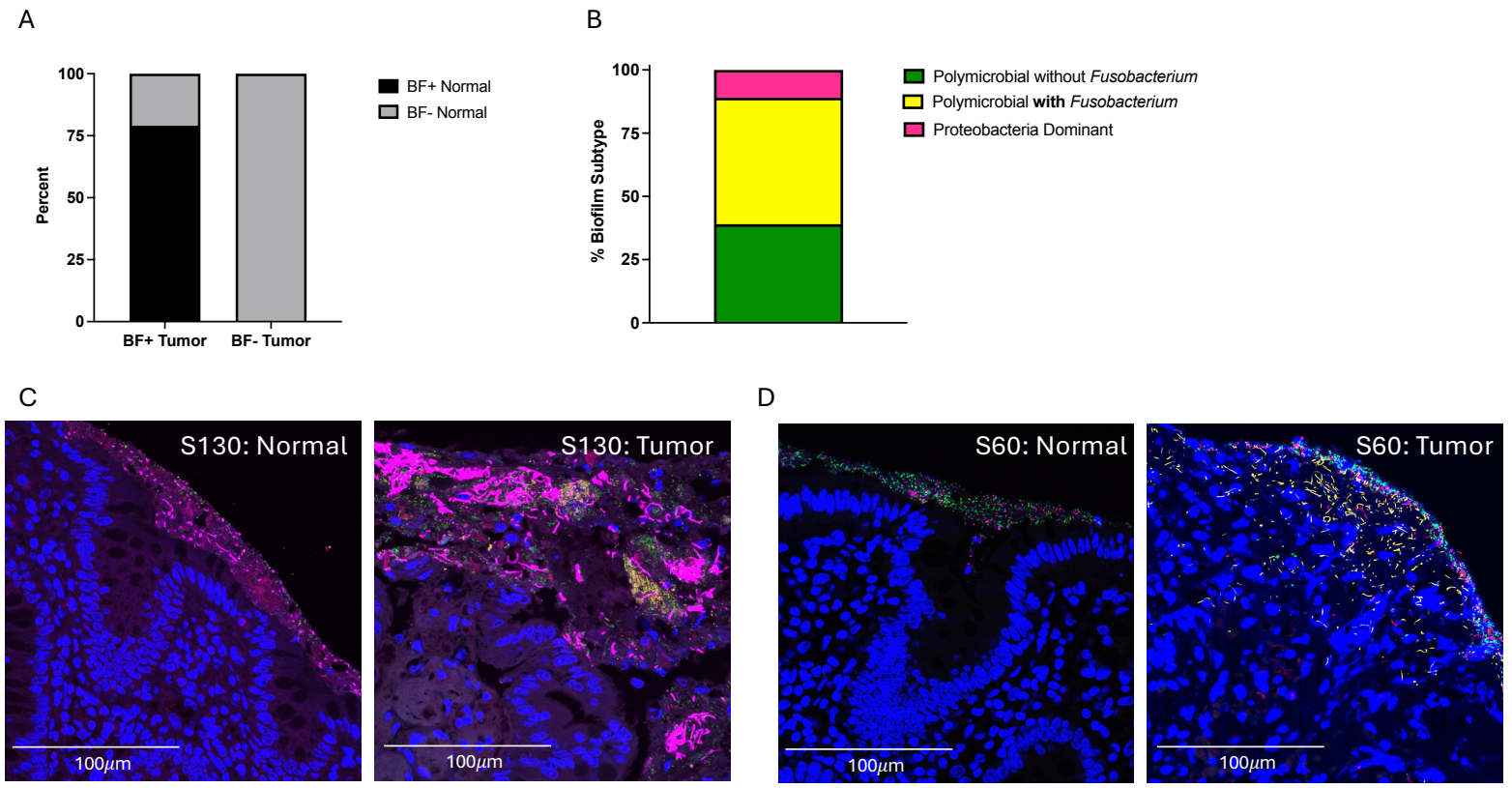

Figure S1. Biofilms on Normal Tissues. A) Percent of biofilm positive (black) or biofilm negative (gray) paired normal tissue for biofilm positive or biofilm negative tumors. N= 110 tumor/normal pairs. B) Stacked bar graph depicting percent of tumor biofilms of each subtype screened by multi-probe FISH (n=28). C-D) Representative 40X images of multi-probe FISH of tumor and normal paired tissues stained for DAPI (blue), Bacteroidetes (green), Fusobacterium (yellow), Lachnospiraceae (red), and Proteobacteria (pink).

