nature medicine

ARTICLES

https://doi.org/10.1038/s41591-019-0406-6

Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer

Jakob Wirbel[1,31], Paul Theodor Pyl[2,3,31], Ece Kartal[1,4], Konrad Zych[1], Alireza Kashani[2], Alessio Milanese[1], Jonas S. Fleck[1], Anita Y. Voigt[1,5], Albert Palleja[2], Ruby Ponnudurai[1], Shinichi Sunagawa[1,6], Luis Pedro Coelho[1,30], Petra Schrotz-King[7], Emily Vogtmann[8], Nina Habermann[9], Emma Niméus[3,10], Andrew M. Thomas[11,12], Paolo Manghi[11], Sara Gandini[13], Davide Serrano[13], Sayaka Mizutani[14,15], Hirotsugu Shiroma[14], Satoshi Shiba[16], Tatsuhiro Shibata[16,17], Shinichi Yachida[16,18], Takuji Yamada[14,19], Levi Waldron[20,21], Alessio Naccarati[22,23], Nicola Segata[11], Rashmi Sinha[8], Cornelia M. Ulrich[24], Hermann Brenner[7,25,26], Manimozhiyan Arumugam[2,27,32]*, Peer Bork[1,4,28,29,32]* and Georg Zeller[1,32]*

Association studies have linked microbiome alterations with many human diseases. However, they have not always reported consistent results, thereby necessitating cross-study comparisons. Here, a meta-analysis of eight geographically and technically diverse fecal shotgun metagenomic studies of colorectal cancer (CRC, $n = 768$), which was controlled for several confounders, identified a core set of 29 species significantly enriched in CRC metagenomes (false discovery rate (FDR) $< 1 \times 10^{-5}$). CRC signatures derived from single studies maintained their accuracy in other studies. By training on multiple studies, we improved detection accuracy and disease specificity for CRC. Functional analysis of CRC metagenomes revealed enriched protein and mucin catabolism genes and depleted carbohydrate degradation genes. Moreover, we inferred elevated production of secondary bile acids from CRC metagenomes, suggesting a metabolic link between cancer-associated gut microbes and a fat- and meat-rich diet. Through extensive validations, this meta-analysis firmly establishes globally generalizable, predictive taxonomic and functional microbiome CRC signatures as a basis for future diagnostics.

Metagenomic sequencing technologies have enabled the study of microbial communities that colonize the human body in a culture-independent manner[1]. They have yielded glimpses into the complex, yet incompletely understood, interactions between the gut microbiome—the microbial ecosystem residing primarily in the large intestine—and its host[2]. To explore microbiome–host interactions within a disease context, metagenome-wide association studies (MWAS) have begun to map gut microbiome alterations in diabetes, inflammatory bowel disease, CRC, and many other conditions[3–12]. However, due to the many

[1]Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. [2]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medicine, University of Copenhagen, Copenhagen, Denmark. [3]Division of Surgery, Oncology and Pathology, Department of Clinical Sciences Lund, Faculty of Medicine, Lund University, Lund, Sweden. [4]Molecular Medicine Partnership Unit, Heidelberg, Germany. [5]The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA. [6]Department of Biology, ETH Zürich, Zürich, Switzerland. [7]Division of Preventive Oncology, National Center for Tumor Diseases and German Cancer Research Center, Heidelberg, Germany. [8]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. [9]Genome Biology Unit, European Molecular Biology Laboratory, Heidelberg, Germany. [10]Division of Surgery, Department of Clinical Sciences Lund, Faculty of Medicine, Skane University Hospital, Lund, Sweden. [11]Department CIBIO, University of Trento, Trento, Italy. [12]Biochemistry Department, Chemistry Institute, University of São Paulo, São Paulo, Brazil. [13]IEO, European Institute of Oncology IRCCS, Milan, Italy. [14]School of Life Science and Technology, Tokyo Institute of Technology, Tokyo, Japan. [15]Research Fellow of Japan Society for the Promotion of Science, Tokyo, Japan. [16]Division of Cancer Genomics, National Cancer Center Research Institute, Tokyo, Japan. [17]Laboratory of Molecular Medicine, Human Genome Center, The Institute of Medical Science, The University of Tokyo, Tokyo, Japan. [18]Department of Cancer Genome Informatics, Graduate School of Medicine/Faculty of Medicine, Osaka University, Osaka, Japan. [19]PRESTO, Japan Science and Technology Agency, Saitama, Japan. [20]Graduate School of Public Health and Health Policy, City University of New York, New York, NY, USA. [21]Institute for Implementation Science in Population Health, City University of New York, New York, NY, USA. [22]Italian Institute for Genomic Medicine, Turin, Italy. [23]Department of Molecular Biology of Cancer, Institute of Experimental Medicine, Prague, Czech Republic. [24]Huntsman Cancer Institute and Department of Population Health Sciences, University of Utah, Salt Lake City, UT, USA. [25]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center, Heidelberg, Germany. [26]German Cancer Consortium, German Cancer Research Center, Heidelberg, Germany. [27]Faculty of Healthy Sciences, University of Southern Denmark, Odense, Denmark. [28]Max Delbrück Centre for Molecular Medicine, Berlin, Germany. [29]Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany. [30]Present address: Institute of Science and Technology for Brain-Inspired Intelligence, Fudan University, Shanghai, China. [31]These authors contributed equally: Jakob Wirbel, Paul Theodor Pyl. [32]These authors jointly supervised this work: Manimozhiyan Arumugam, Peer Bork, Georg Zeller. *e-mail: arumugam@sund.ku.dk; bork@embl.de; zeller@embl.de

