Supplementary Table S1: PubMed results of the meta-analysis search and reasons for inclusion/exclusion of studies.
As search term, we used "(Colorectal Cancer) AND (Metagenome) AND (fecal) AND (human)" and collected results up to the 21.06.2018.

Authors	Title	Year	Journal	doi	Exclusion Reason
Liang et al	Fecal Bacteria Act as Novel Biomarkers for Noninvasive Diagnosis of Colorectal Cancer.	2017	Clinical Cancer Research	10.1158/1078-0432.CCR-16-1599	No shotgun metagenomics
Beghini et al	Large-scale comparative metagenomics of Blastocystis, a common member of the human gut microbiome.	2017	The ISME Journal	10.1038/ismej.2017.139	Unrelated
Butto et al	Functional relevance of microbiome signatures: The correlation era requires tools for consolidation.	2017	The Journal of Allergy and Clinical Immunology	10.1016/j.jaci.2017.02.010	Review
Zhou et al	Cancer killers in the human gut microbiota: diverse phylogeny and broad spectra.	2017	Oncotarget	10.18632/oncotarget.17319	Unrelated
Tsoi et al	Peptostreptococcus anaerobius Induces Intracellular Cholesterol Biosynthesis in Colon Cells to Induce Proliferation and Causes Dysplasia in Mice.	2017	Gastroenterology	10.1053/j.gastro.2017.01.009	No shotgun metagenomics
Dubinikina et al	Links of gut microbiota composition with alcohol dependence syndrome and alcoholic liver disease.	2017	Microbiome	10.1186/s40168-017-0359-2	No CRC samples
Yu et al	Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer.	2017	Gut	10.1136/gutjnl-2015-309800	
Wong et al	Gavage of Fecal Samples From Patients With Colorectal Cancer Promotes Intestinal Carcinogenesis in Germ-Free and Conventional Mice.	2017	Gastroenterology	10.1053/j.gastro.2017.08.022	No human samples
Vogtmann et al	Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing.	2016	PLoS one	10.1371/journal.pone.0155362	
Estaki et al	Cardiorespiratory fitness as a predictor of intestinal microbial diversity and distinct metagenomic functions.	2016	Microbiome	10.1186/s40168-016-0189-7	No CRC samples
Hester et al	Fecal microbes, short chain fatty acids, and colorectal cancer across racial/ethnic groups.	2015	World Journal of Gastroenterology	10.3748/wjg.v21.i9.2759	No shotgun metagenomics
Feng et al	Gut microbiome development along the colorectal adenoma-carcinoma sequence.	2015	Nature Communications	10.1038/ncomms7528	
Zeller et al	Potential of fecal microbiota for early-stage detection of colorectal cancer.	2014	Molecular Systems Biology	10.15252/msb.20145645	

Duboc et al	Connecting dysbiosis, bile-acid dysmetabolism and gut inflammation in inflammatory bowel diseases.	2013	Gut	10.1136/gutjnl-2012-302578	No CRC samples
Couturier-Maillard et al	NOD2-mediated dysbiosis predisposes mice to transmissible colitis and colorectal cancer.	2013	The Journal of Clinical Investigation	10.1172/JCI62236	No CRC samples
Shanahan, F.	The colonic microbiota in health and disease.	2013	Current Opinion in Gastroenterology	10.1097/MOG.0b013e32835a3493	Review
Chen et al	Decreased dietary fiber intake and structural alteration of gut microbiota in patients with advanced colorectal adenoma.	2013	The American Journal of Clinical Nutrition	10.3945/ajcn.112.046607	No shotgun metagenomics
Ou et al	Diet, microbiota, and microbial metabolites in colon cancer risk in rural Africans and African Americans.	2013	The American Journal of Clinical Nutrition	10.3945/ajcn.112.056689	No shotgun metagenomics
Holma et al	Colonic methane production modifies gastrointestinal toxicity associated with adjuvant 5-fluorouracil chemotherapy for colorectal cancer.	2013	Journal of Clinical Gastroenterology	10.1097/MCG.0b013e3182680201	No CRC samples
Brim et al	Microbiome analysis of stool samples from African Americans with colon polyps.	2013	PLoS one	10.1371/journal.pone.0081352	No shotgun metagenomics
Kelly, CP	Aspects of large intestinal health and disease.	2013	Current Opinion in Gastroenterology	10.1097/MOG.0b013e32835af1a6	Review
Walker et al	Therapeutic modulation of intestinal dysbiosis	2013	Pharmacological Research	10.1016/j.phrs.2012.09.008	Review
Castellarin et al	Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma.	2012	Genome Research	10.1101/gr.126516.111	No shotgun metagenomics
Araujo-Perez et al	Differences in microbial signatures between rectal mucosal biopsies and rectal swabs.	2012	Gut Microbes	10.4161/gmic.22157	No shotgun metagenomics
Odamaki et al	Effect of the oral intake of yogurt containing Bifidobacterium longum BB536 on the cell numbers of enterotoxigenic Bacteroides fragilis in microbiota.	2012	Anaerobe	10.1016/j.anaerobe.2011.11.004	No shotgun metagenomics
Wang et al	Structural segregation of gut microbiota between colorectal cancer patients and healthy volunteers.	2012	The ISME Journal	10.1038/ismej.2011.109	No shotgun metagenomics
Thomas et al	New insights into the impact of the intestinal microbiota on health and disease: a symposium report.	2012	The British Journal of Nutrition	10.1017/S0007114511006970	Review
Zella et al	Distinct microbiome in pouchitis compared to healthy pouches in ulcerative colitis and familial adenomatous polyposis.	2011	Inflammatory bowel disease	10.1002/ibd.21460	No CRC samples
Sobhani et al	Microbial dysbiosis in colorectal cancer (CRC) patients.	2011	PLoS one	10.1371/journal.pone.0016393	No shotgun metagenomics

Supplementary Table S2: Extended subject information for the eight included metagenomics studies.
Statistical associations between meta-variables and disease status were assessed using Fisher's Exact test for sex and Wilcoxon tests for BMI and age. Healthy controls in the DE study did not undergo colonoscopy.