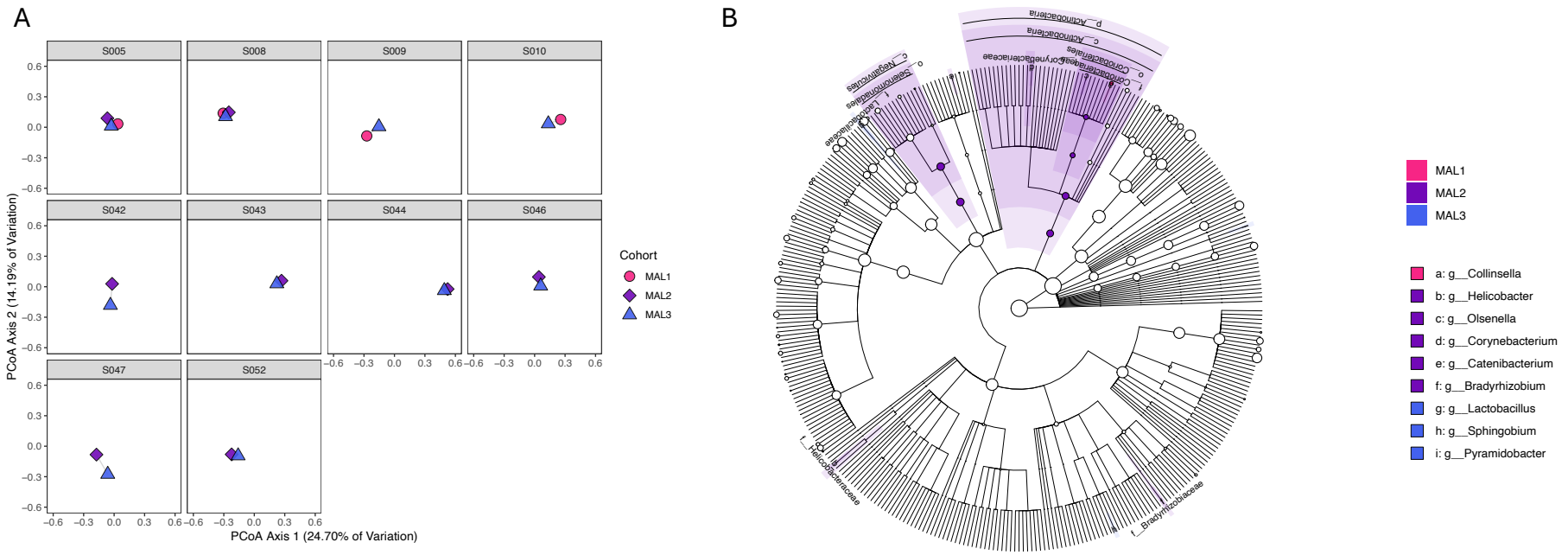

Figure S2. Variation sequencing cohort replicates. A) Principal coordinate analysis plots (PCoA) of Bray Curtis dissimilarity of 10 replicates from MAL1, MAL2, and/or MAL3. Analyzed by PERMANOVA, with $p<0.05$ considered significant. No comparisons between replicates were significant. B) Linear discriminant analysis Effect Size (LEfSe) cladogram of 10 replicates from MAL1, MAL2, and/or MAL3. Shaded regions and labeled taxa are significantly enriched ($p<0.01$) in the indicated cohort relative to the others.

Figure S3. Relative abundance of sequencing cohort replicates. Relative abundance by 16S rRNA amplicon sequencing of 10 tumors sequenced in replicate between cohorts MAL1, MAL2, and/or MAL3 at the A) phylum, B) order, C) Class, D) family, and E) genus level.

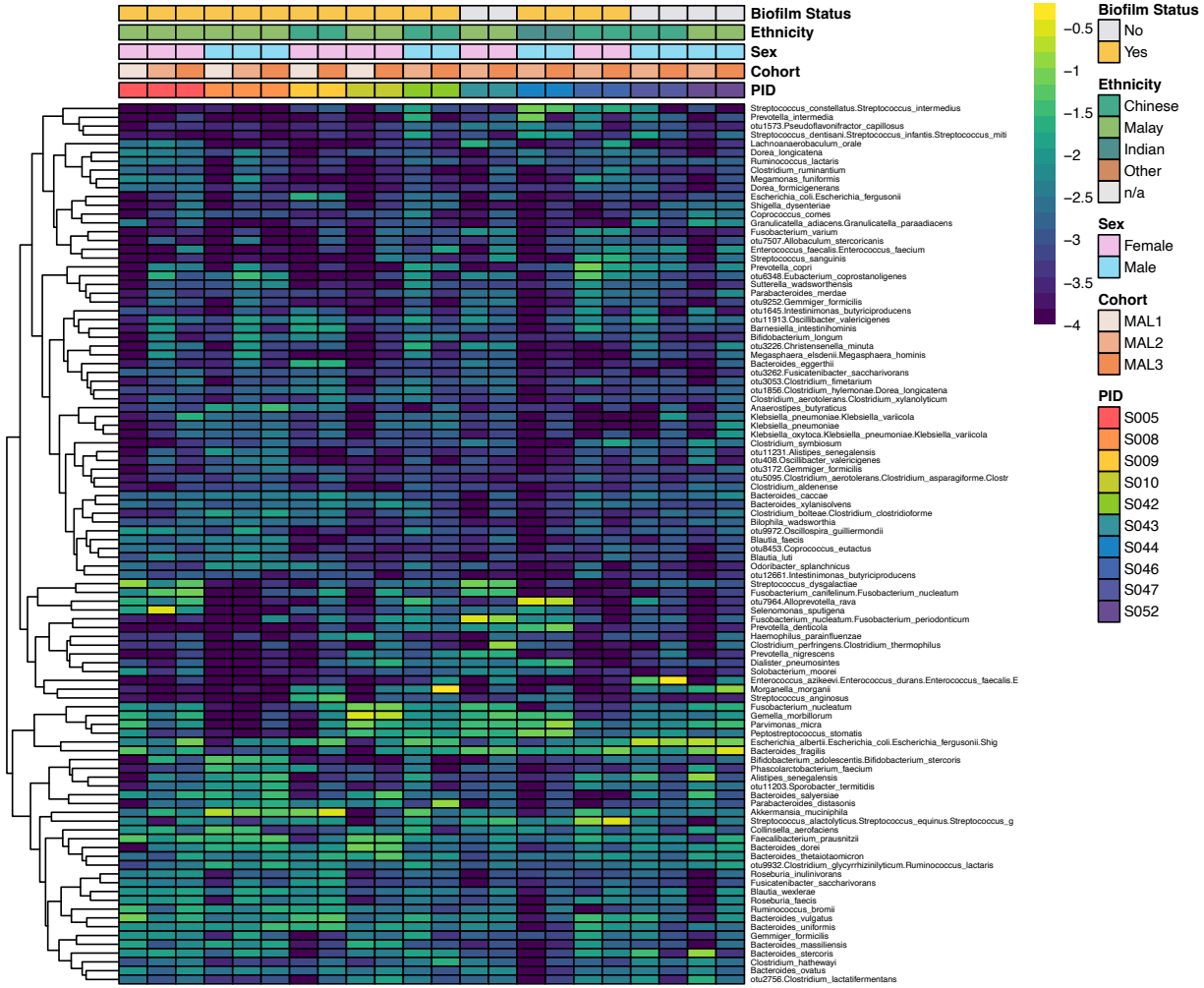

Figure S4. Species abundance of sequencing cohort replicates. Heatmap depicting relative abundance by 16S rRNA amplicon sequencing of 10 tumors sequenced in replicate between cohorts MAL1, MAL2, and/or MAL3 at the species level, with metadata color-coded as depicted in the legend for biofilm status, ethnicity, sex, sequencing cohort, and patient identification number (PID).