biological factors that may influence gut microbiome composition in addition to the condition studied, a current challenge for MWAS are confounders, which can cause false associations[13,14]. This issue is further aggravated by a lack of standards in metagenomic data generation and processing, making it difficult to disentangle technical from biological effects[15].

The robustness of microbiome disease associations can be assessed through comparisons across multiple metagenomic case-control studies, that is, meta-analyses. The aim of meta-analyses is to identify associations that are consistent across studies and thus less likely to be attributable to biological or technical confounders. Most informative are meta-analyses of populations from diverse geographic and cultural regions. Previous microbiome meta-analyses based on 16S ribosomal RNA (rRNA) gene amplicon data found stark technical differences between studies; the reported taxonomic disease associations were either of low effect size or not well resolved[16–18]. In contrast, shotgun metagenomics have enabled analyses with higher taxonomic resolution as well as analyses of gene functions, which have improved the statistical power needed to fine-map disease-associated strains and aid in the interpretation of host-microbial co-metabolism. However, thus far, the meta-analyses of shotgun metagenomic data have either reported on the features of general dysbiosis in comparisons across multiple diseases[19], or have left it unclear how well microbiome signatures generalize across studies of the same disease when data are rigorously separated to avoid overoptimistic evaluations of their prediction accuracy[20].

In this study, we present a meta-analysis of eight studies of CRC, including fecal metagenomic data from 386 cancer cases and 392 tumor-free controls (CTRLs). After consistent data reprocessing, we examined an initial set of five studies for CRC-associated changes in the gut microbiome. First, we investigated potential confounders; then, we identified (univariate) microbial species associations, and inferred species co-occurrence patterns in CRC. Second, we trained multivariable classification models to recognize CRC status, from both taxonomic and functional microbiome profiles, and tested how accurately these models generalized to data from studies not used for training. Moreover, we evaluated the performance improvements achieved by pooling data across studies and the disease specificity of the resulting classification models. Third, the targeted investigation of virulence and toxicity genes as candidate functional biomarkers for CRC revealed several of these to be enriched in CRC metagenomes, which is indicative of their prevalence and potential relevance in CRC patients. Three additional, more recent studies were finally used to independently validate these taxonomic and functional CRC signatures.

Results

Consistent processing of published and new data for the meta-analysis of CRC metagenomes. In this meta-analysis, we included four published studies that used fecal shotgun metagenomics to characterize CRC patients compared to healthy CTRLs (see Table 1, Supplementary Table 1, and Methods for the inclusion criteria). For an additional fifth study population, we generated new fecal metagenomic data from samples collected in Germany; a subset of samples from this patient collective were published previously (see Table 1 and Methods[8]). These five studies were conducted on three continents and differed in sampling procedures, sample storage, and DNA extraction protocols. Notably, the fecal specimens of the United States study were freeze-dried and stored at −80 °C for more than 25 years before DNA extraction and sequencing[10]. However, in all studies, samples were collected before treatment, thus excluding cancer therapy as a potential confounding effect[14,21]. Most samples were taken before bowel preparation for colonoscopy, with some exceptions in the Germany, China, and United States studies (Supplementary Table 2). To ensure consistency in the bioinformatic analyses, all raw sequencing data were reprocessed using the mOTUs2 tool for taxonomic profiling[22] and MOCAT2 (metagenomic analysis toolkit) for functional profiling[23].

