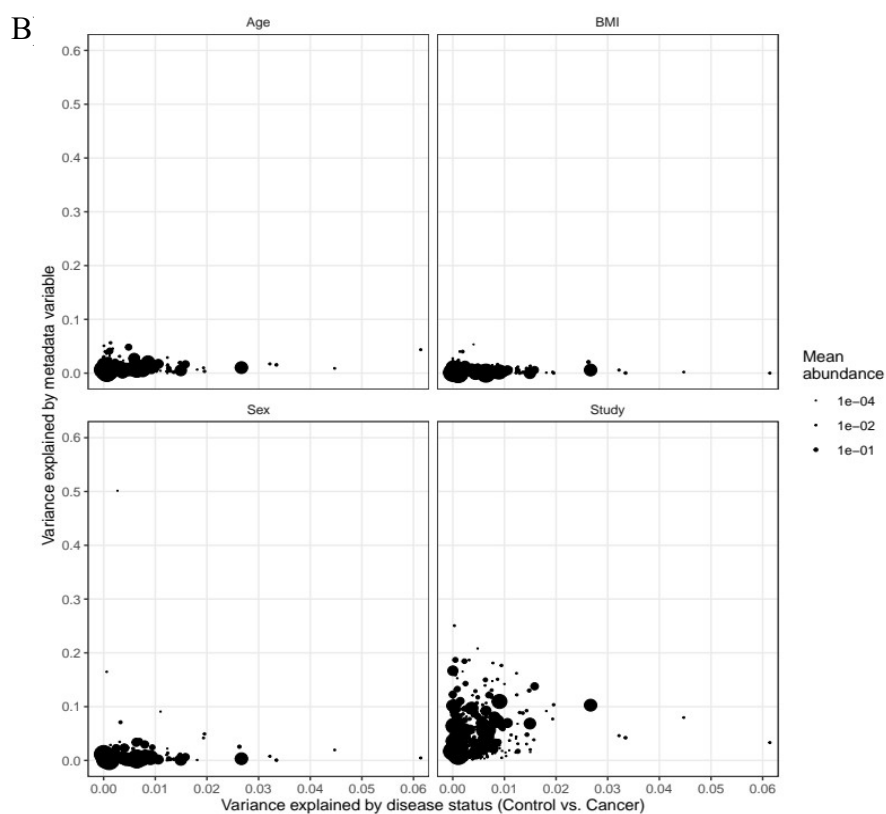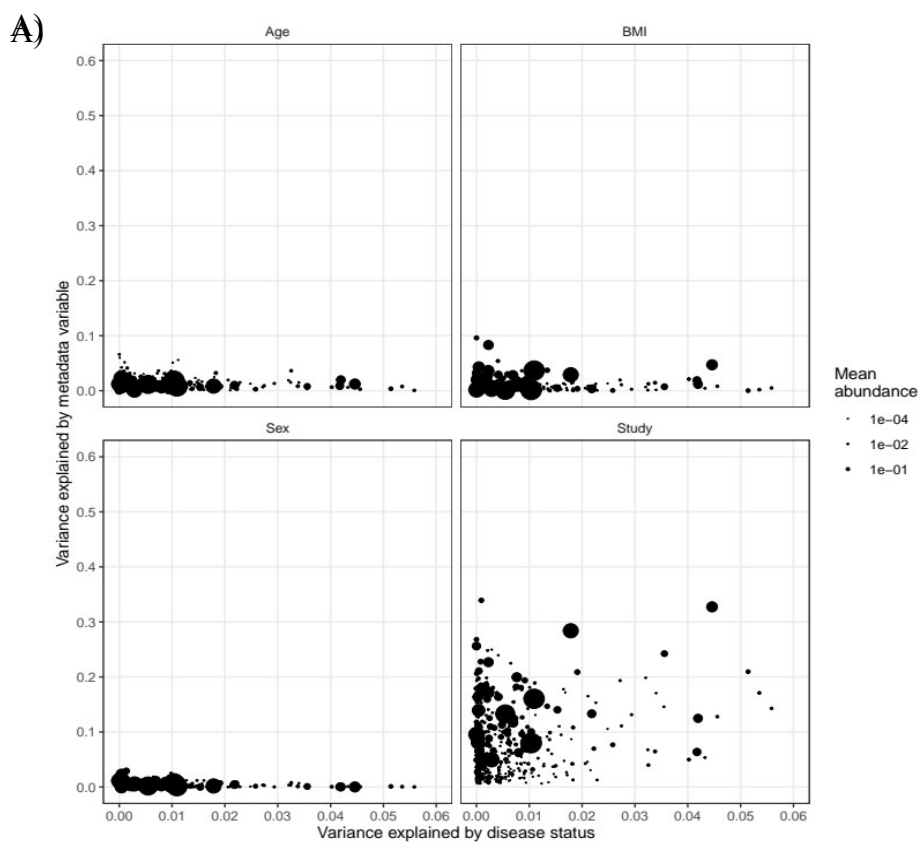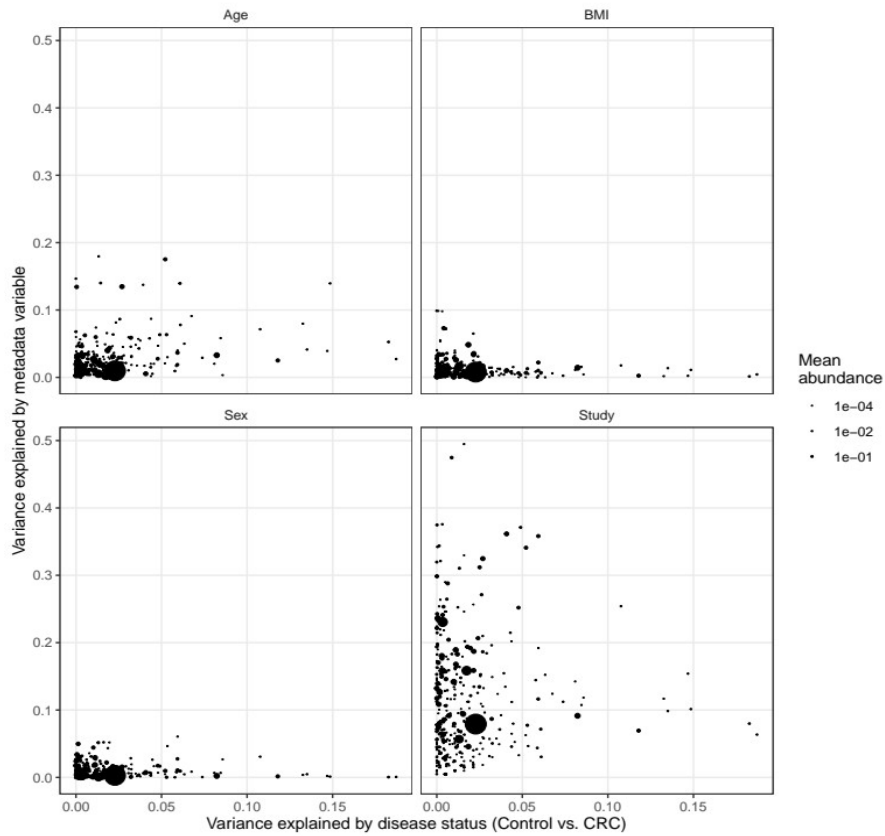**Species-Level Identification of Enterotype-Specific Microbial Markers for Colorectal Cancer and Adenoma**

