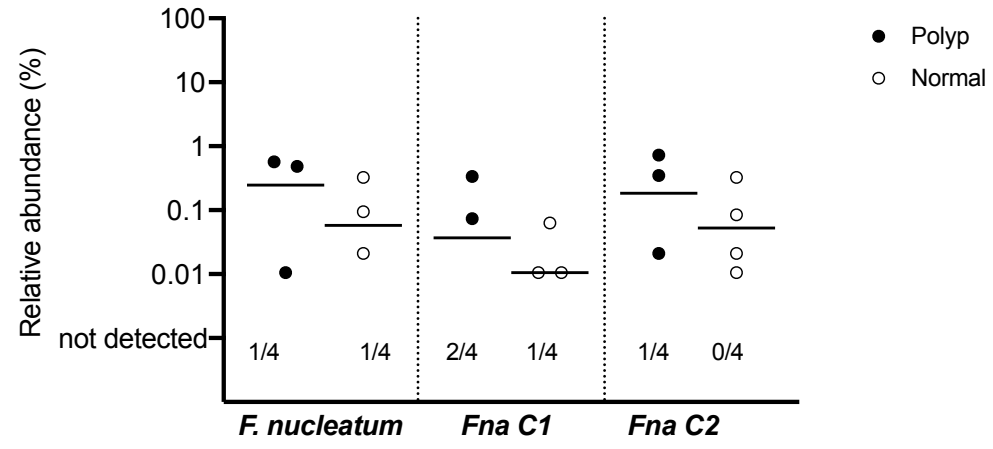
27 **Figure S7. Abundance of oral biofilm organisms in CRC tumors/normal pairs.** Relative abundance by 16S rRNA amplicon
28 sequencing of each depicted species in pairs of tumor (blue circle) and normal (clear circle) tissues. Bars indicate median. Numbers
29 below each graph depict the number of samples out of the total in which each species was not detected by sequencing. Analyzed by two-
30 tailed Wilcoxon matched-pairs signed rank test, with $p < 0.05$ considered significant.