Table 1 | Fecal metagenomic studies of CRC included in this meta-analysis

Country code	Reference	No. of cases	No. of controls
France	Zeller et al.[8]	53	61
Austria	Feng et al.[9]	46	63
China	Yu et al.[11]	74	54
United States	Vogtmann et al.[10]	52	52
Germany	The current study	60	60
External validation cohorts			
Italy 1	Thomas et al.[27]	29	24
Italy 2	Thomas et al.[27]	32	28
Japan	Courtesy of T. Yamada et al.	40	40

See the Methods for the inclusion criteria and Supplementary Table 2 for the extended metadata. For a detailed description of patient recruitment and data generation for the German study, see the Methods. The data for 38 samples from the German study has been published previously as part of an independent validation cohort in Zeller et al.[8].

Univariate meta-analysis of species associated with CRC. The first aim of the meta-analysis was to determine the gut microbial species that are enriched or depleted in CRC metagenomes in a consistent manner across the five study populations. However, since these studies differed from one another in many biological and technical aspects, we first quantified the effect of study-associated heterogeneity on microbiome composition. We contrasted this with other potential confounders (patient age, body mass index (BMI), sex, sampling after colonoscopy, and library size; additionally, smoking status, type 2 diabetes comorbidity, and vegetarian diet where available; Extended Data Fig. 1 and Supplementary Table 3). This analysis revealed the factor ‘study’ to have a predominant impact on species composition, which is supported by a recent comparison of DNA extraction protocols, since these typically differ between studies[15]. An analysis of microbial alpha and beta diversity showed that study heterogeneity also had a larger effect on overall microbiome composition than CRC in our data (Extended Data Fig. 2).

Parametric effect size measures are not well established for the identification of microbial taxa significantly differing in abundance in CRC because microbiome data is characterized by non-Gaussian distributions with extreme dispersion; thus, we used a generalization of the fold change (Extended Data Fig. 3) and non-parametric significance testing. In this permutation test framework[24] (herein referred to as blocked (univariate) Wilcoxon tests), differential abundance in CRC can be assessed while accounting for ‘study’ as a confounding effect that is treated as a blocking factor; additionally, motivated by our confounder analysis, we also blocked for ‘colonoscopy’ in all analyses (Methods and Extended Data Fig. 1). To rule out spurious associations due to the compositional nature of microbial relative abundance data, we additionally compared the results of this test with a method[25] that employs log-ratio transformation and found highly correlated results (Supplementary Fig. 1 and Supplementary Table 4).

At a meta-analysis FDR of 0.005, we identified 94 microbial species to be differentially abundant in the CRC microbiome out of 849 species consistently detected across studies (Supplementary Table 4 and Methods). Among these, we focused on a core set of the 29 most significant markers (FDR $< 1 \times 10^{-5}$; Fig. 1a) for further analysis. The latter included members of several genera previously associated

