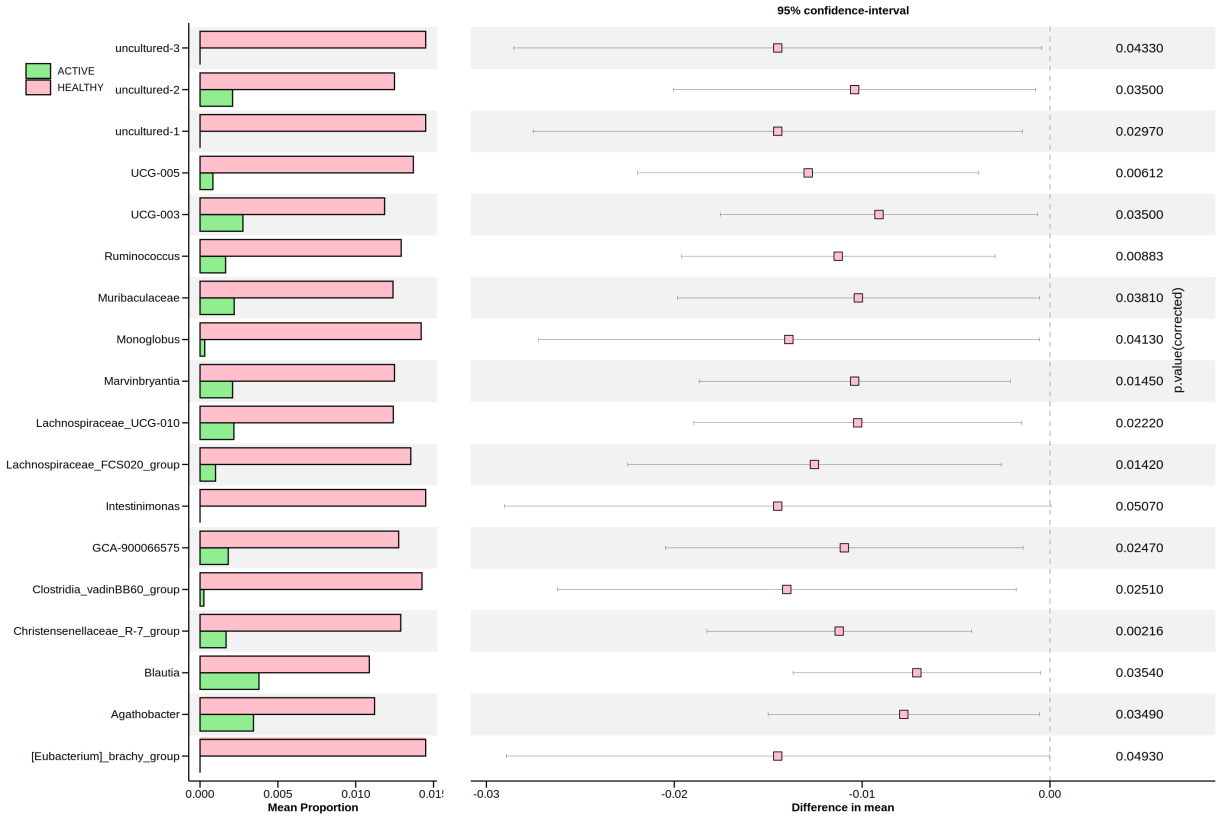
I. DNA extraction from tissue and stool samples and library preparation