		FR		AT		CN		US		DE		Independent validation cohorts	IT1		IT2		JP	
		CTR	CRC	CTR	CRC	CTR	CRC	CTR	CRC	CTR	CRC		CTR	CRC	CTR	CRC	CTR	CRC
N		61	53	63	46	54	74	52	52	60	60		24	29	28	32	40	40
Sex	Female	33	24	26	18	21	26	15	15	28	24		11	6	12	9	16	19
	Male	28	29	37	28	33	48	37	37	32	36		13	23	16	23	24	21
	p-value	0.45		0.85		0.71		1		0.58			0.11		0.28		0.63	
BMI	Mean (SD)	24.7 (3.2)	25.6 (3.7)	27.6 (3.7)	26.5 (3.5)	23.5 (3.0)	24.0 (3.2)	25.3 (4.3)	24.9 (4.2)	24.9 (3.2)	26.2 (4.0)		25.6 (3.7)	25.7 (4.1)	24.6 (3.8)	26.8 (4.3)	23.1 (2.4)	22.4 (2.7)
	p-value	0.49		0.10		0.11		0.63		0.04			0.84				0.26	
Age	Mean (SD)	60.6 (11.4)	66.8 (10.9)	67.1 (6.3)	67.1 (10.9)	61.8 (5.7)	66.0 (10.6)	61.2 (11.0)	61.8 (13.6)	57.6 (11.1)	63.5 (12.6)		67.9 (7.0)	71.4 (8.2)	57.8 (8.3)	58.4 (8.4)	63.2 (12.2)	59.0 (12.8)
	p-value	0.01		0.63		0.01		0.72		0.01			0.13				0.09	
CRC Stage	0/I		15		24		18		0		18							
	II		7		9		22		12		20							
	III		10		9		26		21		9							40 (stage III-IV)
	IV		21		3		8		18		13							
Localisation	Left Colon		27		11		26		19		16			10				
	Right Colon		17		8		11		15		15			9				
	Rectum		9		27		28		13		26			9				
Sampling relative to Colonoscopy		All before		All before		30 before 24 after	21 before 52 after	All after		none	after							
Original Publication		Zeller et al. *MSB* 2014 [1]		Feng et al. *Nat Commun* 2015 [2]		Yu et al. *Gut* 2017 [3]		Vogtmann et al. *PLoS One* 2016 [4]		This study			Thomas et al. [5]				Courtesy of T. Yamada	

Supplementary Table S5: Metabolomic studies performed in colorectal cancer tissue and fecal matter support the enrichment of gut metabolic modules (GMMs) in colorectal cancer metagenomes

For the CRC-associated gut metabolic modules (two-sided blocked Wilcoxon, n=574, see Fig. **4a**), differential abundance of the input molecules was checked in four metabolomics studies. Three studies (Hirayama et al. 2009 [6], Denkert et al. 2008 [7], Mal et al. 2012 [8]) performed experiments using CRC and healthy control tissue, whereas one study (Weir et al. 2013 [9]) assessed metabolites in fecal matter of CRC patients and healthy controls. An upward arrow (↑) indicates that the metabolite was found to be significantly enriched in CRC tissue or fecal matter of CRC patients and a downward arrow (↓) a significant depletion of the metabolite. Gut metabolic modules are ordered by significance and direction of change, corresponding to Fig. **4a**.

	GMM	Name	Input	Output	Experiments in CRC tissue			Experiments in fecal matter
					Hirayama et al.	Denkert et al.	Mal et al.	Weir et al.
Modules down-regulated in CRC metagenomes	MF0015	fructose degradation	[fructose 1-phosphate]	[fructose 1,6-bisphosphate]		fructose 1-phosphate ↓		
	MF0102	sulfate reduction (dissimilatory)	[sulfate, H2,]	[hydrogen sulfide]				
	MF0078	lactaldehyde degradation	[lactaldehyde]	[propane-1,2-diol]				
	MF0002	fructan degradation	[fructan]	[fructose]				
	MF0017	galactose degradation	[galactose]	[glucose 6-phosphate]		galactose ↓	galactose ↓	
	MF0058	lysine degradation II	[lysine]	[CO2, cadaverine]				
	MF0092	lactate production	[pyruvate]	[lactate]	pyruvate ↓			
	MF0047	glutamine degradation II	[glutamine, 2-oxoglutarate]	[glutamate]		glutamine ↓		
	MF0090	ethanol production I	[pyruvate]	[formate, ethanol]	pyruvate ↓			
	MF0010	sucrose degradation I	[sucrose]	[fructose 6-phosphate, glucose 6-phosphate]				
	MF0001	arabinoxylan degradation	[arabinoxylan]	[arabinose, xylose]				
Modules up-regulated in CRC metagenomes	MF0016	fucose degradation	[fucose]	[lactaldehyde, glyceraldehyde 3-phosphate]				
	MF0028	aspartate degradation I	[aspartate, 2-oxoglutarate]	[glutamate, oxaloacetate]	aspartate ↑		aspartate ↑	aspartate ↑
	MF0044	cysteine degradation I	[cysteine]	[NH3, H2S, pyruvate]	cysteine ↑	cysteine ↑		
	MF0072	pyruvate dehydrogenase complex	[pyruvate]	[CO2, acetyl-CoA]				
	MF0027	tyrosine degradation II	[tyrosine]	[phenol, pyruvate, NH3]	tyrosine ↑			
	MF0101	nitrate reduction (dissimilatory)	[nitrate]	[NH3]				
	MF0045	cysteine degradation II	[cysteine, 2-oxoglutarate, hydrogen cyanide]	[pyruvate, glutamate, thiocyanate]	cysteine ↑	cysteine ↑		

