Supplemental legends

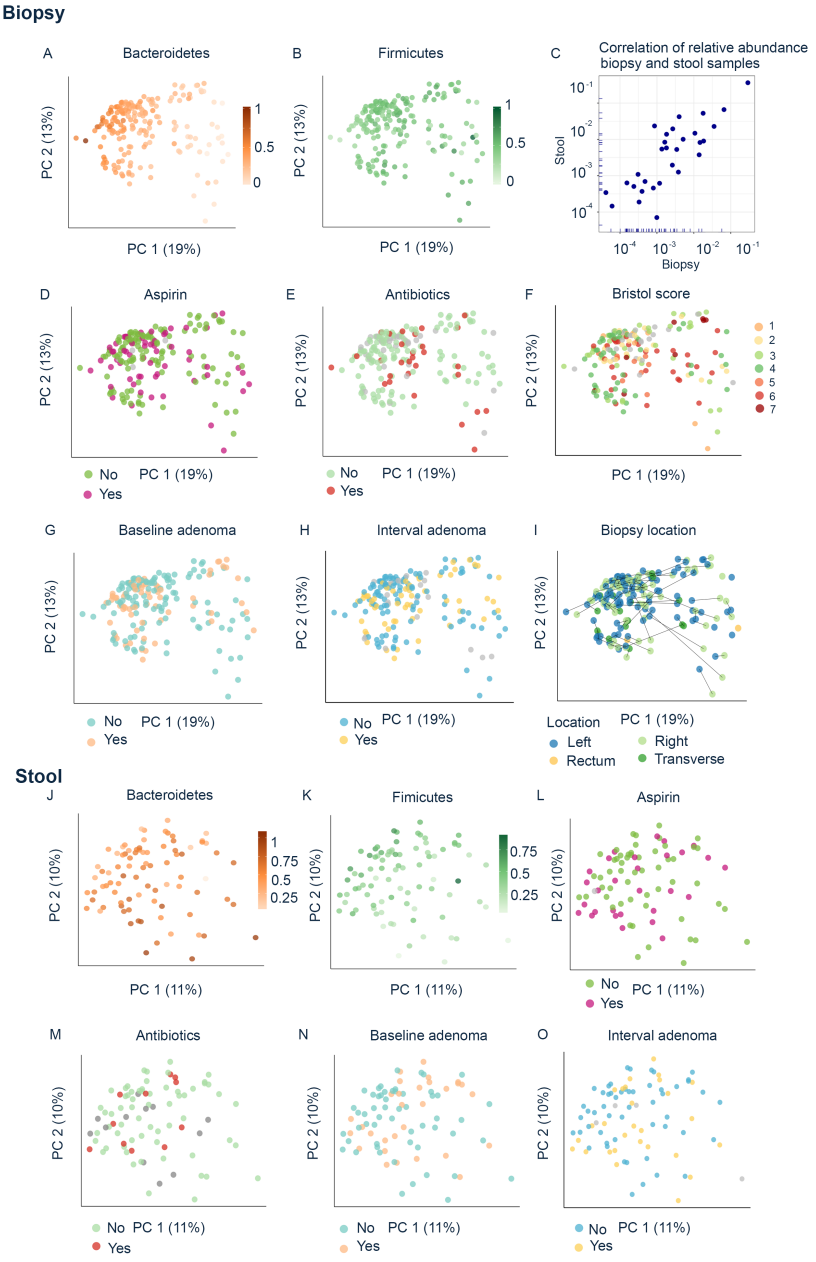

Fig. S1: Related to Figure 2, within- and between-subject structure of mucosal and stool of the Lynch syndrome gut microbiome. Principal coordinate analysis (PCoA) of biopsy genera taxonomic profiles (A-I) and stool species taxonomic profiles (J-O) based on Bray-Curtis dissimilarity matrices. Color gradient represent the relative abundance of two major clades *Bacteroidetes* (**A**) and *Firmicutes* (**B**) in biopsy. Each dot corresponds to the average relative abundance of an OTU across 85 subjects (with both biopsies and stool) (**C**). Clinical covariates were shown in (**D-H**). The location that biopsies were sampled were indicated in (**I**). Marks on the x-axis (vertical lines) or y-axis (horizontal lines) margins represent OTUs with zero measured abundance at one site but non-zero abundance at the other. Color gradient represent the relative abundance of two major clades *Bacteroidetes* (**J**) and *Firmicutes* (**K**) in stool. Clinical covariates were shown in (**L-O**).