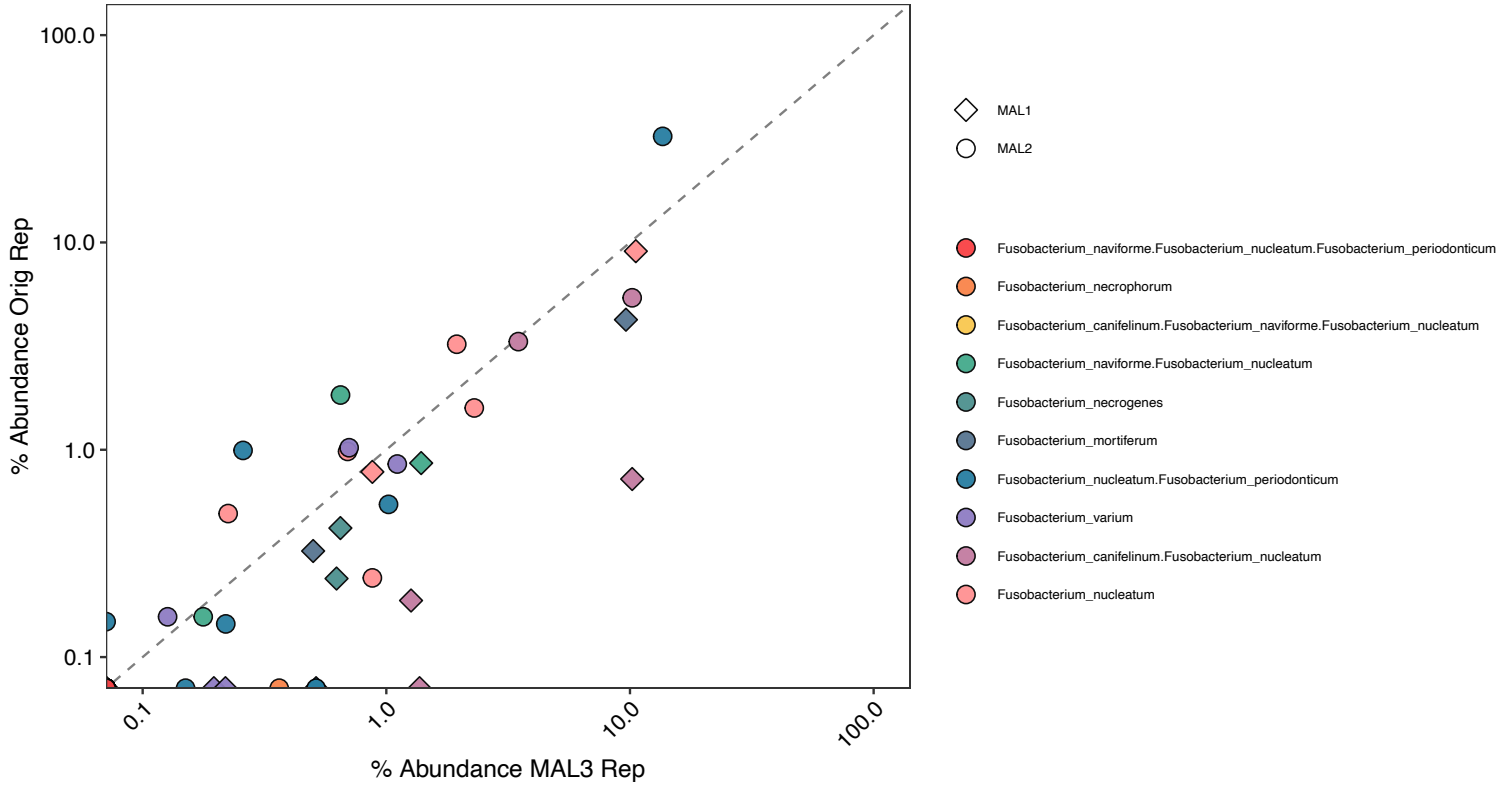

Figure S5. Relative *Fusobacterium species* abundance of sequencing cohort replicates. Relative abundance of *Fusobacterium* species by 16S rRNA amplicon sequencing of 10 tumors sequenced in replicate between cohorts MAL1, MAL2, and/or MAL3 , depicted as percent abundance of the original replicate in MAL1 or MAL2 graphed against percent abundance of the MAL3 replicate. Ambiguous assignments depicted in the legend indicate inability to resolve two or more species (e.g. "Fusobacterium_nucleatum.Fusobacterium_peridonticum" indicates that sequences could not distinguish between *F. nucleatum* and *F. peridonticum*).