	MF0075	acetate to acetyl-CoA	[acetate]	[acetyl-CoA]				acetic acid ↑
	MF0034	alanine degradation II	[2-oxoglutarate, alanine]	[glutamate, pyruvate]	alanine ↑	alanine ↑	alanine ↑	alanine ↑
Modules up-regulated in CRC metagenomes	MF0096	succinate production	[phosphoenolpyruvate, CO2]	[succinate]				
	MF0057	lysine degradation I	[lysine, acetyl-CoA]	[NH3, butanoate, acetoacetyl-CoA]	lysine ↑	lysine ↑		lysine ↑
	MF0040	proline degradation	[proline]	[glutamate]	proline ↑	proline ↑	proline ↑	proline ↑
	MF0031	glutamate degradation II	[glutamate]	[CO2, 4-aminobutanoate]	glutamate ↑	glutamate ↑		glutamate ↑
	MF0103	mucin degradation	[mucin]	[NA]				
	MF0095	propionate production III	[pyruvate, succinyl-CoA]	[propanoyl-CoA, oxaloacetate]				
	MF0088	butyrate production I	[crotonyl-CoA]	[butanoate]				
	MF0083	succinate consumption	[succinate, acetyl-CoA]	[crotonyl-CoA, acetate]	succinate ↑		succinate ↑	
	MF0046	glutamine degradation I	[glutamine]	[glutamate, NH3]				
	MF0032	glutamate degradation III	[glutamate]	[NH3, pyruvate, acetate]	glutamate ↑	glutamate ↑		glutamate ↑
	MF0056	histidine degradation	[histidine]	[NH3, formate, glutamate]	histidine ↑			
	MF0035	glycine degradation	[glycine, tetrahydrofolate]	[CO2, NH3, 5,10-methylenetetrahydrofolate]	glycine ↑	glycine ↑	glycine ↑	glycine ↑
	MF0025	tryptophan degradation	[tryptophan]	[indole, NH3, pyruvate]	tryptophan ↑		Picolinic acid ↑	