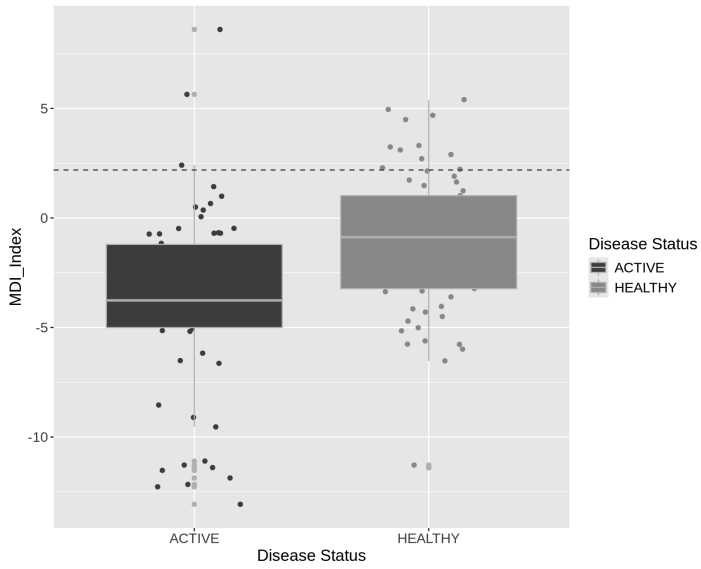
Stool DNA extraction: DNA extraction from stool was done using the QIAamp FastDNA stool mini kit (Qiagen) according to the manufacturer's instructions. The vortexed suspension of Inhibitex buffer and stool (200mg) was added to 2ml microcentrifuge tube containing 370 mg of 0.1 mm zirconia/silica beads (BioSpec) and Cocktail A which included 1.67 µl of lyzozyme (30 mg/ml - Sigma), 13 µl of mutanolysin (11.7 U/µl)(Sigma), and 3 µl of lysostaphin (4.5 U/µl)(Sigma). Samples were then incubated at 37°C on a plate thermo-shaker (PST-60HL-4, Biosan) at 250 rpm for 10 minutes. Following incubation, Cocktail B containing 10 µl of proteinase K (Qiagen), 50 µl of 10% sodium dodecyl sulphate (Sigma) and 20 µl of RNase A (1mg/ml) (Sigma) was added to each sample before second incubation step in water bath at 70°C for 10 minutes. After allowing the samples to cool for 3 minutes, they were vortexed at high speed for 10 minutes followed by microcentrifuging at 13,200 rpm for 1 minute. The supernatant obtained was then transferred into a new microcentrifuge tube and centrifuged at 13,200 rpm for 5 minutes. All supernatant (approximately 600 µl) was transferred into a new 2 ml tube, and the pellet was discarded. The supernatant was centrifuged at 13,200 rpm for 3 minutes, then 400 µl added to a new 2 ml tube containing 400 µl of buffer AL. 400 µl of ethanol (96–100%) was added to the lysate and mixed by vortexing. The manufacturer's instructions were then resumed until final elution in 50 µl of AE Buffer. An extraction control was carried out with each batch of DNA extraction.

Tissue DNA extraction: DNA extraction from tissue was carried out using the QIAamp DNA mini kit (Qiagen) according to the manufacturer's instructions. 180 µl of ATL buffer and 20 µl of proteinase K were added to a 1.5ml microcentrifuge tube containing ≤25mg of tissue specimen. The sample was lysed at 56°C in a water bath for 1-4 hrs with intermittent vortexing (until the specimen was completely lysed). To the lysed sample, 200 µl of AL buffer was added, mixed by vortexing and incubation in a water bath at 70°C for 10 minutes. To this, 200 µl of ethanol (96-100%) was added and vortexed briefly for 15 secs. After a short centrifuge step, the mixture was pipetted onto QIAamp mini spin column. Final elution was done in 100 µl of AE Buffer. An extraction control was carried out with each batch of DNA extraction.