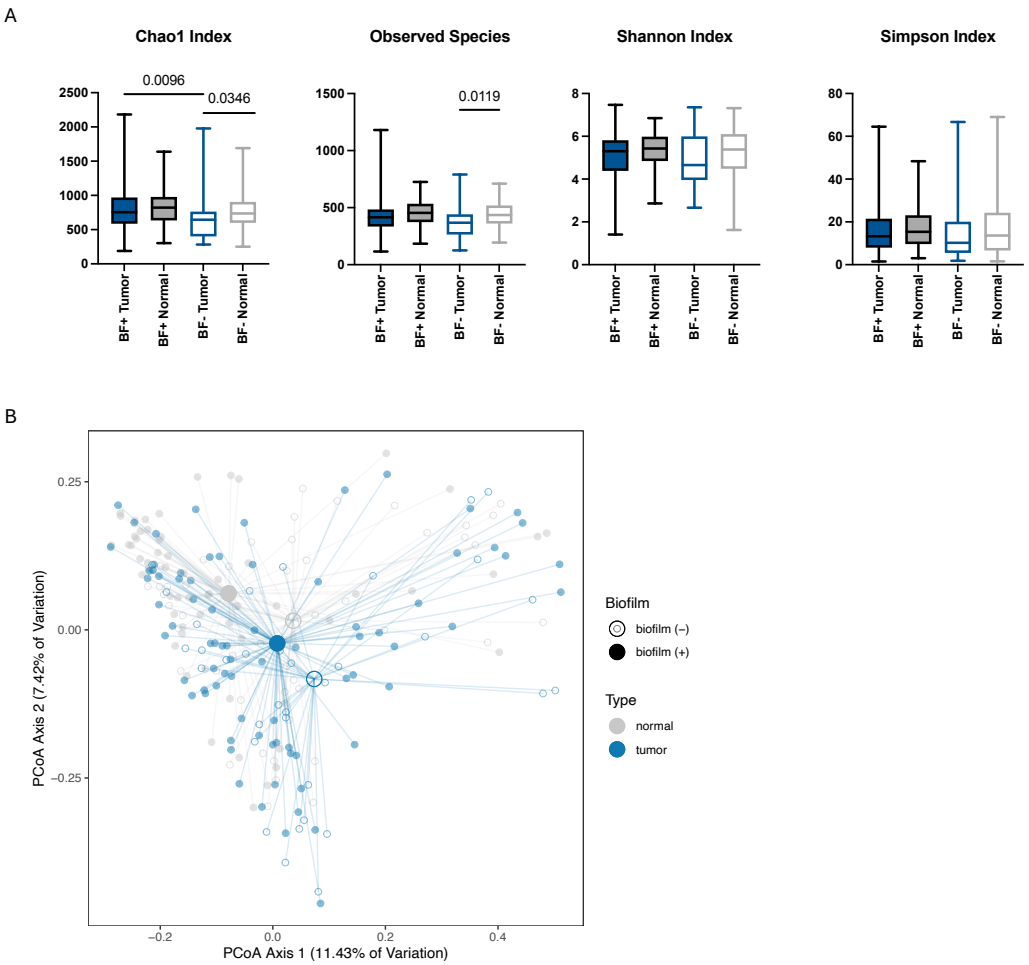

Figure S6. Alpha and beta diversity of tumor and normal tissue by biofilm status. A) Measures of alpha diversity for tumor and normal samples with or without biofilms. BF+ indicates biofilm-positive and BF- indicates biofilm-negative. Box and whisker plots depict min to max. Analyzed by two-tailed Mann Whitney test, with $p<0.05$ considered significant. B) Principal coordinate analysis plots (PCoA) of Bray Curtis dissimilarity depicted as centroids. Analyzed by PERMANOVA, with $p<0.05$ considered significant. No comparisons were significant. N= 49 BF- normal, 34 BF- tumor, 62 BF+ normal, and 77 BF+ tumor samples.

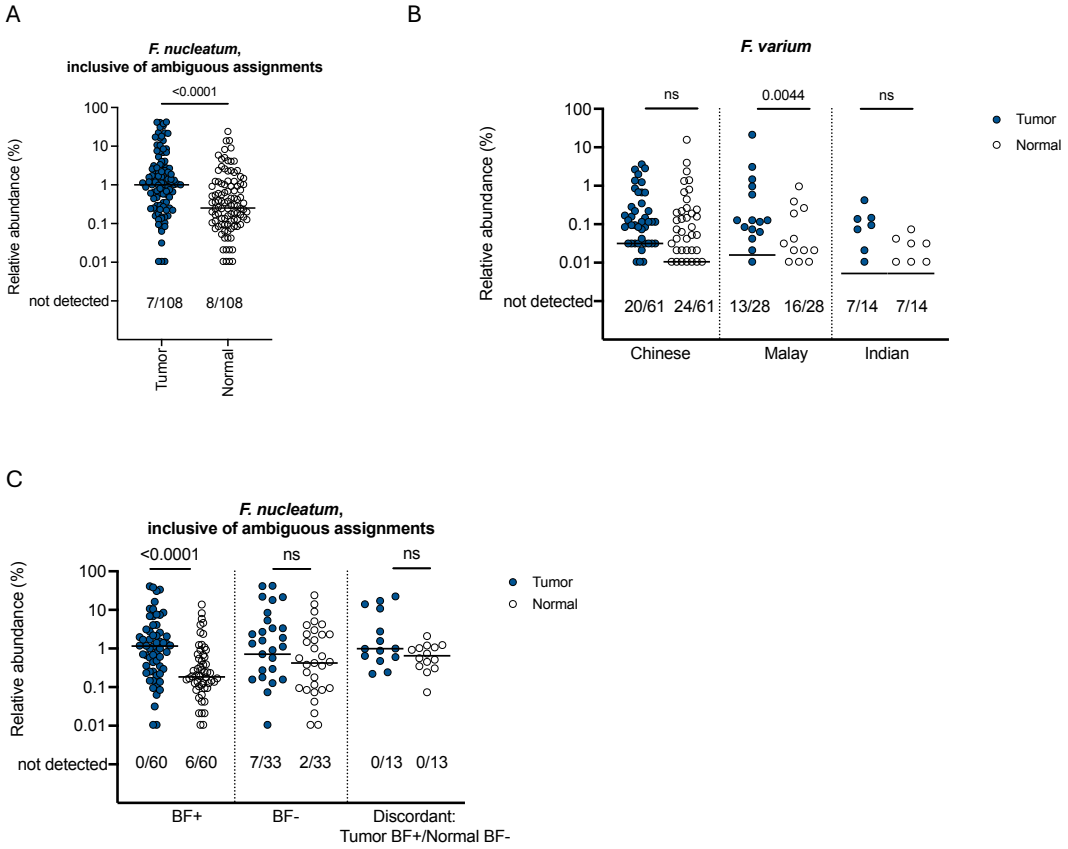

Figure S7. Additional details on *Fusobacterium nucleatum* and *Fusobacterium varium* abundance. A, C) Relative abundance by 16S rRNA amplicon sequencing of each depicted *Fusobacterium nucleatum* as a sum of confident and ambiguous species assignments in pairs of tumor (blue circle) and normal (clear circle) tissues, stratified by biofilm (BF) status in panel C. B) Relative abundance of *Fusobacterium varium* abundance stratified by ethnicity in pairs of tumor (blue circle) and normal (clear circle) tissues. Bars indicate median. Of 112 individuals with CRC in the study, 108 tumor and normal pairs were available for 16S rRNA sequencing analysis. Numbers below each graph depict the number of samples out of the total in which each species was not detected by sequencing. Analyzed by two-tailed Wilcoxon matched-pairs signed rank test, with $p<0.05$ considered significant.