Supplementary Figures

C)

Figure S1. Variance explained by disease status (Control vs. CRC) is plotted against variance explained by metadata variable for microbial species in A) Enterotype 1, B) Enterotype 2 and C) Enterotype 3
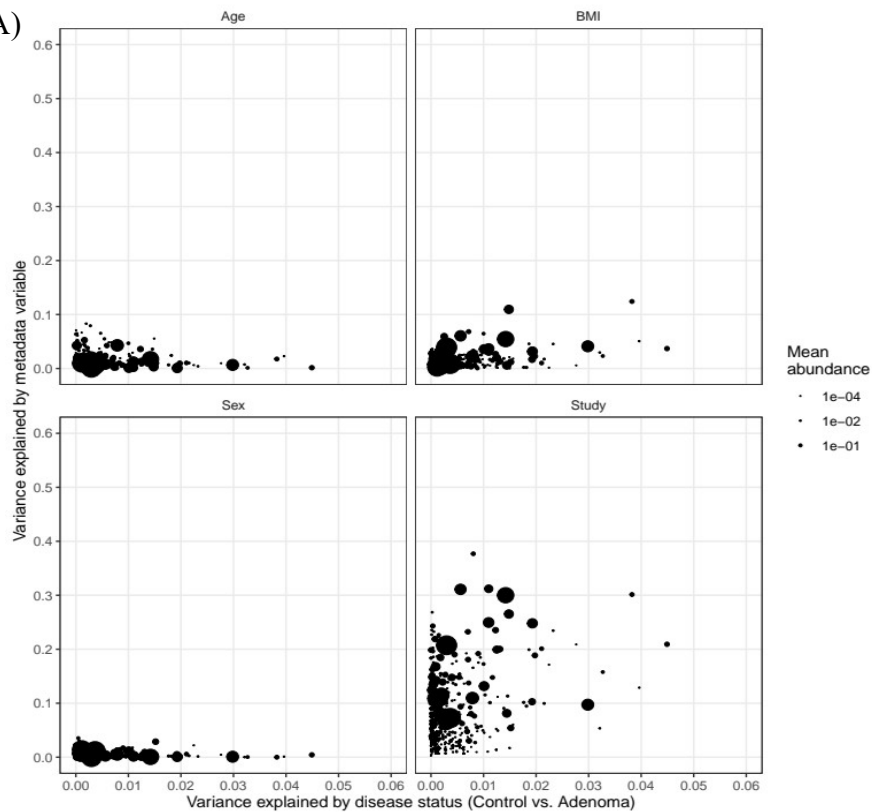
A)