Library Preparation: Amplicons spanning 16S rRNA gene variable regions 3 and 4 (primers 341F 5'-overhang-CCTACGGGNGGCWGCAG-3' and 785R 5'-overhang-GACTACHVGGGTATCTAATCC-3') were produced following the established Illumina protocol (Illumina, 16S Metagenomic Sequencing Library Preparation Protocol Part # 15044223 Rev. B) with following modifications. The details of library preparation are provided in the supplementary material. Extracted DNA from stool and tissue samples were normalized to 5ng/µl and 2.5 µl of normalized stool DNA, and 5 µl of normalized tissue DNA was used as starting template for first-stage PCR using Kapa Hifi Hotstart MM(Roche) with a final primer concentration of 0.2µM per reaction. Cycling conditions for first stage PCR were as follows: initial denaturation at 95°C for 3 minutes; 25 cycles of 95°C for 30 seconds, 55°C for 30 seconds, and 72°C for 30 seconds; and final extension at 72°C for 5 minutes. Amplicons obtained were cleaned using AMPure XP beads at 1.2:1 bead: sample ratio and fragment size estimated by Agilent 4200 Tapestation (expected fragment length around 550bp). To determine volumes required for equimolar pooling, amplicon concentrations were determined by Qubit HS DNA kit on Qubit 3 fluorometer (ThermoFisher Scientific), and fragment size checked by High Sensitivity D1000 DNA kit on Agilent 4200 Tapestation(expected fragment size around 630bp). 5 µl each of 4nM normalized amplicons were pooled together, and the library diluted post-denaturation to a final loading concentration of 8 pM. Libraries were sequenced with 20% PhiX (Illumina) spiked in using the 2×300 bp v3 chemistry on Ilumina MiSeq platform (Next generation sequencing – NGS).

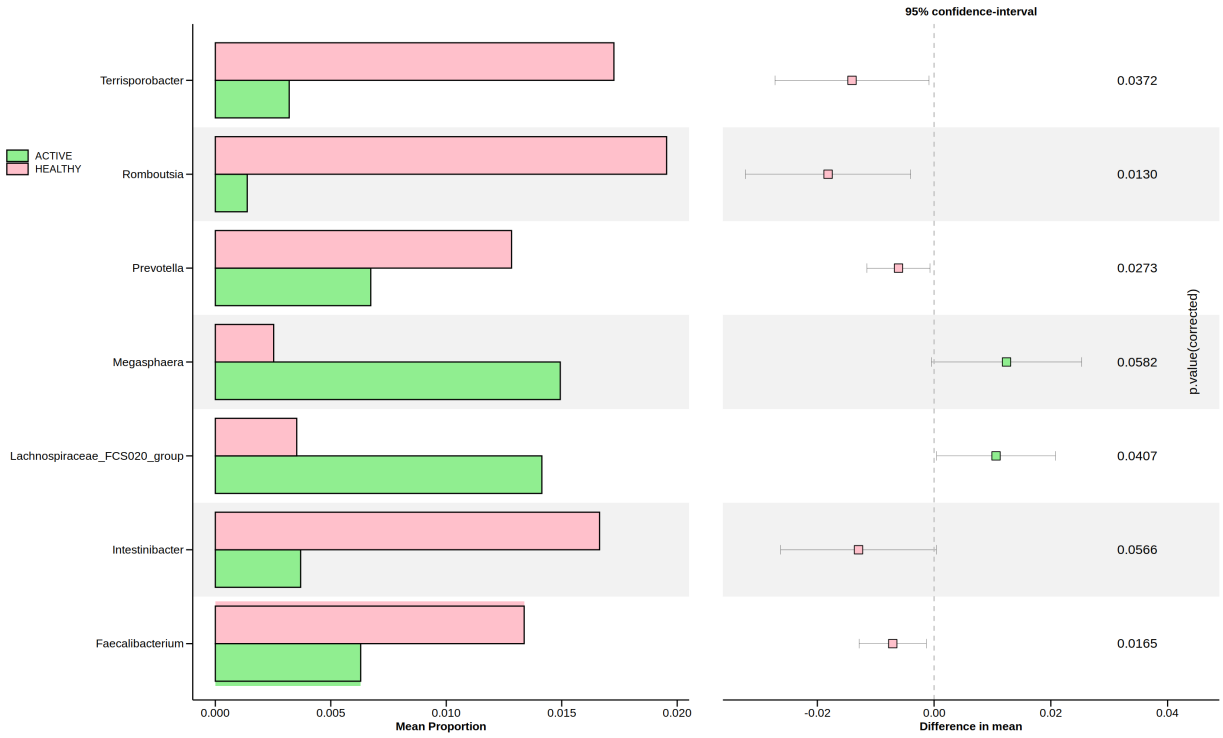


Supplementary Figure 1: Bacterial genera with significant differences in relative abundance between tissue samples of patients with active Crohn's disease and healthy controls.