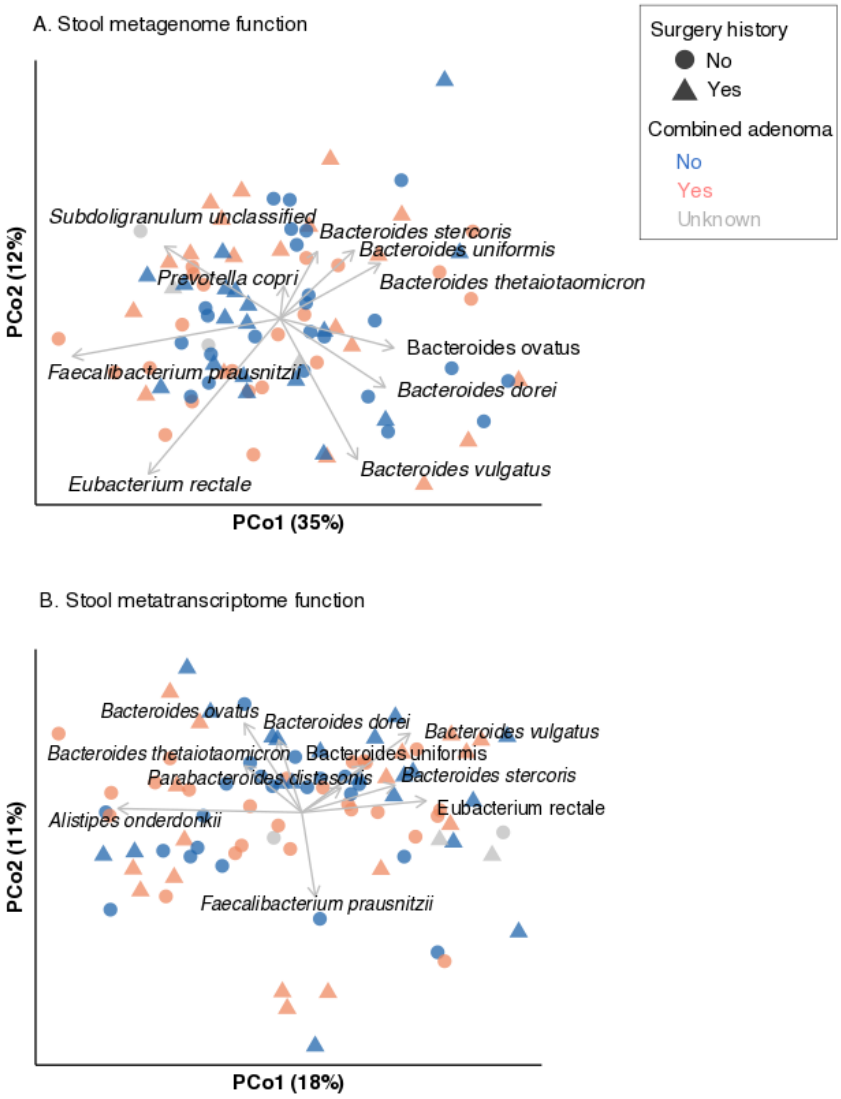

Fig. S2: Related to Figure 2, within- and between-subject structure and function of the Lynch syndrome gut microbiome. Principal coordinate analysis (PCoA) of stool (**A**) metagenomic profiles, (**B**) metatranscriptomic profiles. Biplot overlays show the top 10 most prominent species based on the relative abundance across subjects.