Figure S2. Variance explained by disease status (Control vs. ADN) is plotted against variance explained by metadata variable for microbial species in A) Enterotype 1, B) Enterotype 2 and C) Enterotype 3.
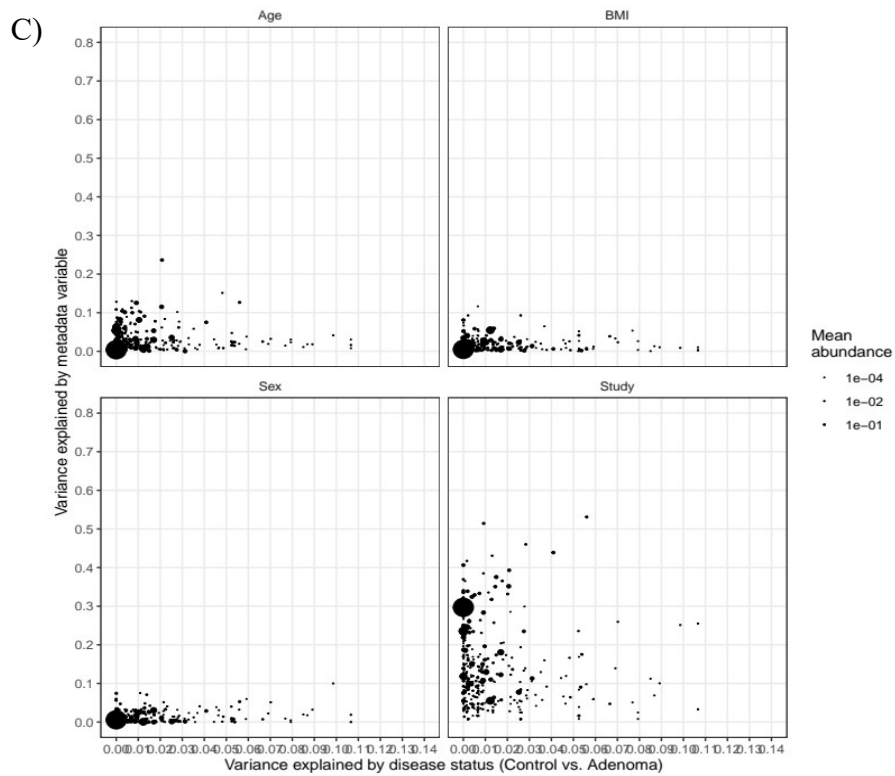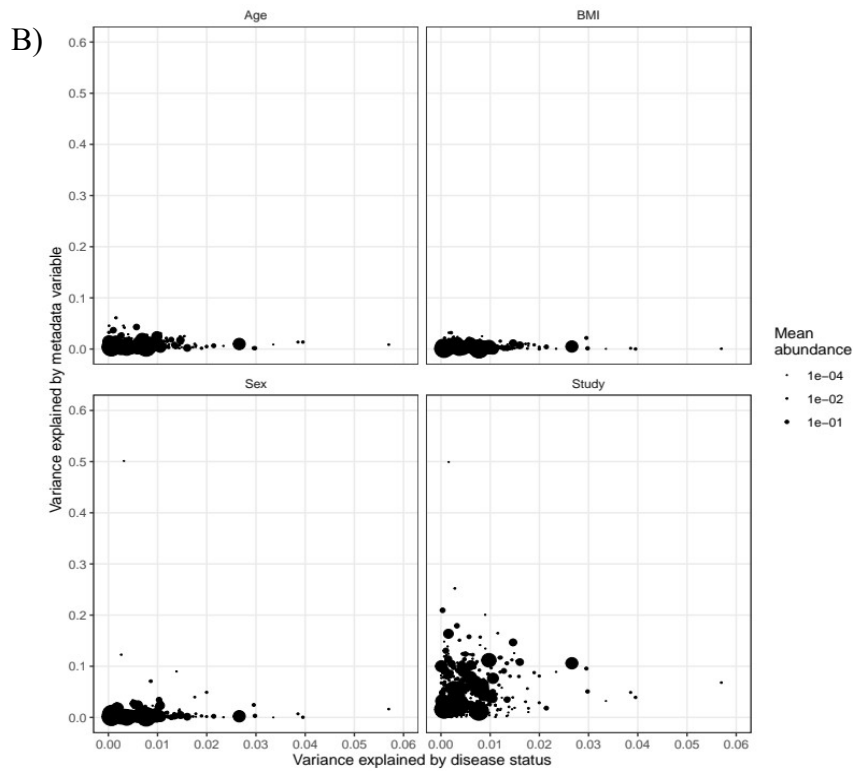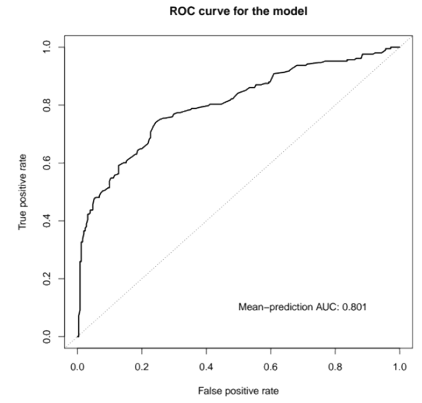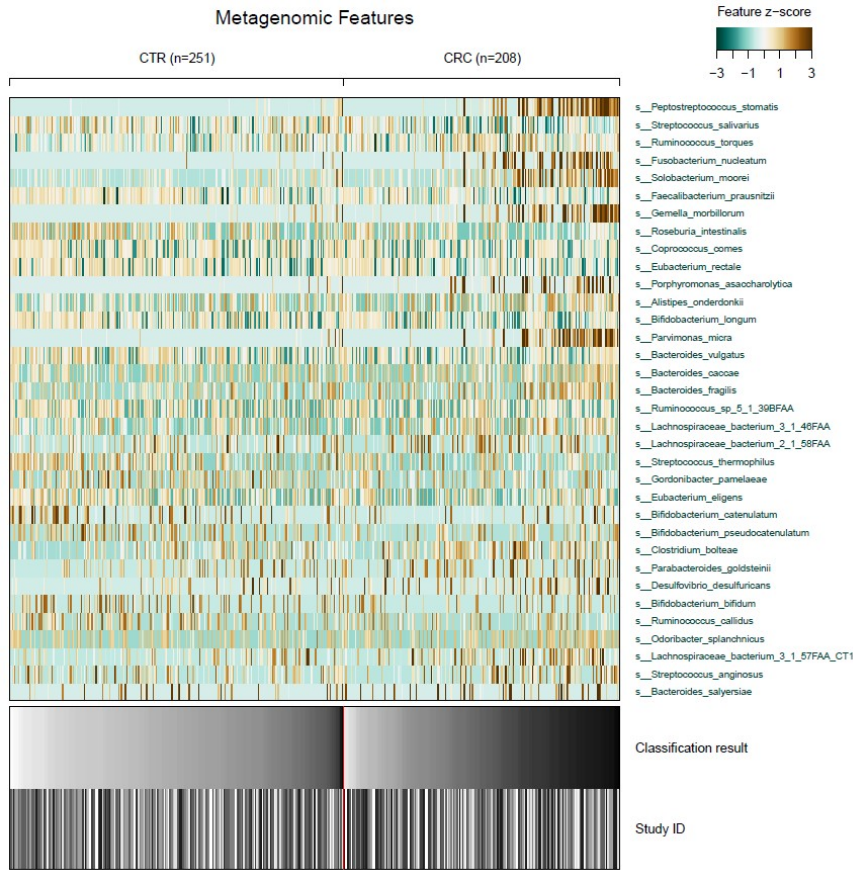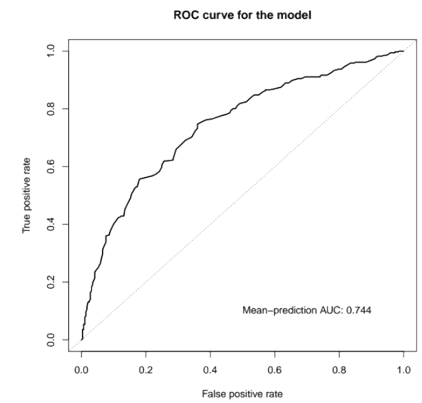
A)