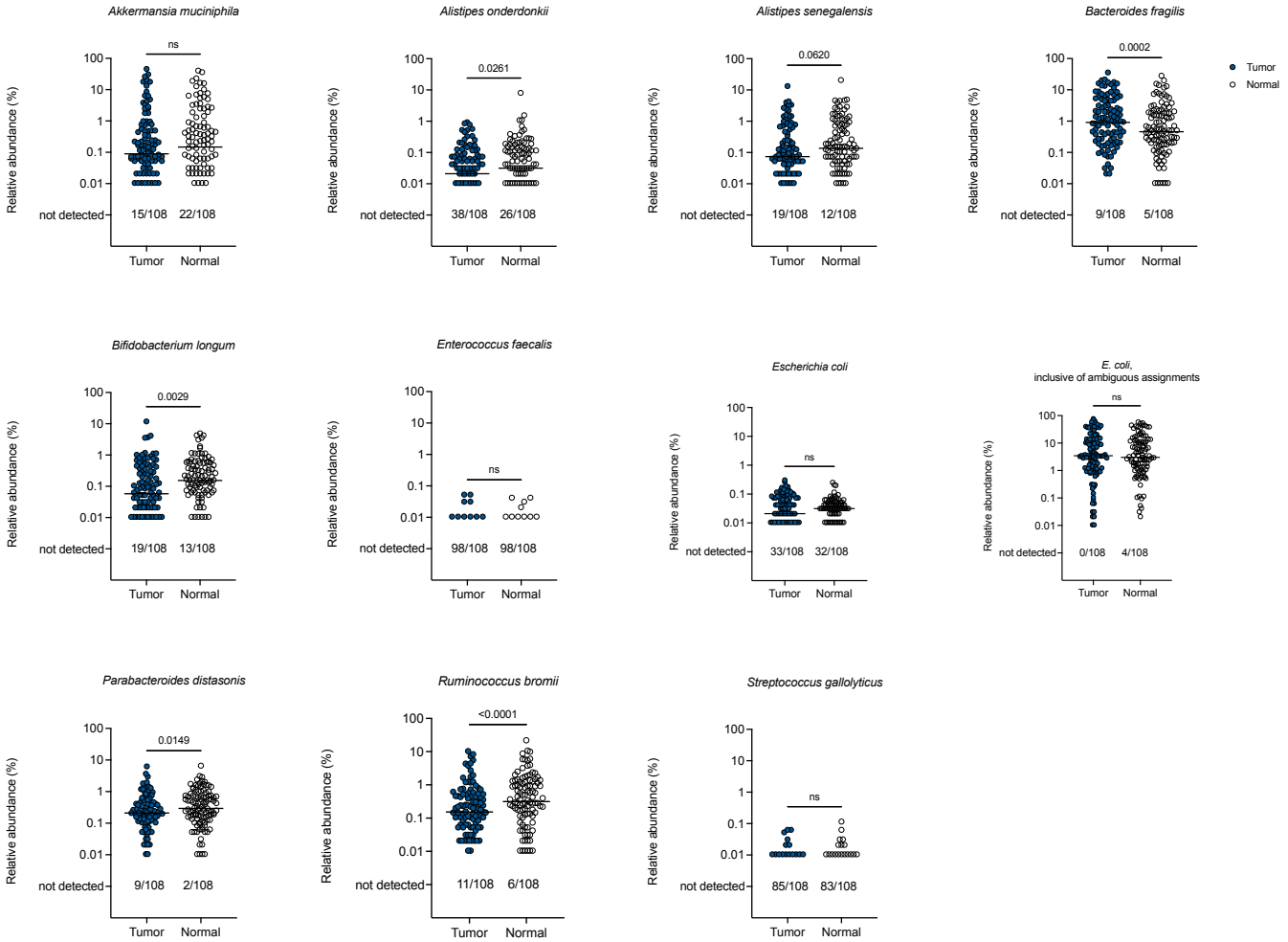

Figure S8. **Abundance of other CRC-associated species.** Relative abundance by 16S rRNA amplicon sequencing of each depicted species in pairs of tumor (blue circle) and normal (clear circle) tissues. Bars indicate median. Of 112 individuals with CRC in the study, 108 tumor and normal pairs were available for 16S rRNA sequencing analysis. Numbers below each graph depict the number of samples out of the total in which each species was not detected by sequencing. Analyzed by two-tailed Wilcoxon matched-pairs signed rank test, with $p<0.05$ considered significant.