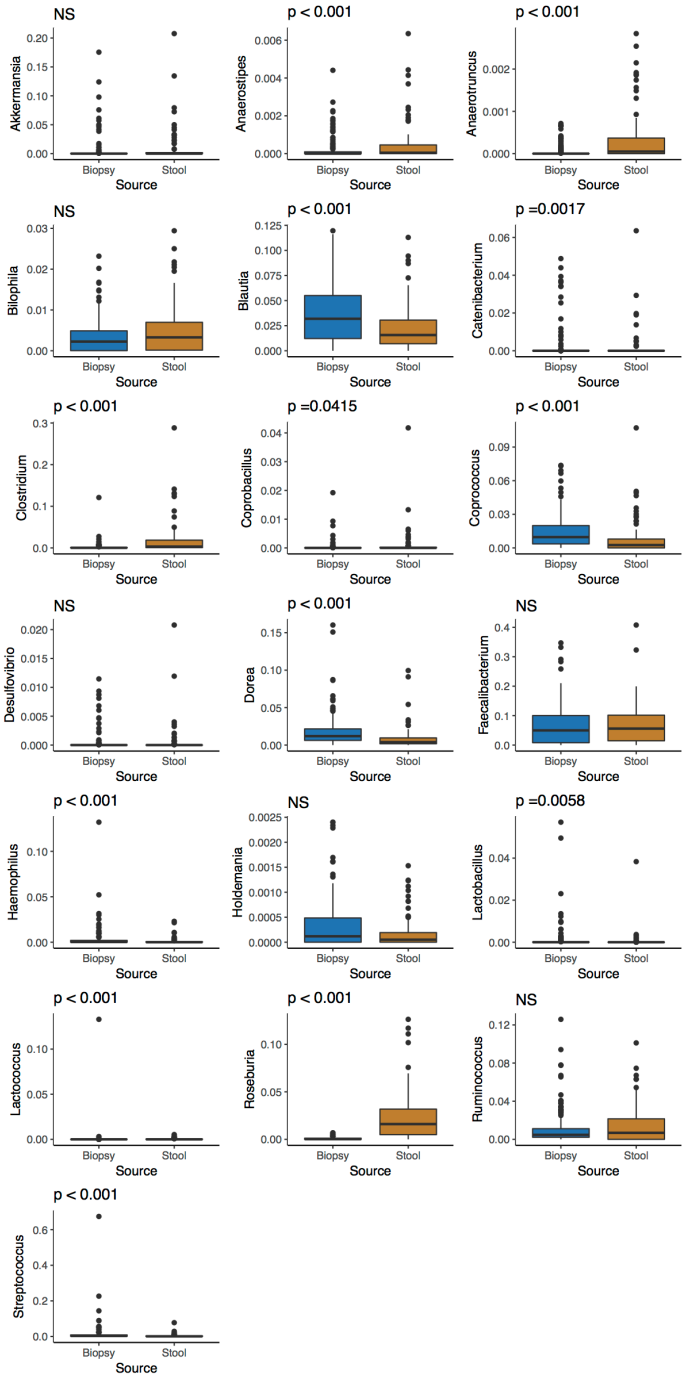

Fig. S3: Related to Figure 2, influence of sample source on the intersection genus of Lynch syndrome microbiome. 19 genera were present in both stool and biopsies samples (i.e. intersection genera). Kruskal-wallis tests analysis identified 13 taxa that were differentially abundant between the source of either mucosa and stool samples.