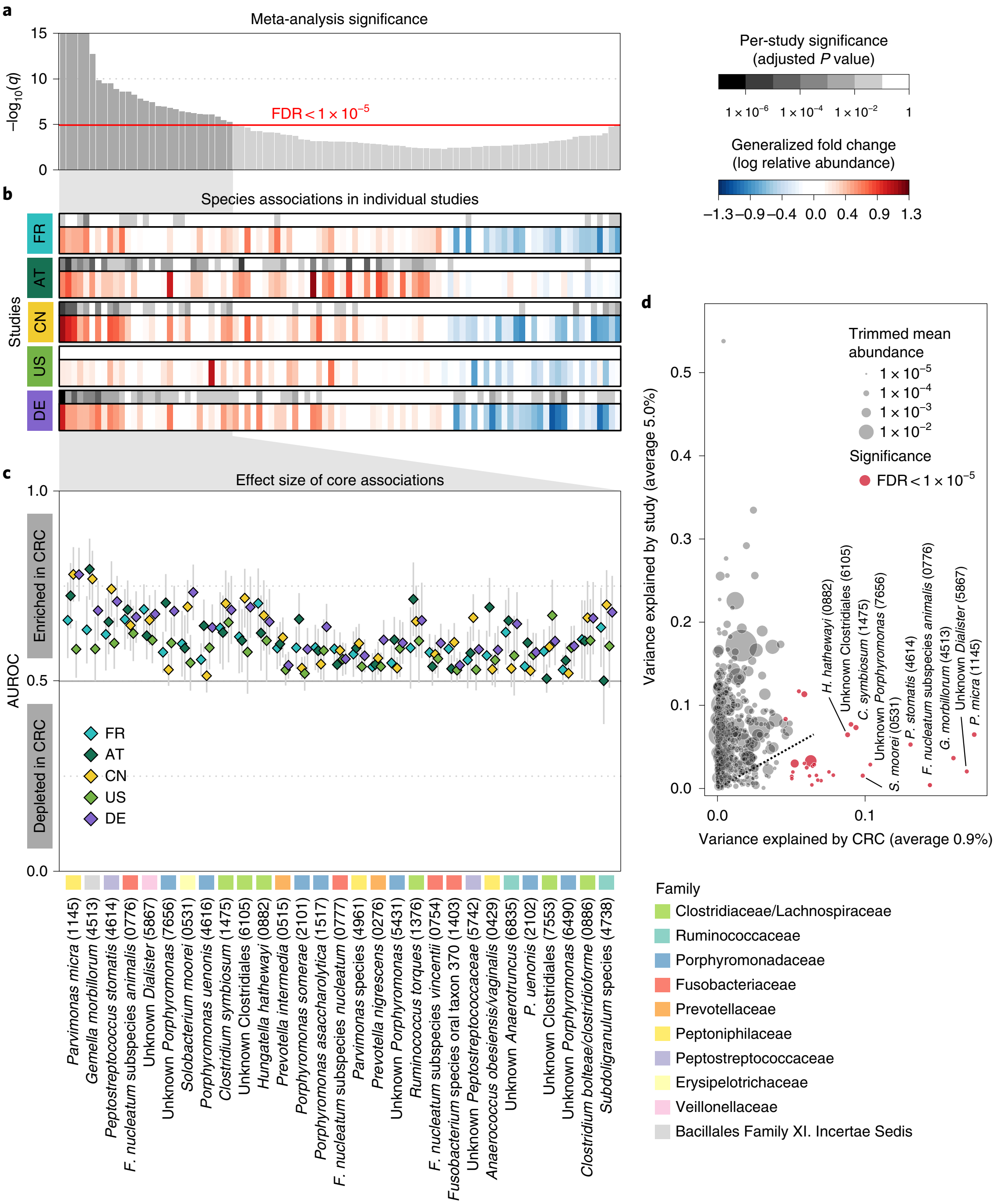


Fig. 1 | Despite study differences, meta-analysis identifies a core set of gut microbes strongly associated with CRC. **a**, The meta-analysis significance of gut microbial species derived from blocked Wilcoxon tests ($n = 574$ independent observations) is given by the bar height (FDR = 0.005). **b**, Underneath, species-level significance, as calculated with a two-sided Wilcoxon test (FDR-corrected P value), and the generalized fold change (Methods) within individual studies are displayed as heatmaps in gray and in color, respectively (see color bars and Table 1 for details on the studies included). Species are ordered by meta-analysis significance and direction of change. AT, Austria; CN, China; DE, Germany; FR, France; US, United States. **c**, For a core of highly significant species (meta-analysis FDR = 1×10^{-5}), association strength is quantified by the AUROC across individual studies (color-coded diamonds), and the 95% confidence intervals are indicated by the gray lines. Family-level taxonomic information is color-coded above the species names (the numbers in brackets are mOTUs2 species identifiers; see Methods). **d**, Variance explained by disease status (CRC versus CTRLs) is plotted against variance explained by study effects for individual microbial species with dot size being proportional to abundance (see Methods); core microbial markers are highlighted in red.

with CRC, such as *Fusobacterium*, *Porphyromonas*, *Parvimonas*, *Peptostreptococcus*, *Gemella*, *Prevotella*, and *Solobacterium* (Fig. 1b)[8–11], and 8 additional species without genomic reference sequences (meta-mOTUs; Milanese et al.[22]; see Methods) mostly from the *Porphyromonas* and *Dialister* genera and the Clostridiales order (see Extended Data Fig. 4 and Supplementary Table 4 for genus-level associations). Collectively, these 29 core CRC-associated species show a previously underappreciated diversity of 11 Clostridiales species to be enriched in CRC (Fig. 1b). In contrast to the majority of species that are more strongly affected by study heterogeneity than by CRC status, 26 out of the 29 CRC-associated species varied more according to disease status (Fig. 1d).