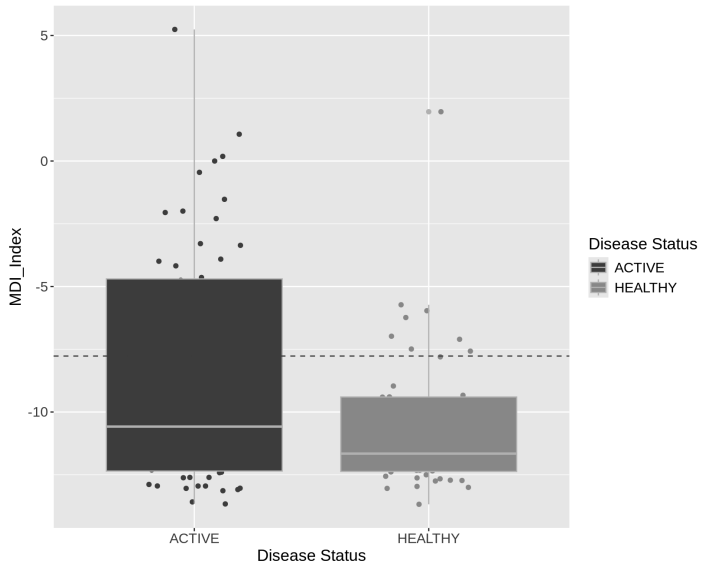


	Dysbiotic	Healthy
ACTIVE	3	63
HEALTHY	11	58

Supplementary Figure 2: Microbial dysbiosis index in tissue samples of active Crohn's disease and healthy controls