Metagenomic Features

CTR (n=251) CRC (n=208)

Feature z-score

−3 −1 1 3

s__Peptostreptococcus_stomatis
s__Streptococcus_salivarius
s__Ruminococcus_torques
s__Fusobacterium_nucleatum
s__Solobacterium_moorei
s__Faecalibacterium_prausnitzii
s__Gemella_morbillorum
s__Roseburia_intestinalis
s__Coprococcus_comes
s__Eubacterium_rectale
s__Porphyromonas_asaccharolytica
s__Alistipes_onderdonkii
s__Bifidobacterium_longum
s__Parvimonas_micra
s__Bacteroides_vulgatus
s__Bacteroides_caccae
s__Bacteroides_fragilis
s__Ruminococcus_sp_5_1_39BFAA
s__Lachnospiraceae_bacterium_3_1_46FAA
s__Lachnospiraceae_bacterium_2_1_58FAA
s__Streptococcus_thermophilus
s__Gordonibacter_pamelaeae
s__Eubacterium_eligens
s__Bifidobacterium_catenulatum
s__Bifidobacterium_pseudocatenulatum
s__Clostridium_bolteae
s__Parabacteroides_goldsteinii
s__Desulfovibrio_desulfuricans
s__Bifidobacterium_bifidum
s__Ruminococcus_callidus
s__Odoribacter_splanchnicus
s__Lachnospiraceae_bacterium_3_1_57FAA_CT1
s__Streptococcus_anginosus
s__Bacteroides_salyersiae

Classification result

Study ID

ROC curve for the model

True positive rate

False positive rate

Mean−prediction AUC: 0.801

B)

Metagenomic Features

CTR (n=289) CRC (n=336)

Feature z-score

−3 −1 1 3

s__Ruminococcus_torques
s__Faecalibacterium_prausnitzii
s__Solobacterium_moorei
s__Peptostreptococcus_stomatis
s__Flavonifractor_plautii
s__Bacteroides_caccae
s__Eubacterium_ventriosum
s__Roseburia_intestinalis
s__Holdemania_filiformis
s__Fusobacterium_nucleatum
s__Streptococcus_thermophilus
s__Anaerotruncus_colihominis
s__Alistipes_onderdonkii
s__Gemella_morbillorum
s__Eubacterium_eligens
s__Odoribacter_splanchnicus
s__Bacteroides_fragilis
s__Veillonella_atypica
s__Burkholderiales_bacterium_1_1_47
s__Clostridium_scindens
s__Coprococcus_catus
s__Granulicatella_adiacens
s__Bacteroides_intestinalis
s__Coprobacter_fastidiosus
s__Alistipes_sp_AP11
s__Eubacterium_biforme
s__Prevotella_copri
s__Subdoligranulum_sp_4_3_54A2FAA
s__Megamonas_hypermegale
s__Bacteroides_nordii
s__Prevotella_stercorea