All of the core CRC-associated species were enriched in patients and were often undetectable in metagenomes from non-neoplastic CTRLs. While previous studies were contradictory in the reported proportion of positive versus negative associations[8,9,17,20], our meta-analysis results are more easily reconciled with a model in which—potentially many—gut microbes contribute to or benefit from tumorigenesis than with the opposing model where a lack of protective microbes contributes to CRC development (Fig. 1c). Although these core taxonomic CRC associations were highly significant and consistent, individual studies showed marked discrepancies in the species identified as significant (Fig. 1b). Retrospective examination of the precision and sensitivity with which individual studies detected this core of CRC-associated species showed relatively low sensitivity for the United States study (consistent with the original report[10]) and low precision of the Austrian study due to associations that were not replicated in other studies (Supplementary Fig. 2).

Analyzing patient metagenomes for co-occurrences among the core set of 29 species that are strongly enriched in the CRC microbiome revealed four species clusters with distinct taxonomic composition (Fig. 2a and Extended Data Fig. 5; Methods). Two of them showed strong taxonomic consistency: cluster 1 exclusively comprised *Porphyromonas* species and cluster 4 only contained members of the Clostridiales order. In contrast, the other two clusters were taxonomically more heterogeneous, with cluster 3 grouping together the species with the highest prevalence in CRC cases (all among the ten most highly significant markers), consistent with a co-occurrence analysis of one of the data sets included here[11]. Cluster 2 contained species with intermediate prevalence.

Investigating whether these four clusters were associated with different tumor characteristics, we found the *Porphyromonas* cluster 1 to be significantly enriched in rectal tumors (Fig. 2b), consistent with the presence of superoxide dismutase genes in *Porphyromonas* genomes possibly conferring tolerance to a more aerobic milieu in the rectum (Extended Data Fig. 5). The Clostridiales cluster 4 was significantly more prevalent in female CRC patients. All species clusters showed a slight tendency toward late-stage CRC (that is, American Joint Committee on Cancer stages 3 and 4), but this was only significant for cluster 3. Associations with patient age and BMI were weaker and not significant (Extended Data Fig. 5). To rule out secondary effects due to differences in patient characteristics among studies, all of these tests were corrected for study effects (by blocking for ‘study’ and ‘colonoscopy’; see Methods). At the level of individual species, significant stage-specific enrichments could not be detected, suggesting CRC-associated microbiome changes to be less dynamic during cancer progression than previously postulated[26], although fecal material may be less suitable to address this question than tissue samples.

Metagenomic CRC classification models. To establish metagenomic signatures for CRC detection across studies in the face of geographic and technical heterogeneity, we developed multivariable statistical modeling workflows with rigorous external validation to avoid prevailing issues of overfitting and overoptimistic reports of model accuracy[19]. As a precaution against overoptimistic evaluation, these workflows are independent of the differential abundance analysis described earlier. Instead, least absolute shrinkage and selection operator (LASSO) logistic regression classifiers were employed to select predictive microbial features and eliminated uninformative ones (see Methods).

In a first step, we used abundance profiles from five studies including the 849 most abundant microbial species and assessed how well classifiers trained in cross-validation on one study generalize in evaluations on the other four studies (study-to-study transfer of classifiers; Fig. 3a). Within-study cross-validation performance, as quantified by the area under the receiver operating characteristics curve (AUROC), ranged between 0.69 and 0.92 and was generally maintained in study-to-study transfer (AUROC dropping by 0.07 ± 0.12 on average) with two notable exceptions. First, in line with the univariate analysis of species associations, CRC detection accuracy in the United States study was lower than for the other studies, both in cross-validation and in study-to-study transfer. This could potentially be explained by the United States fecal specimens, unlike the other studies, being freeze-archived for > 25 years before metagenomic sequencing[10]. Second, classifiers trained on the Austrian study did not generalize as well to the other studies, consistent with low study precision seen in univariate meta-analysis (Supplementary Fig. 2). Given the microbial co-occurrence clusters described earlier, we wondered whether species–species interactions would provide additional information relevant for CRC recognition that is not contained in the species abundance profiles. However, non-linear classifiers able to exploit such interactions did not yield significantly better accuracies (Supplementary Fig. 3; see also Thomas et al.[27]), suggesting that the linear model based on few biomarkers (on average 17 species accounted for more than 80% of the total classifier weights; Extended Data Fig. 6) is near-optimal for CRC prediction.