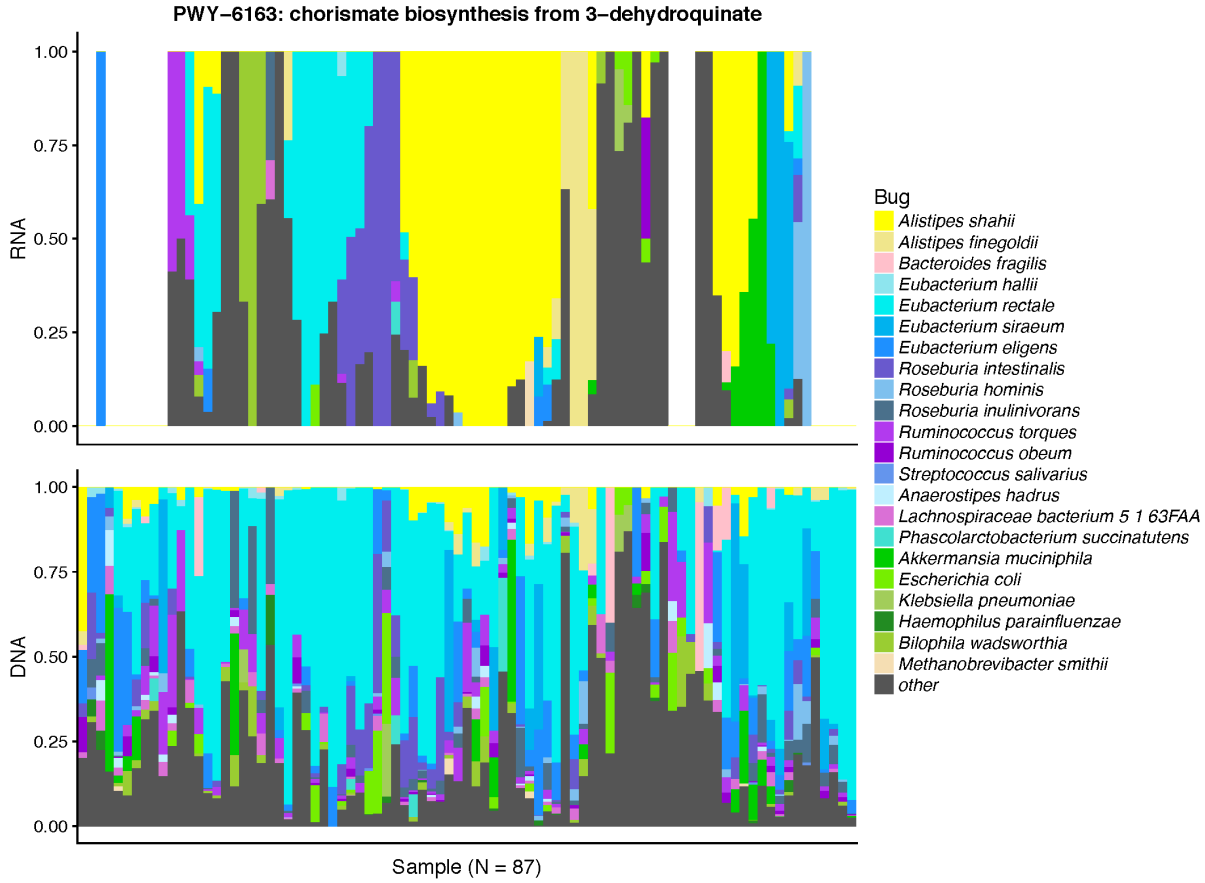

Fig. S4: Related to Figure 3, species-stratified distributions of metagenomic potential (DNA) and metatranscriptomic activity (RNA) for pathway-6163 with non-zero abundance in at least 10% of samples.