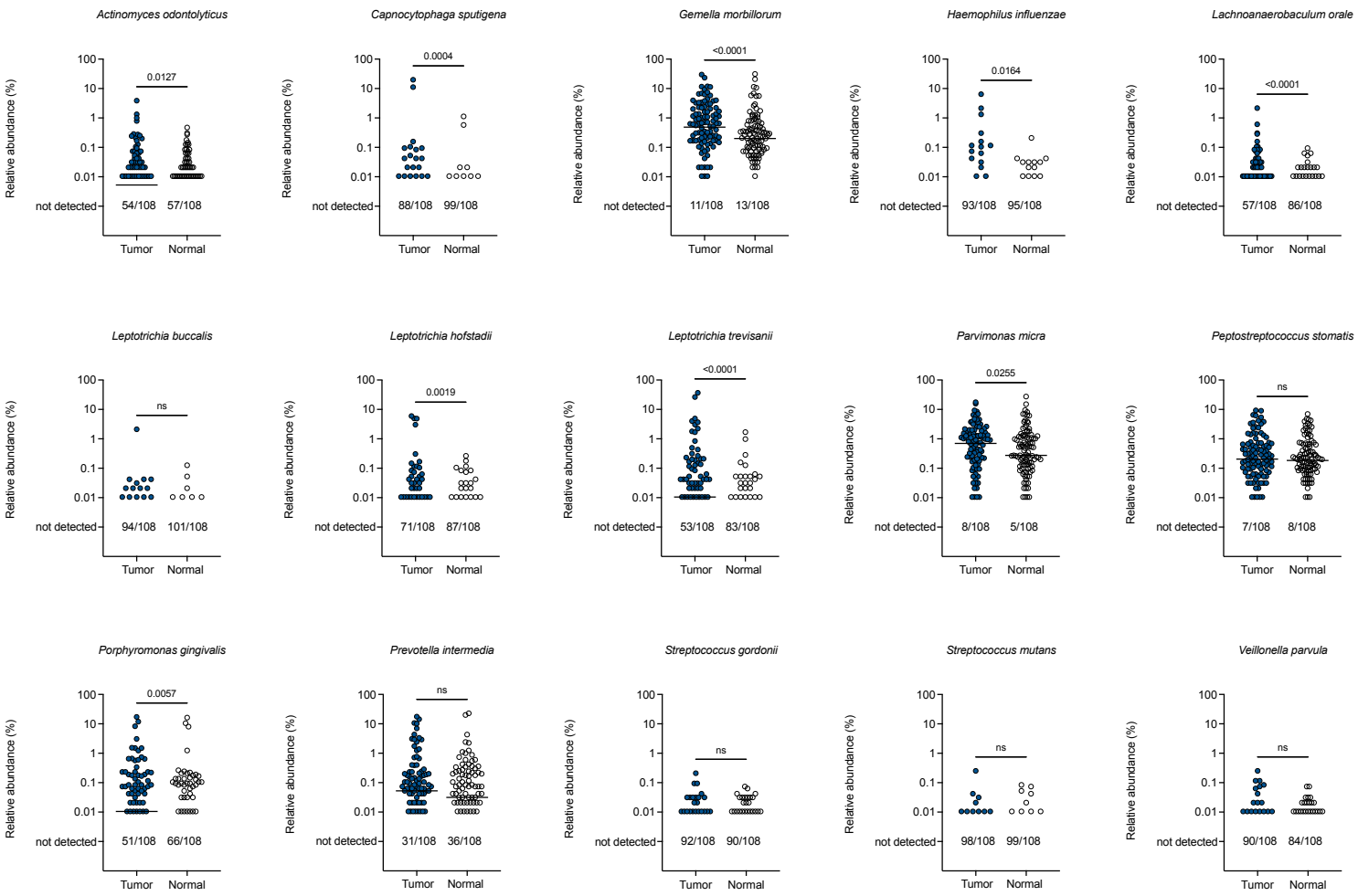

Figure S9. Abundance of oral biofilm organisms in CRC tumors/normal pairs. Relative abundance by 16S rRNA amplicon sequencing of each depicted species in pairs of tumor (blue circle) and normal (clear circle) tissues. Bars indicate median. Of 112 individuals with CRC in the study, 108 tumor and normal pairs were available for 16S rRNA sequencing analysis. Numbers below each graph depict the number of samples out of the total in which each species was not detected by sequencing. Analyzed by two-tailed Wilcoxon matched-pairs signed rank test, with $p<0.05$ considered significant.

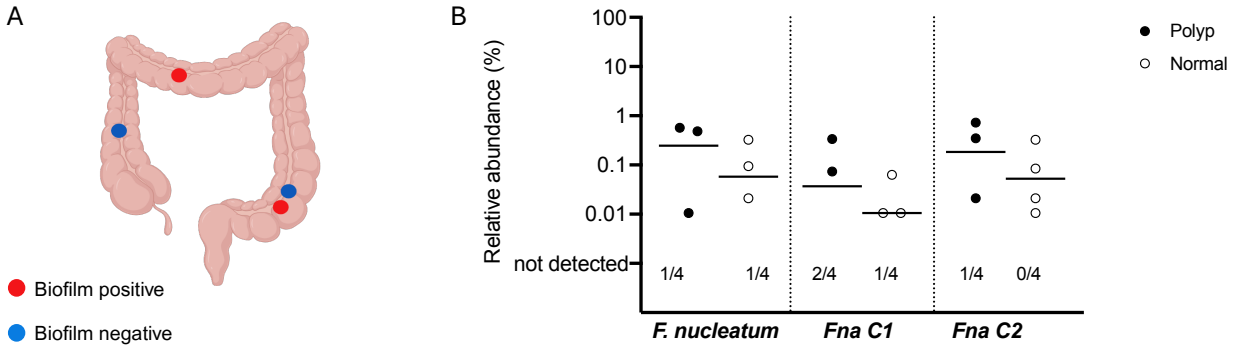

Figure S10. Adenomas in the cohort. A) Diagram of adenoma (polyp) location in the colon and biofilm status (red circle = biofilm-positive; blue circle = biofilm-negative). B) Relative abundance by 16S rRNA amplicon sequencing of *Fusobacterium nucleatum*, *subspecies animalis* clade 1 (Fna C1) and clade 2 (Fna C2) in pairs of polyp (black circle) and normal (clear circle) tissues. Bars indicate median. N = 4 polyp:normal pairs. Numbers below each graph depict the number of samples out of the total in which each species was not detected by sequencing.

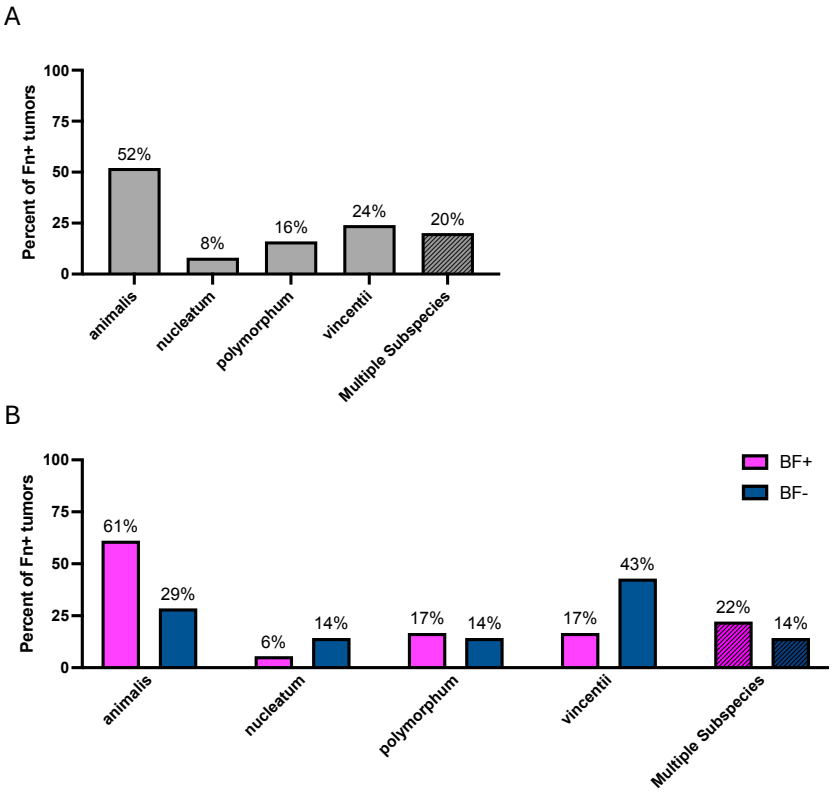

Figure S11. *Fusobacterium nucleatum* subspecies detected in tumors by PCR. Tumor DNA samples positive for *F. nucleatum* (Fn+) using primers for *nusG* (n=25) were screened with subspecies-specific PCRs [16]. A) Percent of these tumors in which bands were detected for each subspecies or for multiple subspecies, and B) stratified by the biofilm (BF) status of the tumor.