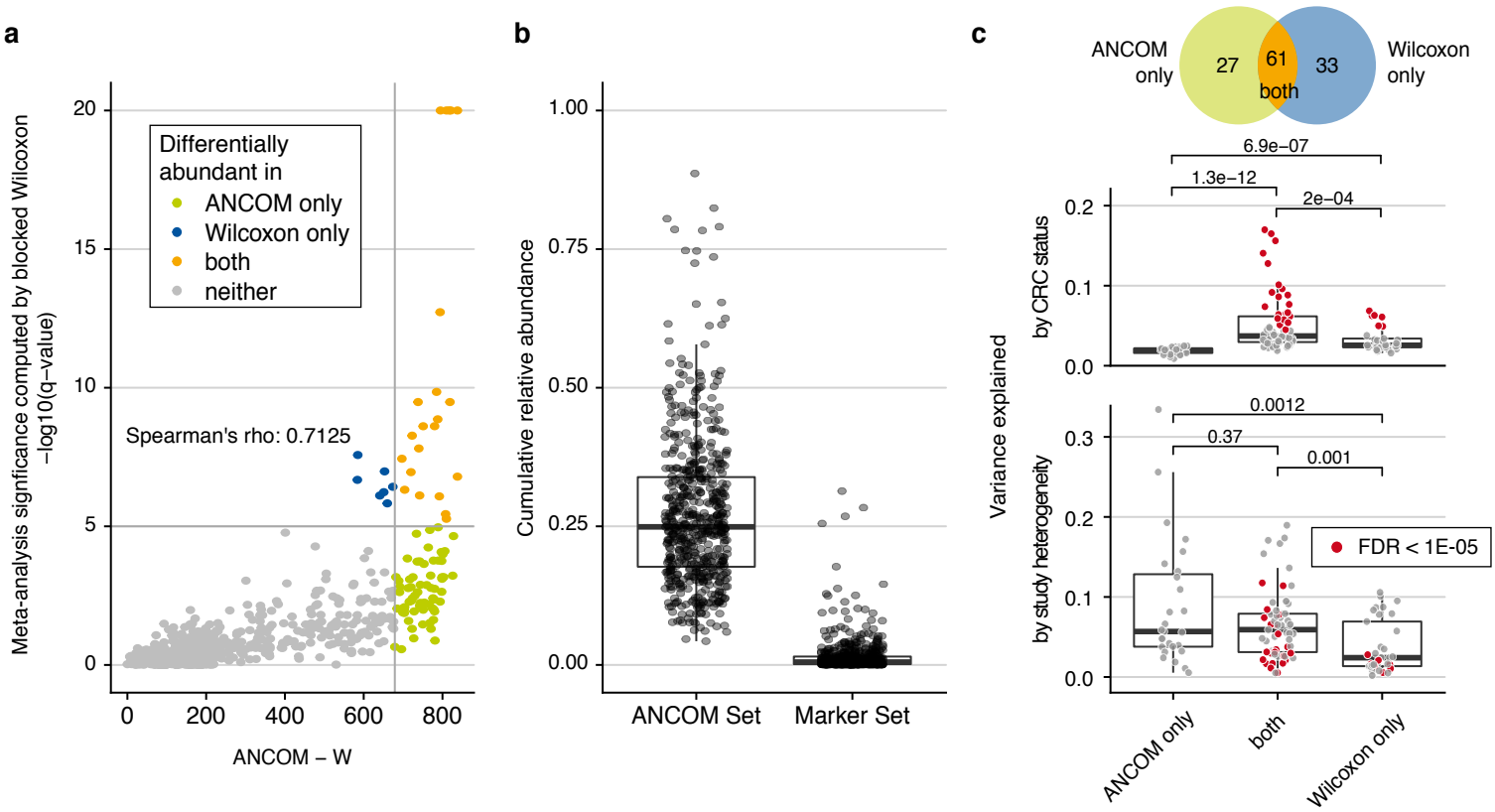

Figure S1: Results from blocked Wilcoxon test are unlikely to be confounded by compositionality
(a) CRC enrichment was tested by confounder-aware ANCOM (Mandal et al. *Microb Ecol Health Dis* 2015 [10]) and the two-sided blocked Wilcoxon test (n=574 independent observations, see Methods). As ANCOM does not output a p-value but rather an effect size called W, the values for W are plotted against the q-values determined by the blocked Wilcoxon for all gut microbial species. In order to call a species differentially abundant between the two groups (CRC vs control), ANCOM determines a decision boundary on the W values. Dots are color-coded according to whether the Wilcoxon test returned significant results or ANCOM's W was higher than the respective decision boundary. **(b)** The plots shows for the set of species called as differentially abundant by ANCOM (ANCOM set) and the set of marker species (FDR < 1E-05, see **Fig. 1c**) the cumulative relative abundance in all samples (n=575). In the set of marker species, the cumulate relative abundance is below 3% in 90% of the samples, indicating that these are unlikely to have changed solely by compositional effects. **(c)** The size of the intersection and set differences between the ANCOM set and the extended set of marker species (FDR < 0.005, see **Fig. 1ab**) are shown in the Venn diagram. For the species in each group, the boxplots below show the variance explained by CRC status and by study heterogeneity (see Methods, **Fig. 1d**). The species uniquely identified by ANCOM seem to be more confounded by study heterogeneity. P-values were computed by two-sided Wilcoxon test. In all boxplots, boxes denote interquartile ranges (IQR) with the median as thick black line and whiskers extending up to the most extreme points within 1.5-fold IQR.

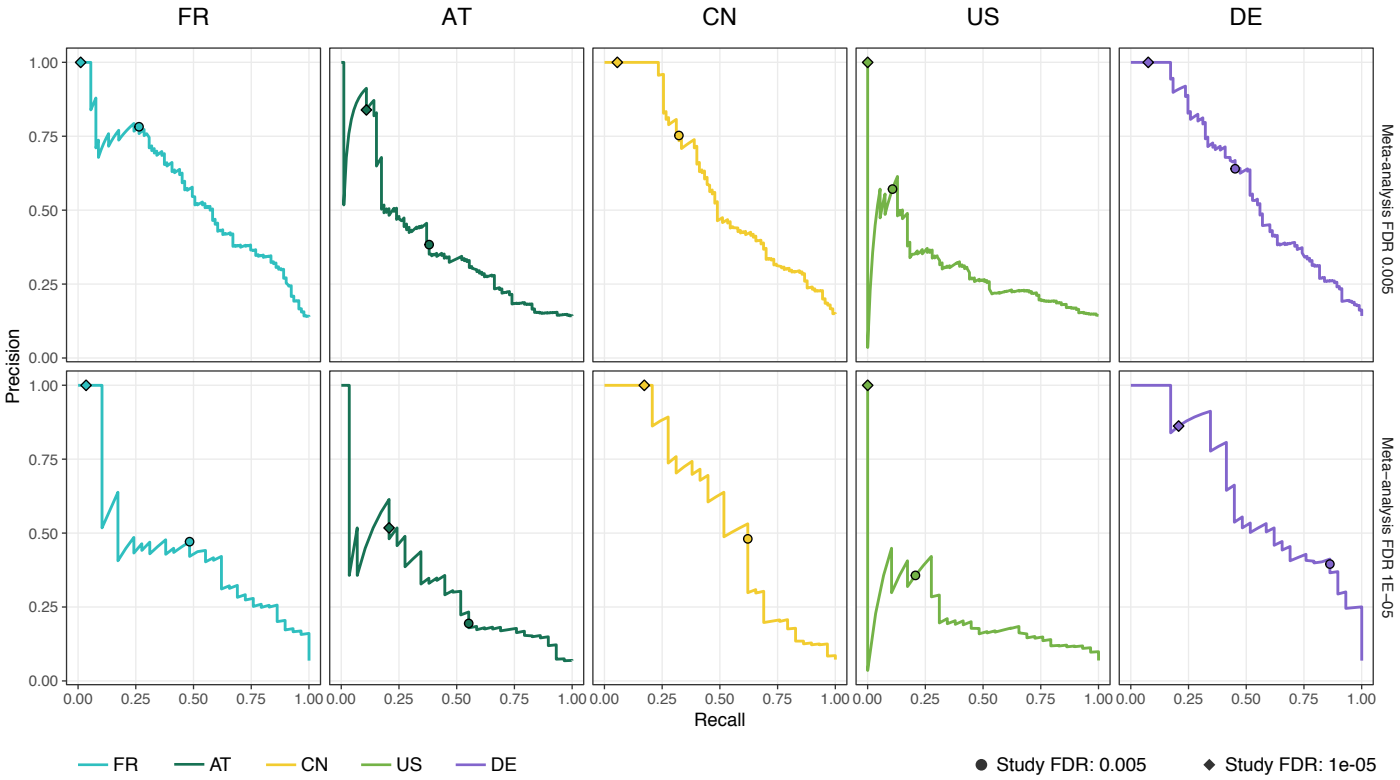

Figure S2: Retrospective analysis of study sensitivity and precision for detection of CRC-associated microbial species identified in meta-analysis.
Precision-recall plots for the different studies using the meta-analysis set of associated species at FDR 0.005 (n=94, top row) and 1E-05 (n=29, bottom row) as "true" set (tested by two-sided blocked Wilcoxon test, see Methods) and the naive (uncorrected) within-study significance (tested by two-sided Wilcoxon test) as predictor. Circles and diamonds indicate precision and recall of within-study significance cutoffs of 0.005 and 1E-05, respectively.