We further assessed if including data from all but one study in model training improves prediction on the remaining hold-out study (leave-one-study-out (LOSO) validation). The LOSO performance of species-level models ranged between 0.71 and 0.91; when the United States study was disregarded as an outlier, it was ≥0.83 (Fig. 3b). This corresponds to a LOSO accuracy increase of 0.076 ± 0.03 compared to study-to-study transfer. These results suggest that one can expect a CRC detection accuracy ≥0.8 (AUROC) for any new CRC study using similarly generated metagenomic data. Moreover, we verified that metagenomic CRC classification models trained on species composition were not biased for clinical subgroups. With the exception of slightly more sensitive detection of late-stage CRC ($P = 0.04$, mostly originating from the United States study; Extended Data Fig. 7), we did not observe any classification bias by patient age, sex, BMI, or tumor location. Taken together, this suggests that these metagenomic classifiers are unlikely to be strongly confounded by the clinical parameters recorded.

Several previous studies comparing microbiome changes across multiple diseases reported primarily general dysbiotic alterations and highlighted the need to examine the disease specificity of microbiome signatures[17,19]. Therefore, we assessed the false positive predictions of our metagenomic CRC classifiers on the fecal metagenomes of type 2 diabetes[4,5], Parkinson’s disease[12], ulcerative colitis, and Crohn’s disease[6,7] patients, reasoning that classifiers relying on biomarkers for general dysbiosis would yield an excess of false positives on these cohorts. However, our LOSO classification models calibrated to have a false positive rate (FPR) of 0.1 on CRC data sets in fact maintained similarly low FPRs on other disease data sets ranging from 0.09 to 0.13 (Fig. 3c). Interestingly, the disease specificity of LOSO models was significantly improved over that observed for classifiers trained on a single study, indicating that inclusion of multiple studies in the training set of a classifier can substantially improve its specificity for a given disease.

Functional metagenomic signatures for CRC. Since shotgun metagenomics data, unlike 16S rRNA gene amplicon data, allow for a direct analysis of the functional potential of the gut microbiome, we examined how predictive the metabolic pathways and orthologous gene families differing in abundance between CRC patients and CTRLs would be of CRC status. When applying the same classification workflow as stated earlier to eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) orthologous gene family abundances[28], CRC detection accuracy was very similar to that observed for the taxonomic models (Fig. 3d,e). AUROC