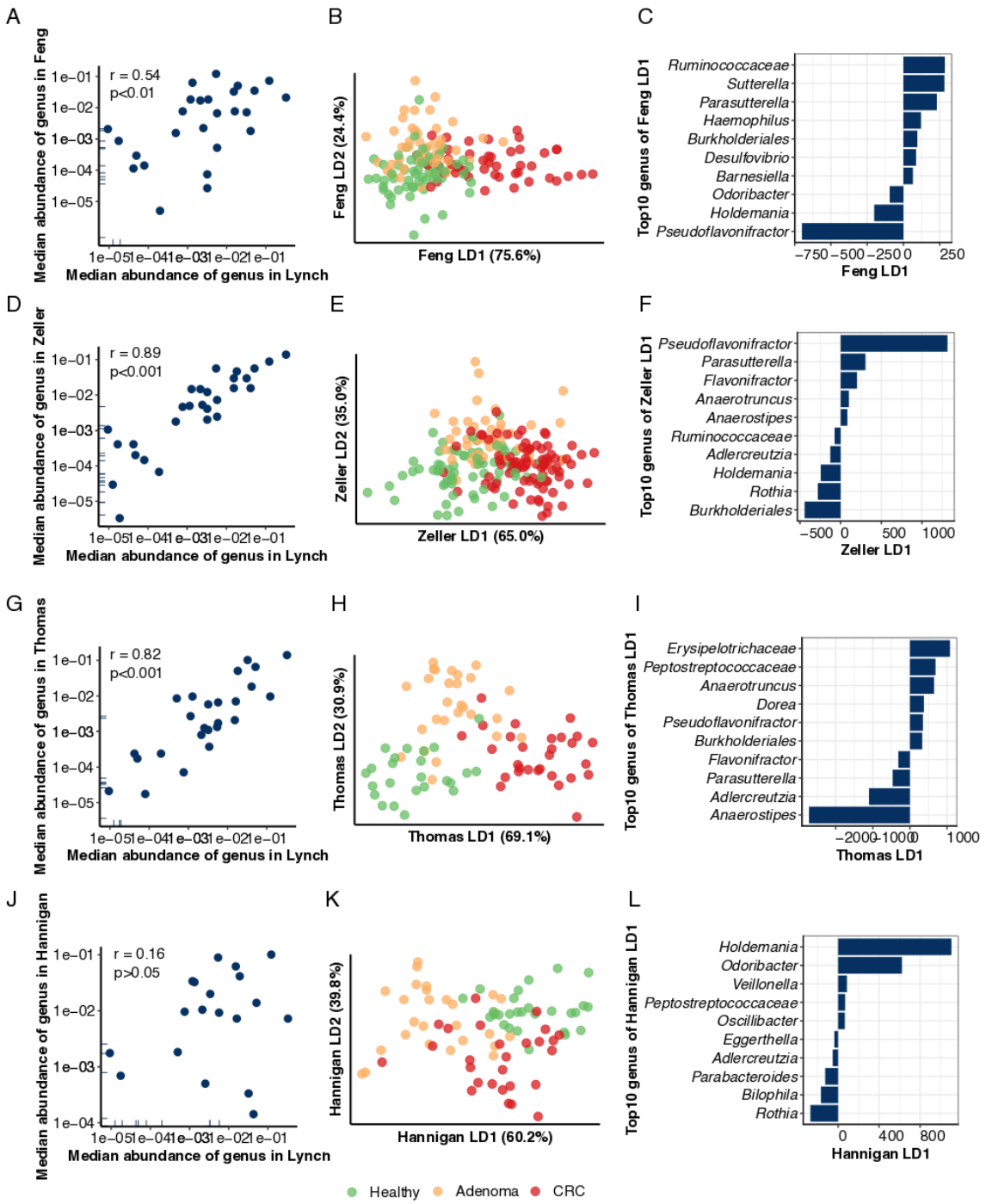

Fig. S5: Related to Figure 4, correlation of gut microbiome performance for CRC detection across cohorts. Genera associated with CRC were highly correlated between our Lynch population and the four datasets from '*curatedMetagenomicData*' (Pasolli et al., 2017) of Feng (**A-C**), Zeller (**D-E**), Thomas (**G-I**), and Hannigan (**J-L**), respectively. Linear discriminant analysis (LDA) was typically able to weakly distinguish control, adenomatous, and cancer patients in each of these microbiome studies (**B**), (**E**), (**H**) and (**K**). The 10 genera most highly loaded by LDA for this classification are shown in (**C**), (**F**), (**I**) and (**L**).