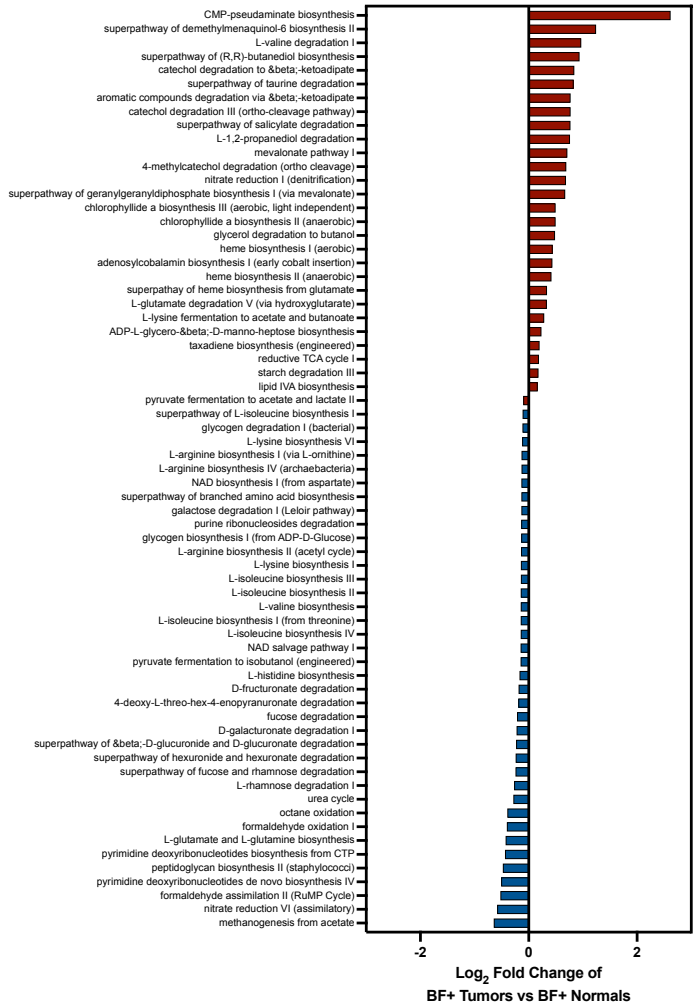

Figure S12. PICRUSt2 predictions of metabolic function associated with tissue type. All PICRUSt2 pathways with significant differences between tumors and normal tissues in individuals with biofilms (N=79) . This data matches the left side of the heatmap in Fig 6A. Y-axis depicts the Log_2 Fold change between groups (>0 indicates higher in tumors). Analyzed by Mann Whitney test, with $p<0.05$ considered significant.

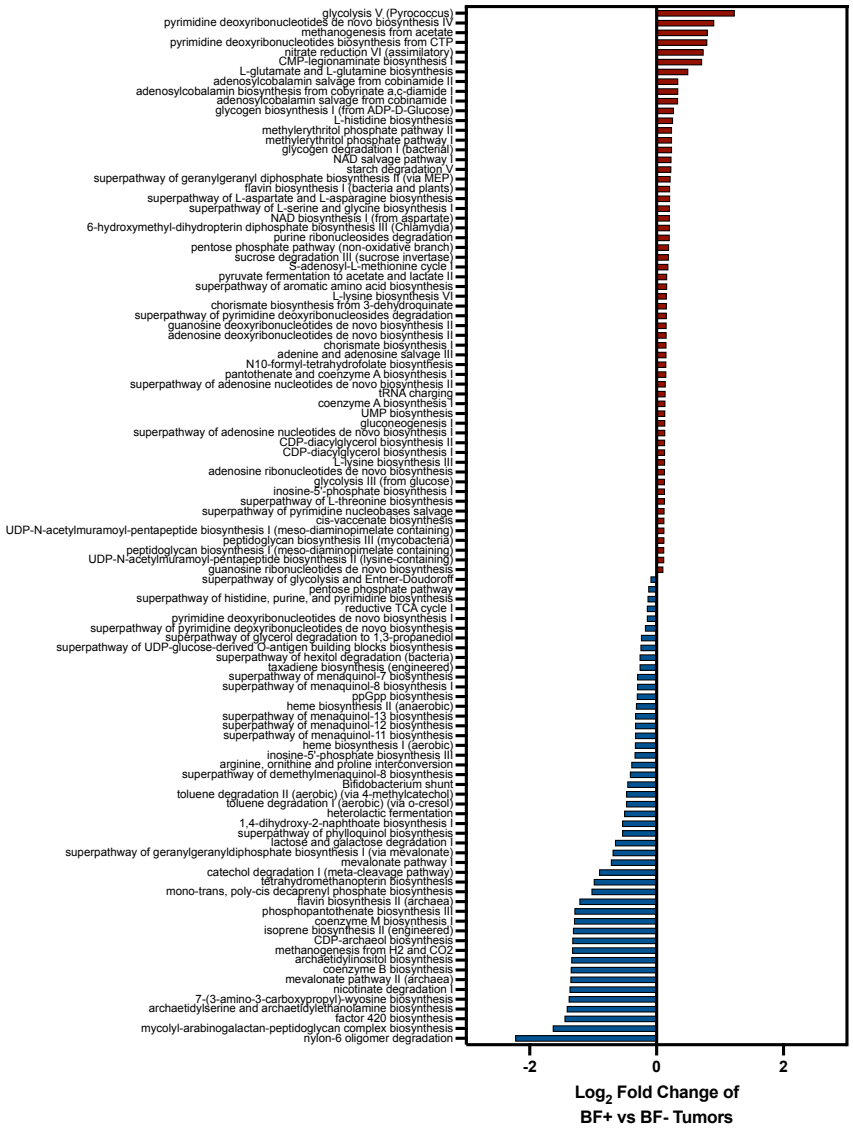

Figure S13. PICRUSt2 predictions of metabolic function associated with biofilm status in tumors. All PICRUSt2 pathways with significant differences between biofilm positive tumors (N=79) and biofilm-negative tumors (N=34). Y-axis depicts the Log_2 Fold change between groups (>0 indicates higher in BF+ tumors). Analyzed by Mann Whitney test, with $p<0.05$ considered significant.