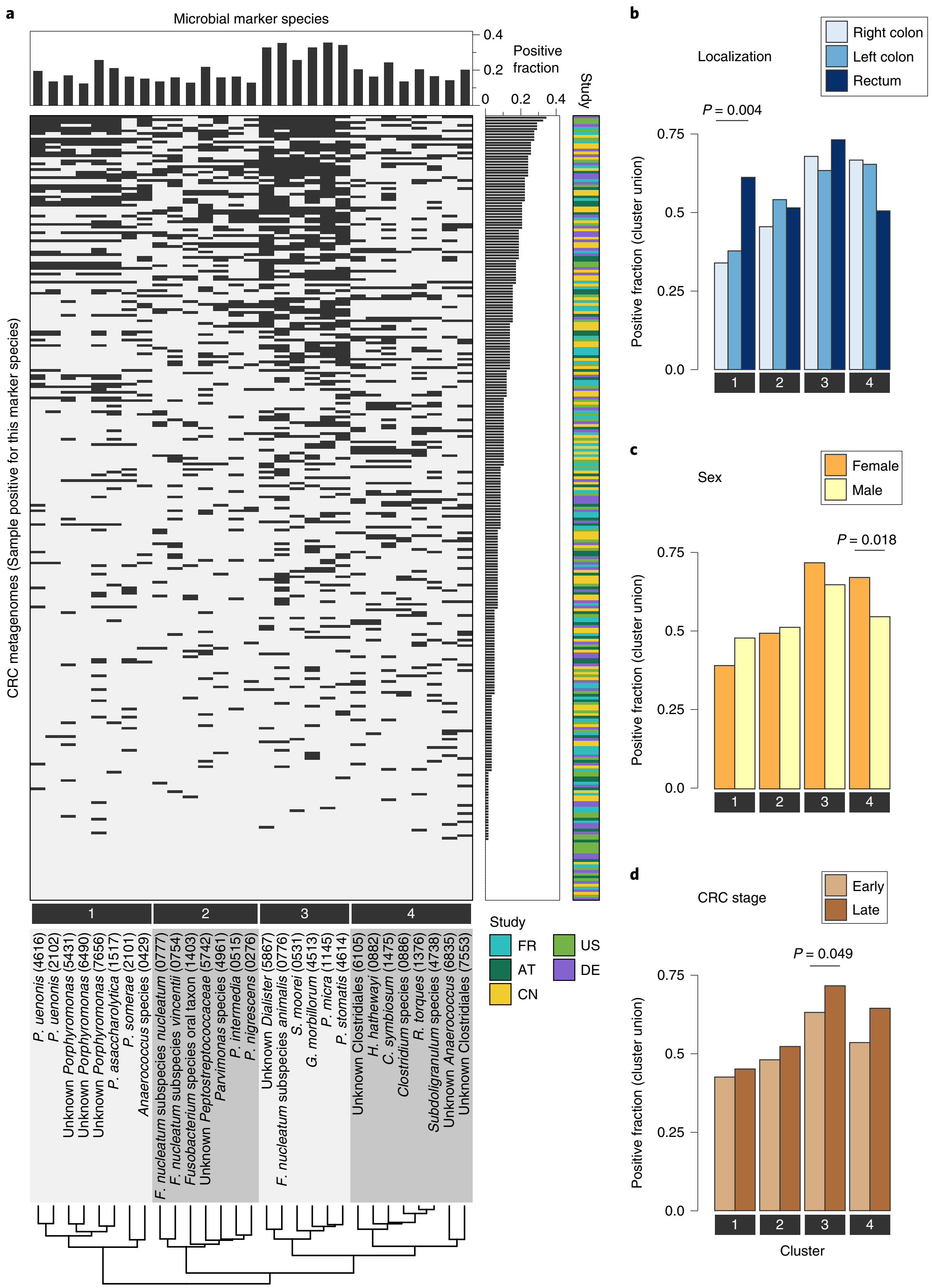


Fig. 2 | Co-occurrence analysis of CRC-associated gut microbial species reveals four clusters preferentially linked to specific patient subgroups. a, For all CRC patients ($n = 285$ independent samples), the heatmap shows whether the respective sample is positive for each of the core set of microbial marker species (see Methods for adjustment of positivity threshold). Samples are ordered according to the sum of positive markers, and marker species are clustered based on the Jaccard index of positive samples, resulting in four clusters (see Methods). **b**–**d**, The barplots in **b**, **c**, and **d** show the fraction of CRC samples that are positive for marker species clusters (defined as the union of positive marker species) broken down by patient subgroups based on differences in tumor location, sex, or CRC stage, respectively. Statistically significant associations between CRC subgroups and marker species clusters were identified using the Cochran–Mantel–Haenszel test blocked for 'study' and 'colonoscopy' effects and are indicated above the bars ($P < 0.1$). Country codes as in Fig. 1b.