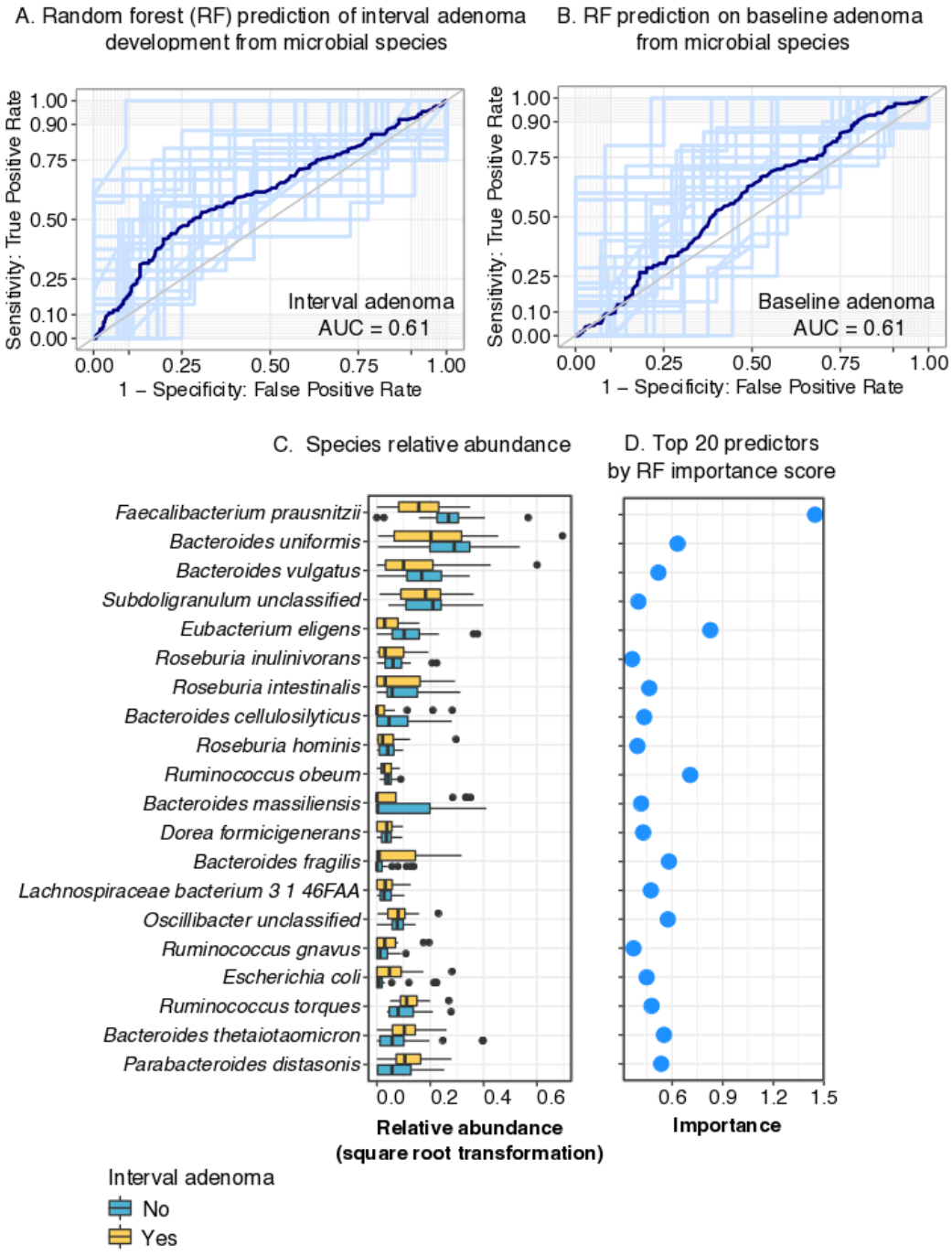

Fig. S6: Related to Figure 7, gut microbial species are not discriminative of baseline or interval adenoma development. Of random forest (RF) classifiers evaluated to predict current or one-year interval adenoma from taxonomic features (species) of the stool microbiome, neither (**A**) interval adenomas nor (**B**) baseline adenomas were predicted significantly better than chance. (**C** and **D**) The 20 species given the highest importance scores by the RF are shown, sorted by differential abundance that were signed by Gini index, along with their abundances in subjects with and without interval adenoma development. Note that this does not exclude overall microbial community configuration, nor individual microbial features, from being significantly associated with adenoma development as discussed in the text, since significance of association is a much lower bar than accuracy of discriminative prediction.