Figure S8

A



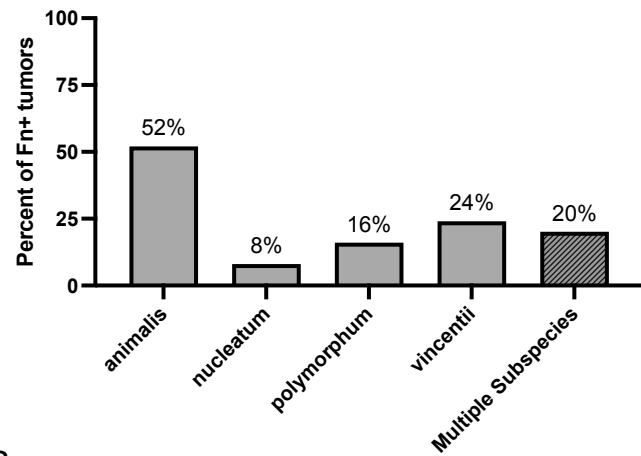
B



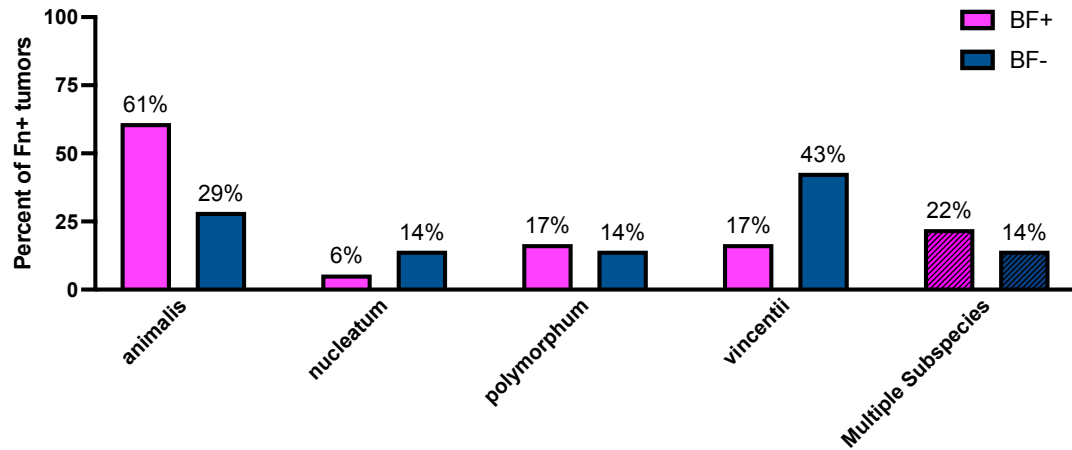
31 **Figure S8. Adenomas in the cohort.** A) Diagram of adenoma (polyp) location in the colon and biofilm status (red circle = biofilm-
32 positive; blue circle = biofilm-negative). B) Relative abundance by 16S rRNA amplicon sequencing of *Fusobacterium nucleatum*,
33 *subspecies animalis* clade 1 (Fna C1) and clade 2 (Fna C2) in pairs of polyp (black circle) and normal (clear circle) tissues. Bars indicate
34 median. Numbers below each graph depict the number of samples out of the total in which each species was not detected by sequencing.