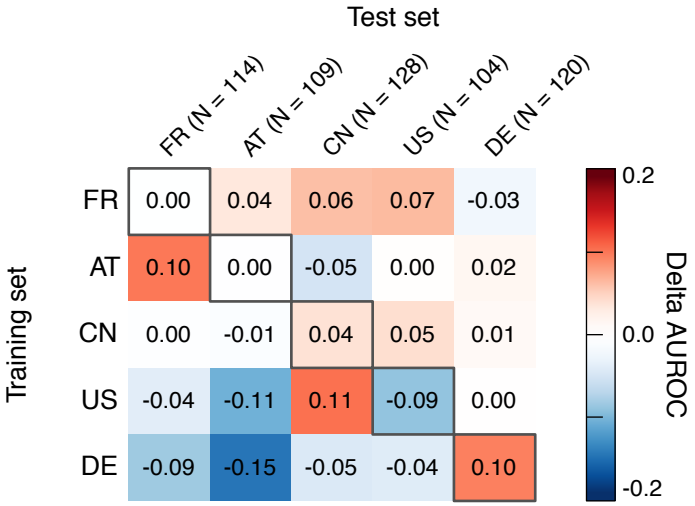

Figure S3: Performance differences between LASSO and Random Forest classification models. Classification accuracy as measured by AUROC was compared between LASSO models (see **Fig. 3a**) and Random Forest models trained on taxonomic abundance profiles. Positive Delta AUROC values (red) indicate better performance for the Random Forest classifier, whereas negative values (blue) represent better performance of the LASSO models.

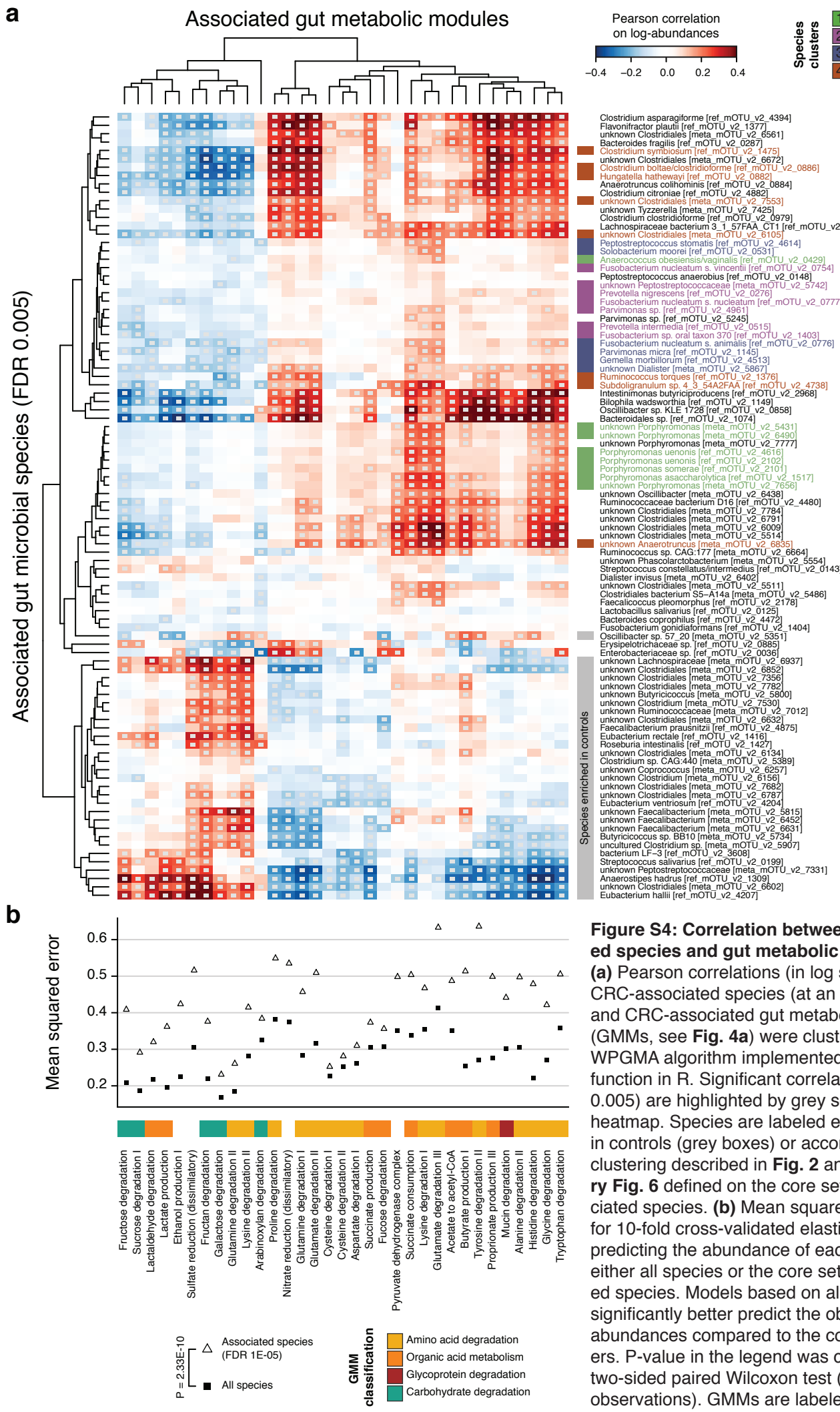

Figure S4: Correlation between CRC-associated species and gut metabolic modules. **(a)** Pearson correlations (in log space) between CRC-associated species (at an FDR of 0.005) and CRC-associated gut metabolic modules (GMMs, see **Fig. 4a**) were clustered using the WPGMA algorithm implemented in the *hclust* function in R. Significant correlations (p-value < 0.005) are highlighted by grey squares inside the heatmap. Species are labeled either as enriched in controls (grey boxes) or according to the clustering described in **Fig. 2** and **Supplementary Fig. 6** defined on the core set of 29 CRC-associated species. **(b)** Mean squared error is shown for 10-fold cross-validated elastic net models predicting the abundance of each GMM based on either all species or the core set of CRC-associated species. Models based on all species can significantly better predict the observed GMM abundances compared to the core set of markers. P-value in the legend was computed by two-sided paired Wilcoxon test (n=66 paired observations). GMMs are labeled with the higher-level classification from **Fig. 4a**.