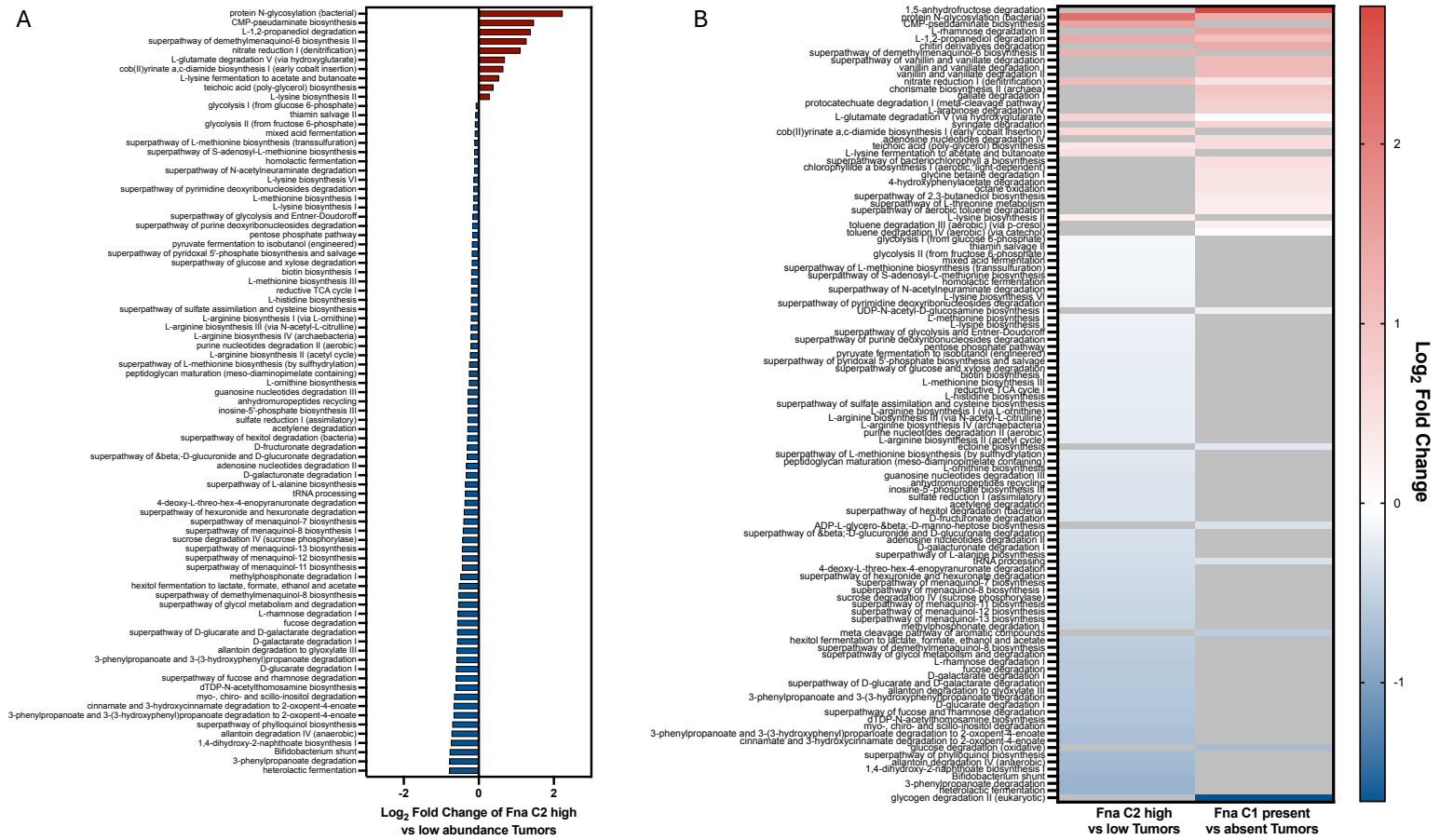

Figure S14. PICRUSt2 predictions of metabolic function associated with Fna C2 abundance. A) All PICRUSt2 pathways with significant differences between high Fna C2 abundance (≥1%) and low abundance (<1%) in tumors (n=112). Y-axis depicts theLog2 Fold change between groups (>0 indicates higher in Fna C2 high abundance tumors). Analyzed by Mann Whitney test, with $p<0.05$ considered significant. B) Heatmap of all statistically significant PICRUSt2 pathways with differences in gene content in n=112 tumors for either high Fna C2 abundance (≥1%) vs low abundance (<1%) (left) or present Fna C1 (>0%) vs absent (0% abundance) tumors (right), as analyzed by Mann Whitney test, with $p<0.05$ considered significant. Data depicted as Log_2 Fold Change.

Supplementary Tables

- Extended Data 1: 16S Sequencing results of 10 replicates
    - Table S1A: pre-processing data
    - Table S1B: phylum
    - Table S1C: class
    - Table S1D: order
    - Table S1E: family
    - Table S1F: genus
    - Table S1G: species/OTU

- Extended Data 2: 16S Sequencing results of full cohort
    - Table S2A: 16S alpha diversity
    - Table S2B: 16S phylum
    - Table S2C: 16S class
    - Table S2D: 16S order
    - Table S2E: 16S family
    - Table S2F: 16S genus
    - Table S2G: 16S species/OTU
    - Table S2H: Fna clades
    - Table S2I: PICRUSt2 pathways

- Table S3: *Fusobacterium* strains isolated by culture

- Extended Data 4: Taxa Contributing to PICRUSt2 pathway enrichment
    - Table S4A: Comparison of biofilm-positive tumor and biofilm-positive normal tissues
    - Table S4B: Comparison of biofilm-positive tumor and biofilm-negative tumor tissues
    - Table S4C: Comparison of Fna C2 high (≥1%) tumor and Fna C2 low (<1%) tumor tissues