Figure S9

A

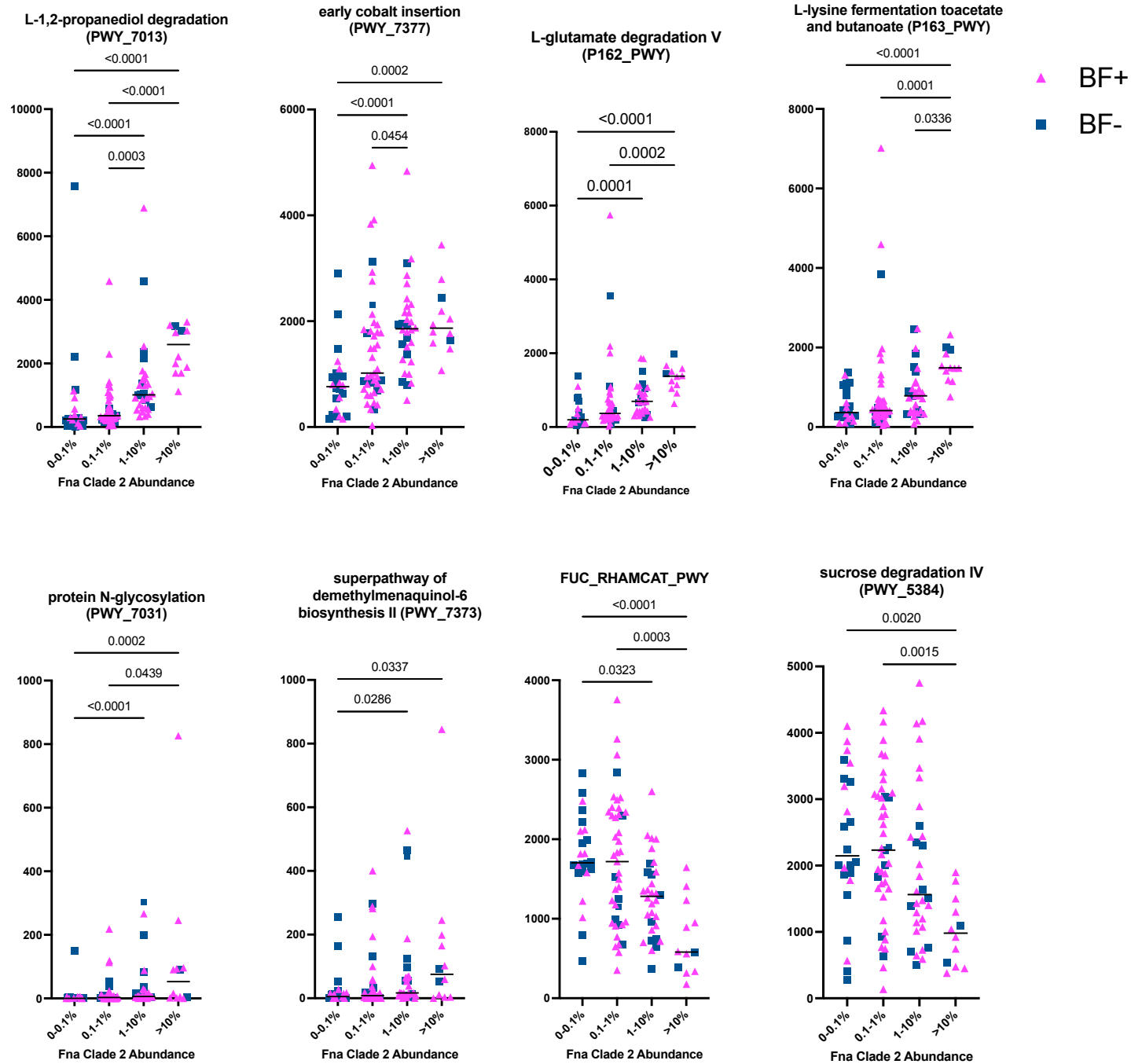


B



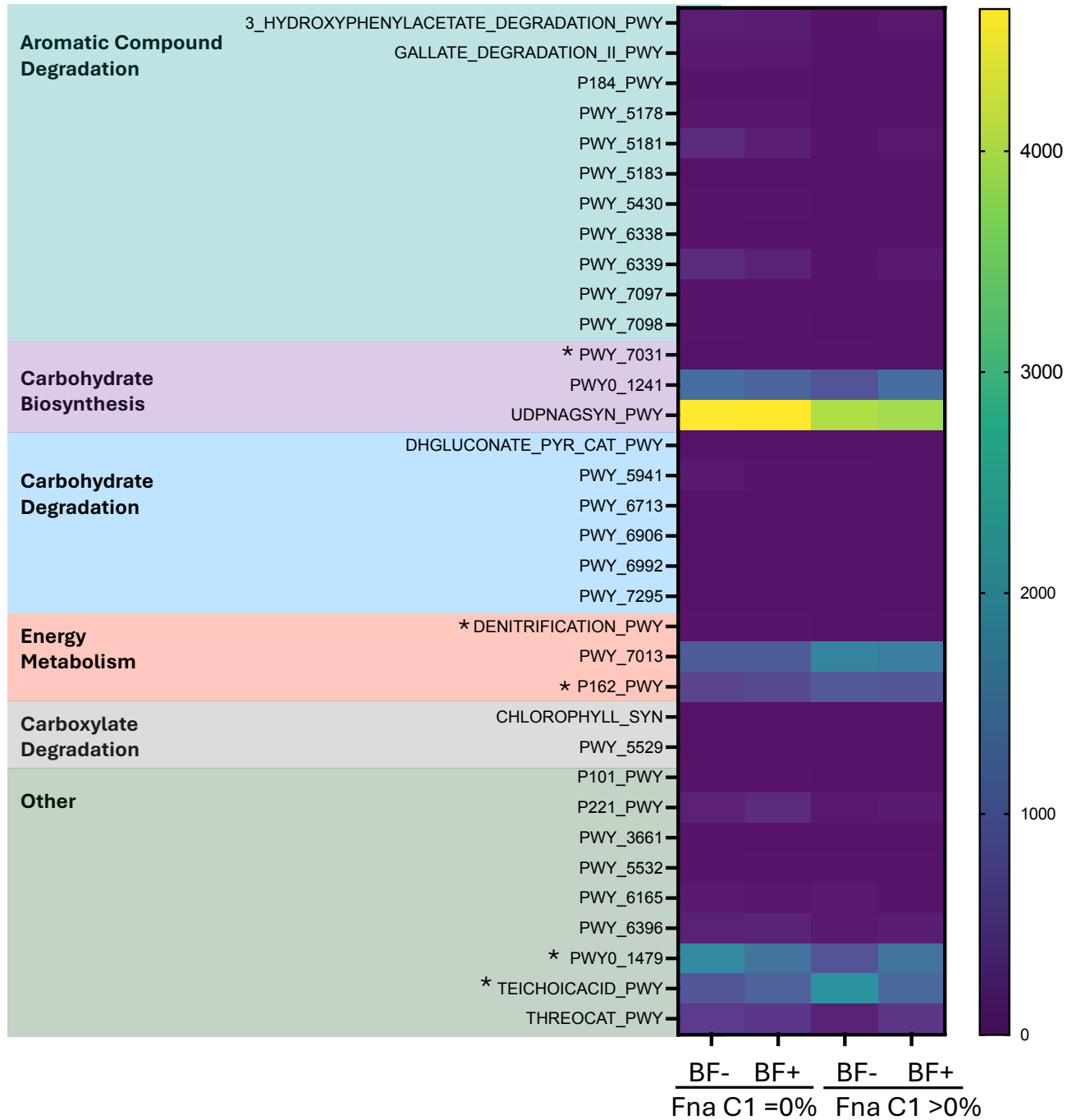
35 **Figure S9. *Fusobacterium nucleatum* subspecies detected in tumors by PCR.** Tumor DNA samples positive for *F. nucleatum* (Fn+)
36 using primers for *nusG* (n=25) were screened with subspecies-specific PCRs ¹⁶. A) Percent of these tumors in which bands were detected
37 for each subspecies or for multiple subspecies, and B) stratified by the biofilm (BF) status of the tumor.