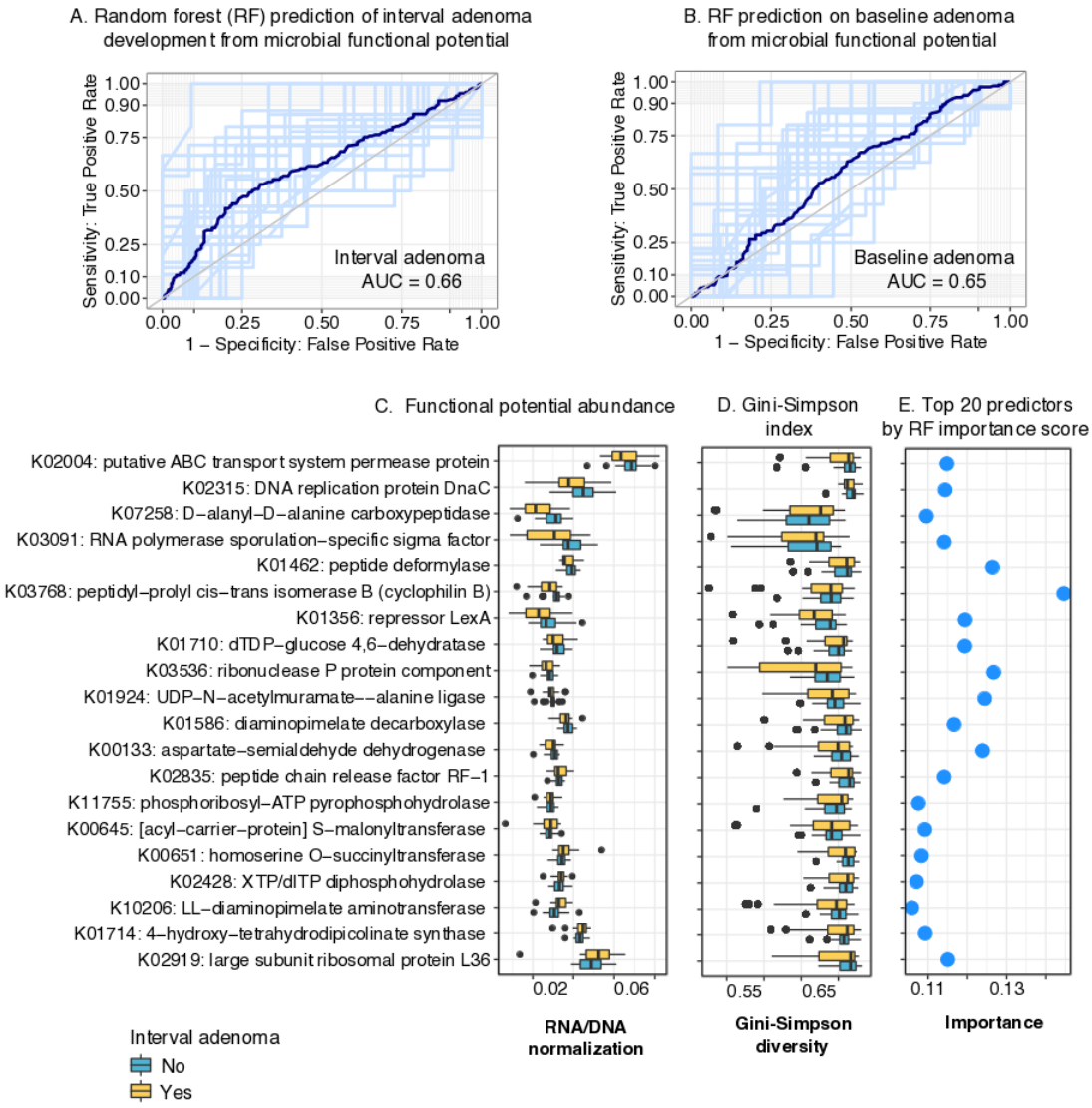

Fig. S7: Related to Figure 7, gut metagenomic functional potential is not discriminative of baseline or interval adenoma development. Of random forest (RF) classifiers evaluated to predict current or one-year interval adenoma from functional features of the stool metagenome, neither (**A**) interval adenomas nor (**B**) baseline adenomas were predicted significantly better than chance. (**C** and **D**) The 20 functional metagenomic features (KOs) given the highest importance scores by the RF are shown, sorted by differential abundance that were signed by Gini index, along with their abundances in subjects with and without interval adenoma development. Note that this does not exclude overall microbial community configuration, nor individual microbial features, from being significantly associated with adenoma development as discussed in the text, since significance of association is a much lower bar than accuracy of discriminative prediction.

Supplementary tables

Table S1, related to Figure 1, demographic and clinical characteristics of Lynch cohort.

Table S1, demographic and clinical characteristics of Lynch cohort	
General demographic (n=100)	
Collection site, *n* (%)	
MSKCC	54 (54%)
MGH	46 (46%)
Gender, *n* (%)	
Male	44 (44%)
Female	56 (56%)
Age	
Mean ± SD	47.9±14.5
Race, *n* (%)	
White	89 (89%)