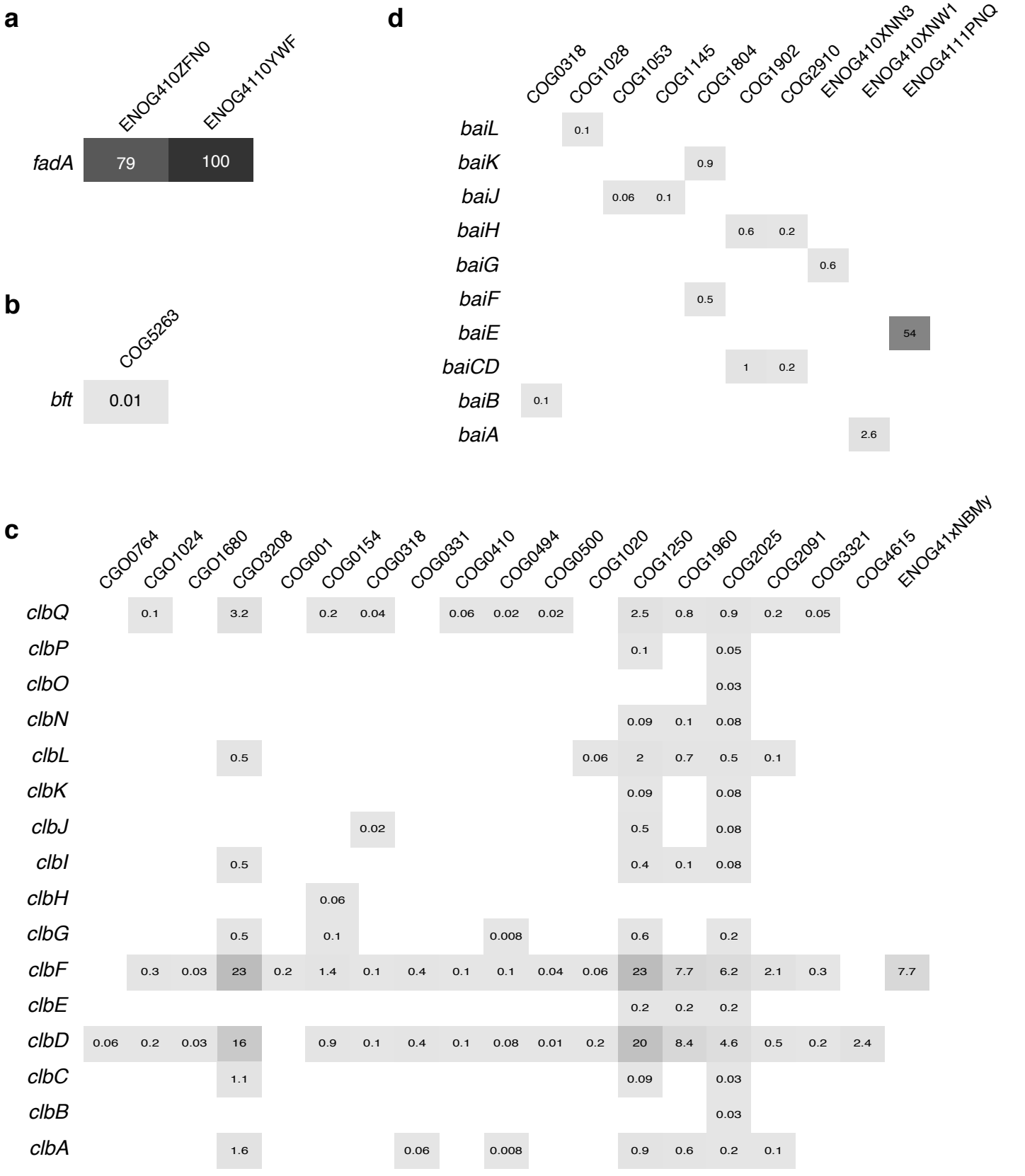

Figure S5: Coverage of HMM hits in the eggNOG database.
For each HMM used to annotate potential virulence factors and toxins (see Methods and **Fig. 4c**), the resulting matches in the IGC were compared to how many genes in the IGC are annotated with overlapping eggNOG orthologous groups. The values in the heatmap represent the percentage of all genes in the IGC annotated with the respective eggNOG orthologous group that were captured by the HMM. Since *fadA* **(a)** and *bft* **(b)** are single-gene factors, a single HMM was trained in both cases, whereas multiple different HMMs were trained for the individual gene functions of *pks* **(c)** and the *bai operon* **(d)**. With the exception of *fadA* and the *baiE* gene, the eggNOG orthologous groups are generally much broader and encompass a magnitude more genes in the IGC than the HMM hits.