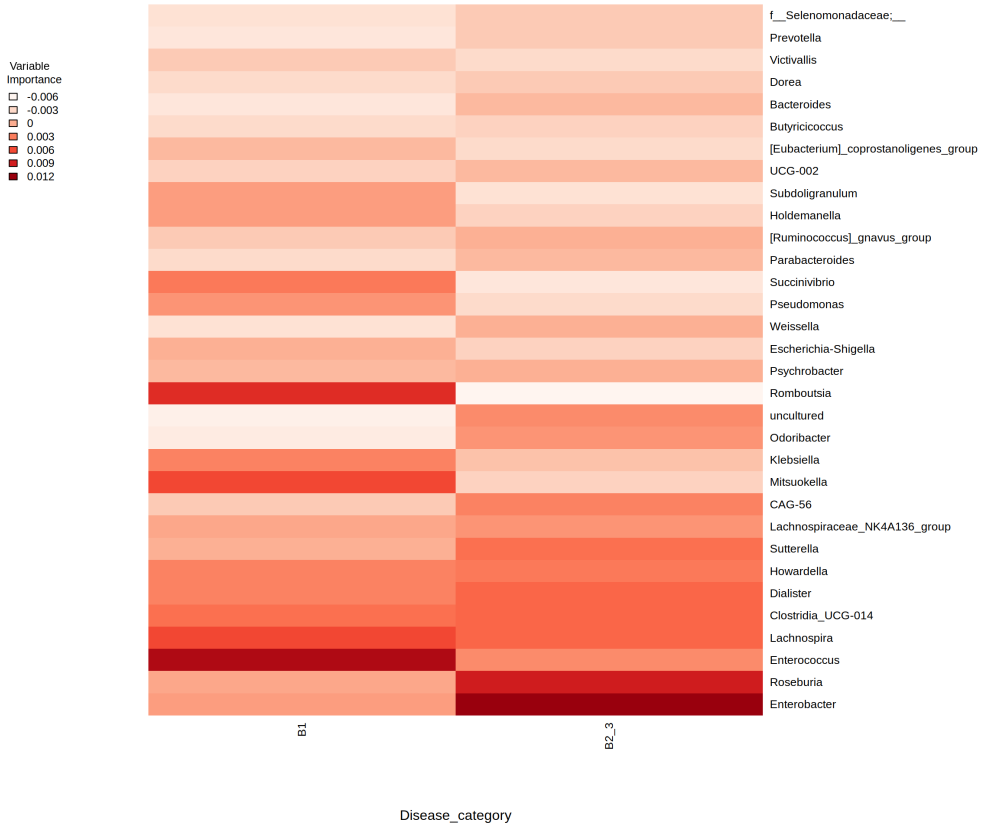


Supplementary Figure 3: Bacterial genera with significant differences in relative abundance between stool samples of patients with active Crohn's disease and healthy controls.

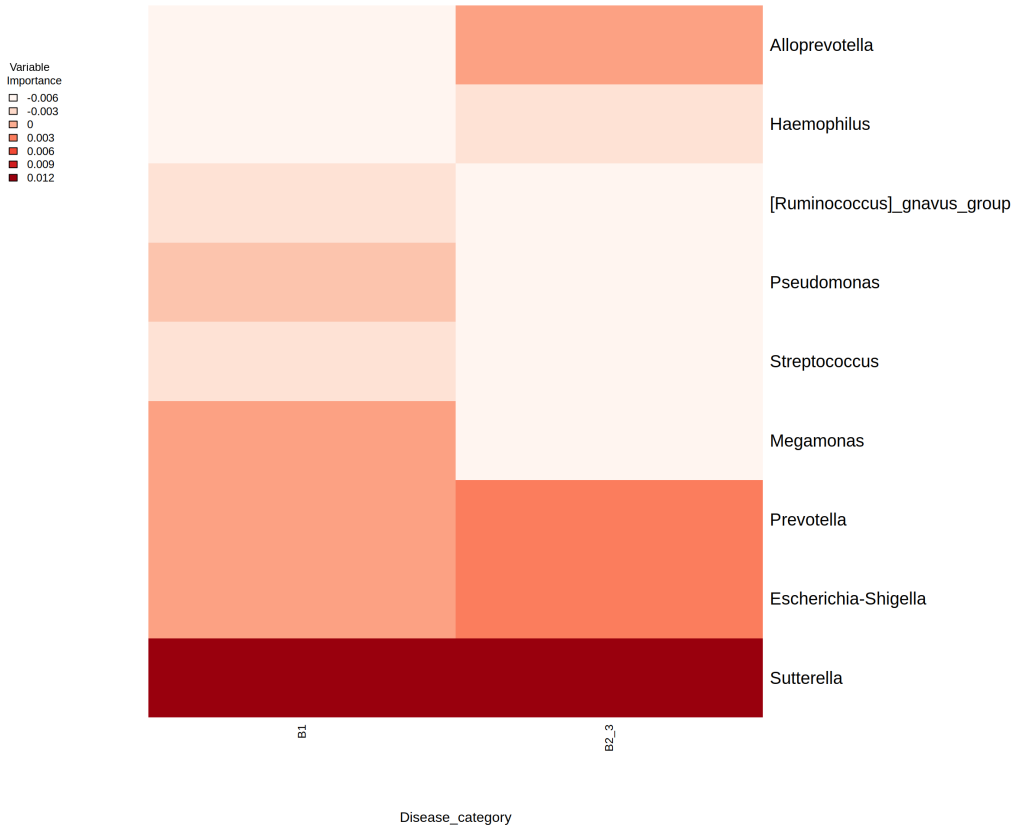


	Dysbiotic	Healthy
ACTIVE	20	39
HEALTHY	8	39

Supplementary Figure 4: Microbial dysbiosis index in stool samples of active Crohn's disease and healthy controls



Supplementary Figure 5: Random forest analysis showing fecal bacterial genera predicting aggressive (B2/B3) vs non-aggressive (B1) disease behavior in patients with active Crohn's disease.



Supplementary Figure 6: Random forest analysis showing tissue bacterial genera predicting aggressive (B2/B3) vs non-aggressive (B1) disease behavior in patients with active Crohn's disease