Non-White	11 (11%)
BMI (kg/m2)	
Mean ± SD	
Biopsy location, *n* (%)	
Left colon	98 (52%)
Right colon	72 (39%)
Rectum	2 (1%)
Transverse	15 (8%)
Aspirin, *n* (%)	
Yes	34 (34%)
No	66 (66%)
Antibiotics, *n* (%)	
Yes	17 (17%)
No	66 (66%)
Mutation type, *n* (%)	
MLH1	30 (30%)
MSH2	32 (32%)
MSH6	23 (23%)
PMS2	13 (13%)
Unknown (a diagnosis of Lynch syndrome)	2 (2%)
Cancer history (CRC history), *n* (%)	
Yes	41 (41%)
No	59 (59%)
Clinical outcomes, *n* (%)	
Baseline adenoma, n (%)	
Yes	33 (33%)
No	67 (67%)
Interval adenoma, n (%)	
Yes	28 (28%)
No	61 (61%)

17 and 11 subjects did not provide data on antibiotic use or interval adenoma, respectively.

Table S8, related to Fig.7, S6 and S7: Gut microbial species, functional potential and transcriptional activity are weak predictor of adenoma development.

Prediction models	Features/predictors	AUC
Baseline adenoma	Species	0.6
	Species + Age	0.7
	Species + Age + Bristol scale	0.69
Interval adenoma	Species	0.61
	Species + Age	0.6
	Species + Age + Bristol scale	0.62
Baseline adenoma	DNA kos	0.65
	DNA kos + Age	0.66
	DNA kos + Age + Bristol scale	0.66
Interval adenoma	DNA kos	0.66
	DNA kos + Age	0.66
	DNA kos + Age + Bristol scale	0.66
Baseline adenoma	RNA/DNA ratio	0.6
	RNA/DNA ratio + Age	0.63
	RNA/DNA ratio + Age + Bristol scale	0.63
Interval adenoma	RNA/DNA	0.72
	RNA/DNA + Age	0.74
	RNA/DNA + Age + Bristol scale	0.72