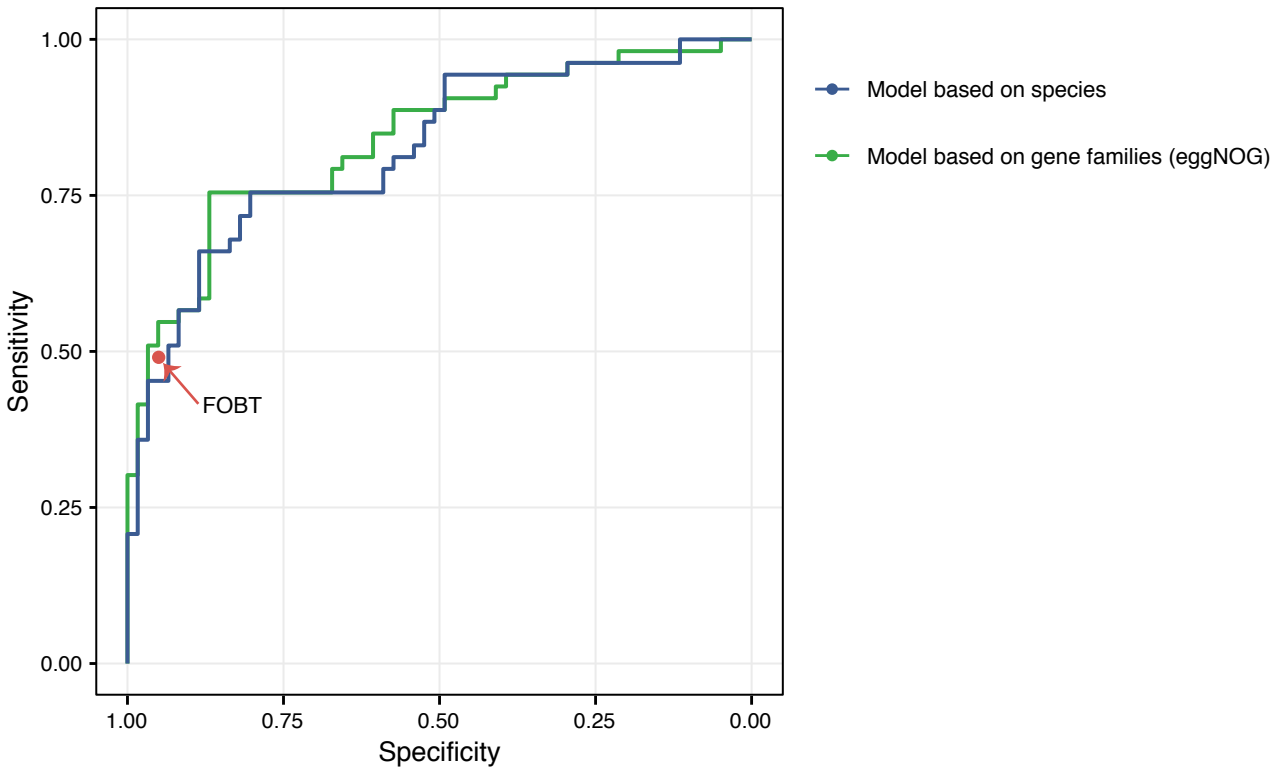

Figure S6: Models based on taxonomic or functional profiles detect CRC as accurately as the fecal occult blood test (FOBT)
Receiver operating characteristics for the leave-one-study-out (LOSO) models based on species abundances and eggNOG gene families for the FR study (the only dataset for which FOBT results are available) are shown in blue and green, respectively. The accuracy of the FOBT test on the same set of patients is indicated by the red dot.

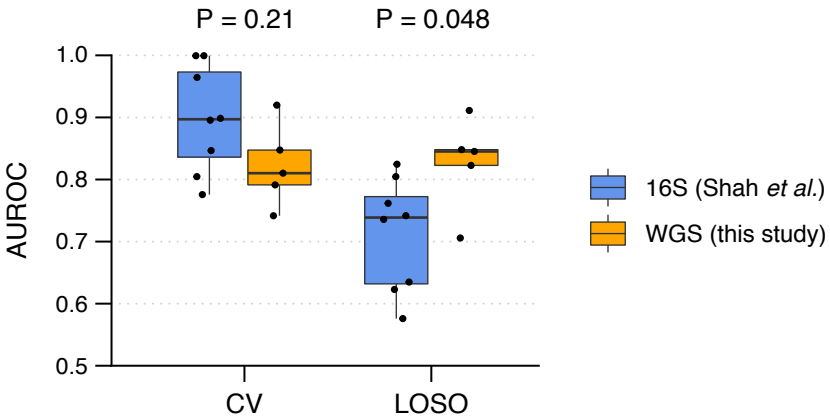

Figure S7: Comparison of classification accuracies in meta-analyses based on16S and whole metagenomic sequencing (WGS)
Model performance as measured by AUROC is compared for cross-validation (CV) and leave-one-study-out (LOSO) models between a meta-analysis based on whole metagenomic sequencing (this study) and another one based on 16S data (see Shah et al., Gut 2018 [11]). For CV, each dot represents the performance of a model trained on a single metagenomic study (included in the respective meta-analysis) in cross-validation, whereas each dot for LOSO shows how well a model trained on all but one study performs on the left-out study. AUROC values for WGS are identifical to **Fig. 3ab**. Model performances for 16S were extraced out of the Supplementary Tables 8 and 9 from Shah et al. Gut 2018. Significance was assessed using a two-sided Wilcoxon test (n=5 studies for WGS and n=9 studies for 16S). In all boxplots, boxes denote interquartile ranges (IQR) with the median as thick black line and whiskers extending up to the most extreme points within 1.5-fold IQR.