Classification result

Study ID

ROC curve for the model

True positive rate
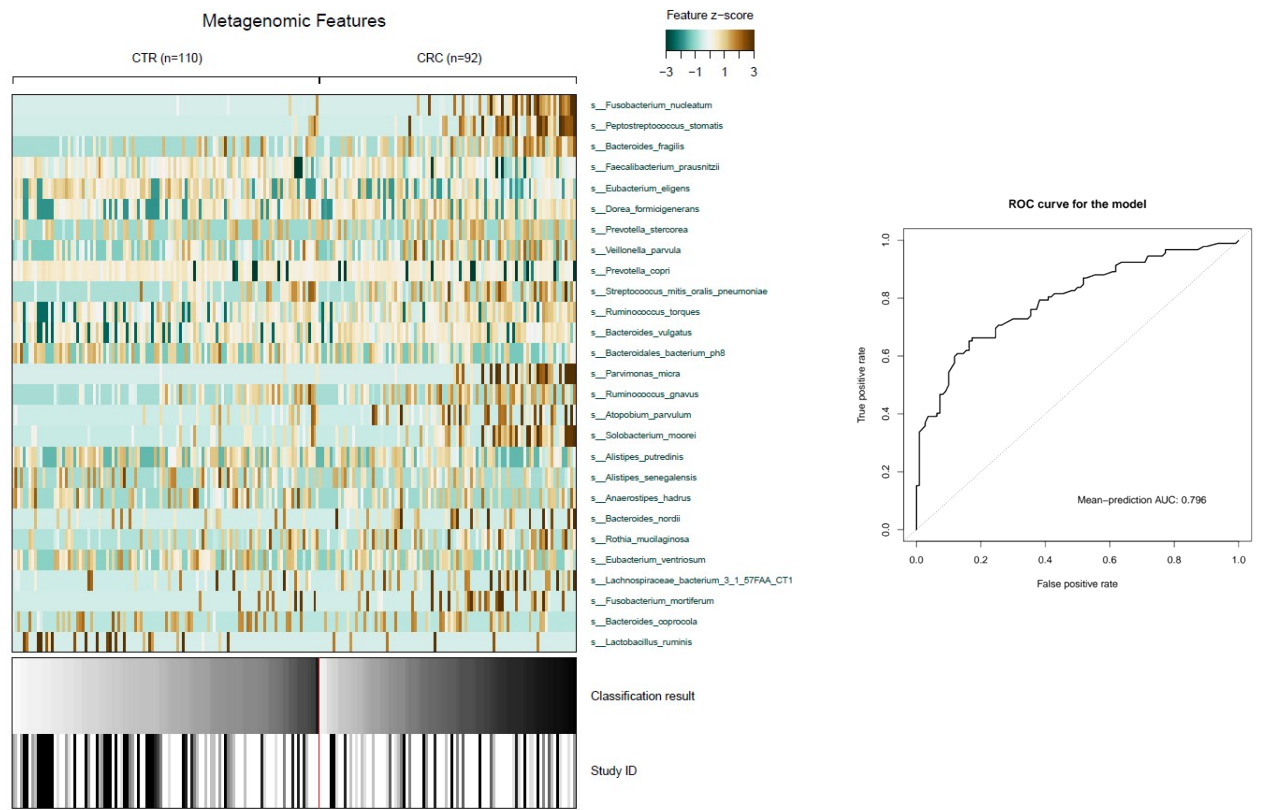