Figure S10



38 **Figure S10. PICRUSt2 predictions of metabolic function associated with Fna C2 abundance in CRC tumors.** Specific PICRUSt2
39 pathways of interest from among those depicted in Fig 6, for biofilm positive (pink triangle) and biofilm negative (blue square) tumors
40 for each Fna C2 abundance category. Bars indicate median. Analyzed by Kruskal-Wallis test and Dunn's multiple comparisons, with
41 $p < 0.05$ considered significant.

Figure S11



42 **Figure S11. PICRUSt2 pathways associated with Fna C1 abundance in CRC tumors.** Heatmap of all statistically significant
43 PICRUSt2 pathways with significant differences in gene content between absent Fna C1 (0% abundance) and present Fna C1 (>0%) in
44 tumors (n=112), as analyzed by Mann Whitney test, with $p < 0.05$ considered significant. Pathways denoted with an asterix (*) were also
45 significantly different between Fna C2 abundance categories, as depicted in Fig 6. Data are further stratified by biofilm (BF) status
46 within each abundance category.

47 **Supplementary Tables**

- 48 • Extended Data 1: 16S Sequencing results of 10 replicates
- 49 ○ Table S1A: pre-processing data
- 50 ○ Table S1B: phylum
- 51 ○ Table S1C: class
- 52 ○ Table S1D: order
- 53 ○ Table S1E: family
- 54 ○ Table S1F: genus
- 55 ○ Table S1G: species/OTU
- 56
- 57 • Extended Data 2: 16S Sequencing results of full cohort
- 58 ○ Table S2A: 16S alpha diversity
- 59 ○ Table S2B: 16S phylum
- 60 ○ Table S2C: 16S class
- 61 ○ Table S2D: 16S order
- 62 ○ Table S2E: 16S family
- 63 ○ Table S2F: 16S genus
- 64 ○ Table S2G: 16S species/OTU
- 65 ○ Table S2H: Fna clades
- 66 ○ Table S2I: PICRUSt2 pathways
- 67
- 68 • Table S3: *Fusobacterium* strains isolated by culture