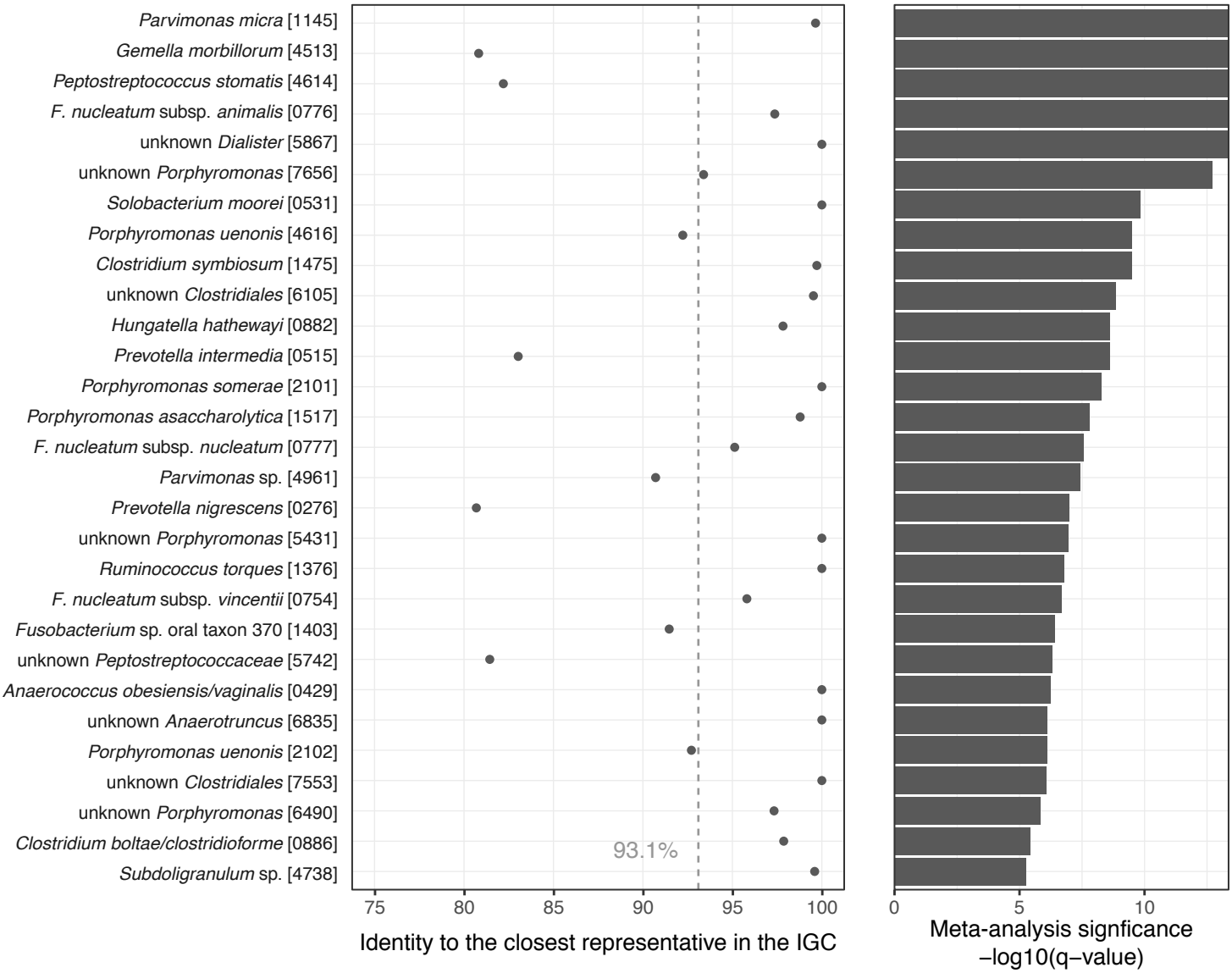

Figure S8: Matches for COG0533 of the CRC-associated mOTUs in the integrated gene catalogue (IGC)
For the core set of CRC-associated marker species (see **Fig. 1**), dots denote the sequence identity of the COG0533 gene used to define the respective mOTU (see Mende et al. *Nature Methods* 2013 [12], Sunagawa et al. *Nature Methods* 2013 [13], Milanese et al. *Nature Communications* 2019 [14], and motu-tool.org) to the closest representative in the integrated gene catalogue (IGC). The dashed line indicates a cutoff of 93.1%, which was determined as species cutoff for this gene (see Supplementary Table 4 in Sunagawa et al. *Nature Methods* 2013). The barplot oin the right represents the meta-analysis significance of the CRC enrichment for the respective species (two-sided blocked Wilcoxon test, n=574 independent samples; see **Fig. 1a**).
All species (rows) to the right of, or close to the dashed line are represented by sequences in the IGC that are likely to originate from the same species (or genus). In contrast, for the remaining five CRC-associated marker species distant from the dashed line, close relatives could not be detected in the IGC indicating that their gene pool (and functional repertoire) is potentially not captured by our functional metagenome annotation workflow due to its reliance on the IGC.

Supplementary References

1. Zeller, G., et al., *Potential of fecal microbiota for early-stage detection of colorectal cancer.* Mol. Syst. Biol., 2014. **10**(11): p. 766.
2. Feng, Q., et al., *Gut microbiome development along the colorectal adenoma-carcinoma sequence.* Nat. Commun., 2015. **6**: p. 6528.
3. Yu, J., et al., *Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer.* Gut, 2017. **66**(1): p. 70-78.
4. Vogtmann, E., et al., *Colorectal Cancer and the Human Gut Microbiome: Reproducibility with Whole-Genome Shotgun Sequencing.* PLoS One, 2016. **11**(5): p. e0155362.
5. Thomas, A.M., et al., *Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation.* co-submitted to Nature Medicine, 2018.
6. Hirayama, A., et al., *Quantitative metabolome profiling of colon and stomach cancer microenvironment by capillary electrophoresis time-of-flight mass spectrometry.* Cancer Res, 2009. **69**(11): p. 4918-25.
7. Denkert, C., et al., *Metabolite profiling of human colon carcinoma--deregulation of TCA cycle and amino acid turnover.* Mol Cancer, 2008. **7**: p. 72.
8. Mal, M., et al., *Metabotyping of human colorectal cancer using two-dimensional gas chromatography mass spectrometry.* Anal Bioanal Chem, 2012. **403**(2): p. 483-93.
9. Weir, T.L., et al., *Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults.* PLoS One, 2013. **8**(8): p. e70803.
10. Mandal, S., et al., *Analysis of composition of microbiomes: a novel method for studying microbial composition.* Microb Ecol Health Dis, 2015. **26**: p. 27663.
11. Mende, D.R., et al., *proGenomes: a resource for consistent functional and taxonomic annotations of prokaryotic genomes.* Nucleic Acids Res., 2017. **45**(D1): p. D529-D534.
12. Shah, M.S., et al., *Leveraging sequence-based faecal microbial community survey data to identify a composite biomarker for colorectal cancer.* Gut, 2018. **67**(5): p. 882-891.
13. Mende, D.R., et al., *Accurate and universal delineation of prokaryotic species.* Nat. Methods, 2013. **10**(9): p. 881-884.
14. Sunagawa, S., et al., *Metagenomic species profiling using universal phylogenetic marker genes.* Nat. Methods, 2013. **10**(12): p. 1196-1199.
15. Milanese, A., et al., *Microbial abundance, activity, and population genomic profiling with mOTUs.* Nature Communications, 2019. **formally accepted for publication**.