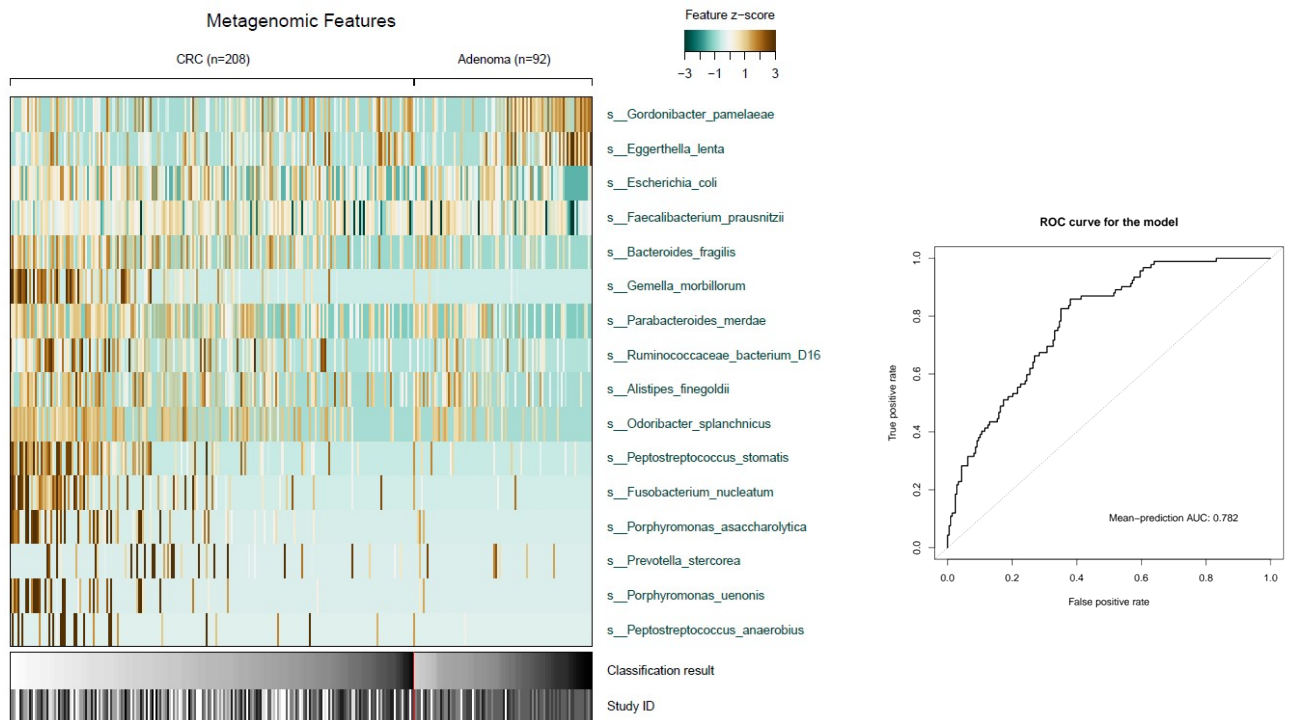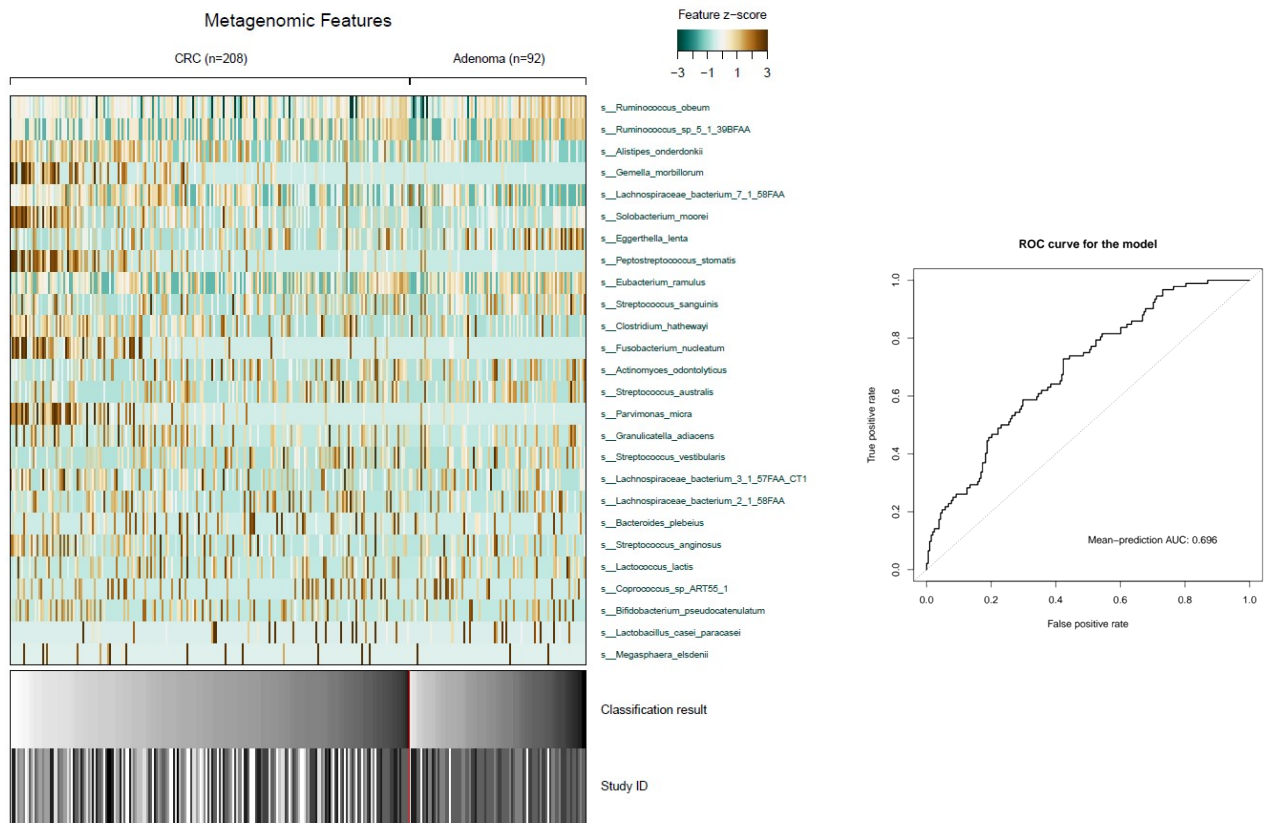
False positive rate

Mean−prediction AUC: 0.744

Figure S3. Random Forest Classifier results for distinguishing CRC from Controls in A) Enterotype 1, B) Enterotype 2, C) Enterotype 3.
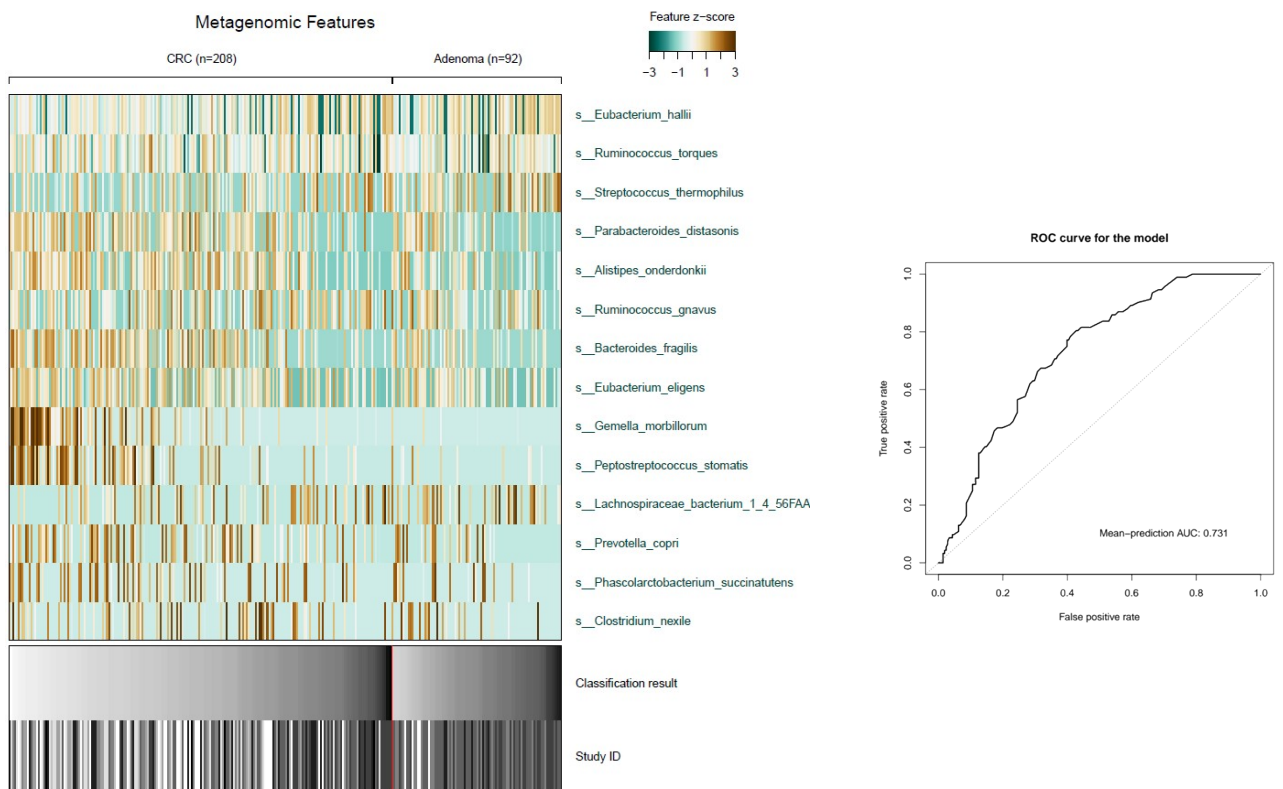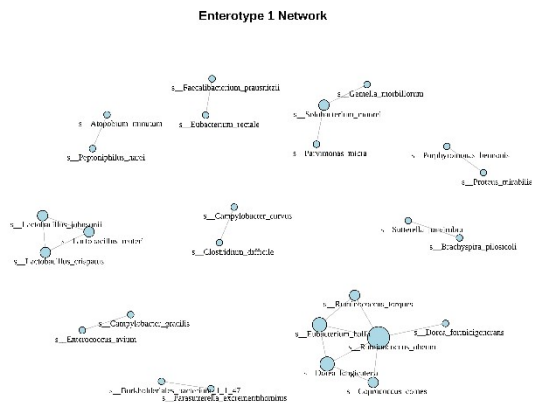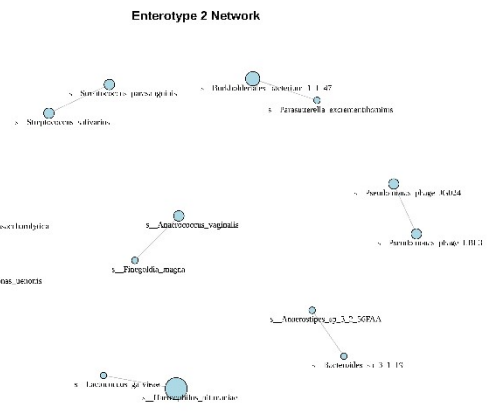
B)

C)

Figure S4. Random Forest Classifier results for distinguishing ADN from CRC in A) Enterotype 1, B) Enterotype 2, C) Enterotype 3.
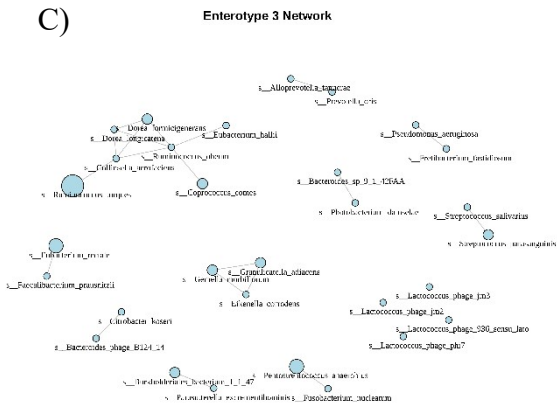
Figure S5. Microbial co-occurence network plots of enterotype classes, A) Enterotype 1, B) Enterotype2,C): Enterotype 3