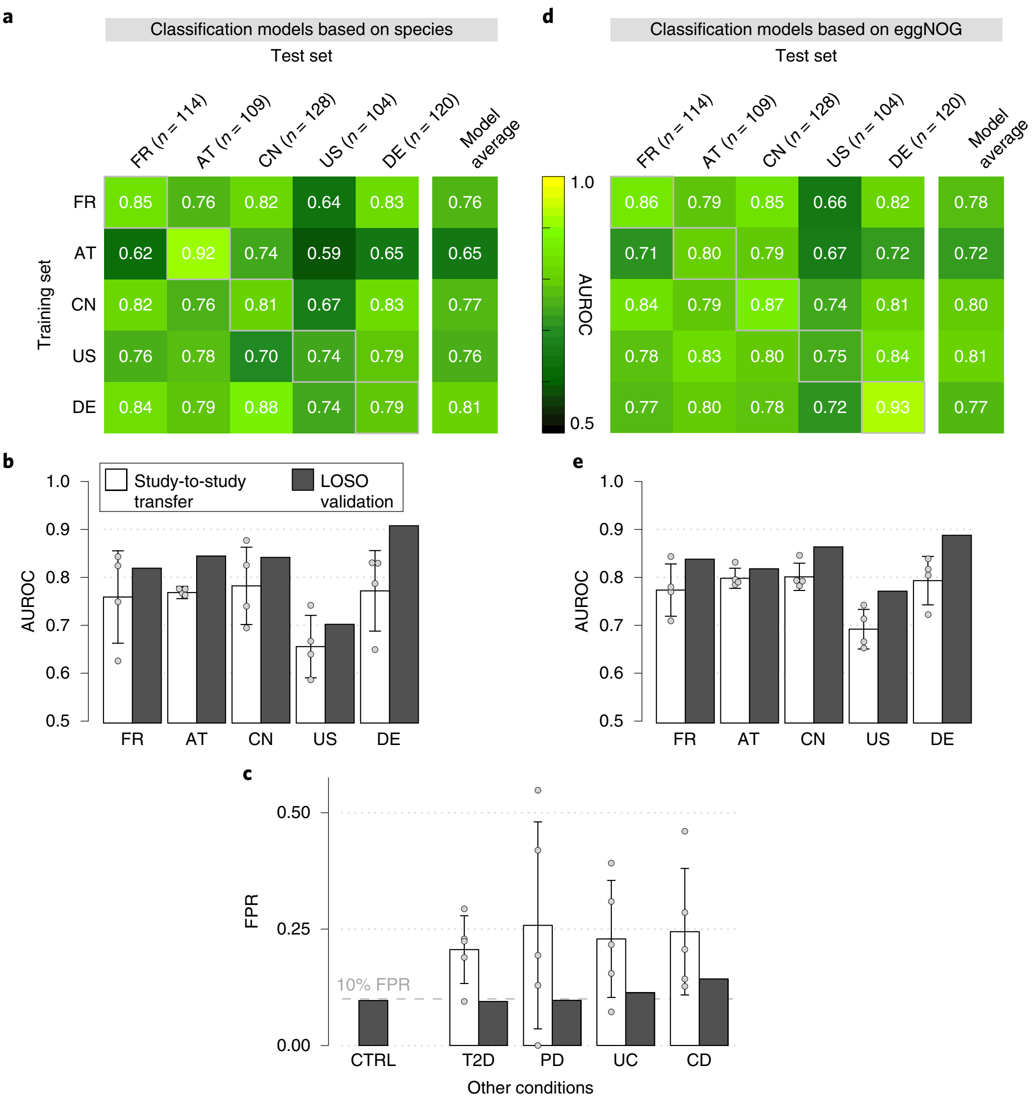


Fig. 3 | Both taxonomic and functional metagenomic classification models generalize across studies, in particular when trained on data from multiple studies. a–e, CRC classification accuracy resulting from cross-validation within each study (gray boxes along the diagonal) and study-to-study model transfer (external validations off-diagonal) as measured by the AUROC for classifiers trained on the species (**a**) and eggNOG gene family (**d**) abundance profiles. The last column depicts the average AUROC across the external validations. Classification accuracy, as evaluated by AUROC on a hold-out study, improves if taxonomic (**b**) or functional (**e**) data from all other studies are combined for training (LOSO validation) relative to models trained on data from a single study (study-to-study transfer, average and s.d. shown by bar height and error bars, respectively, *n* = 4). **c**, Combining training data across studies substantially improves CRC specificity of the (LOSO) classification models relative to models trained on data from a single study (depicted by bar color, as in **c** and **d**) as assessed by the FPR on fecal samples from patients with other conditions (see legend). The bar height for study-to-study transfer corresponds to the average FPR across classifiers (*n* = 5) with the error bars indicating the s.d. of the FPR values observed. T2D, type 2 diabetes; PD, Parkinson's disease; UC, ulcerative colitis; CD, Crohn's disease. Country codes as in Fig. 1b.

values ranged from 0.70 to 0.81 for study-to-study transfer (per-study averages; see Fig. 3e) and from 0.78 to 0.89 in LOSO validation with a pattern of generalization across studies resembling that for taxonomic classifiers. The accuracy of functional signatures did not strongly depend on eggNOG as an annotation source, but was similar when based on other comprehensive functional databases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG)[29] (Extended Data Fig. 8). When using individual gene abundances from metagenomic gene catalogs as a classifier input[30], we observed higher within-study cross-validation AUROC values ≥ 0.96 in all studies, but lower generalization to other studies (AUROC between 0.60 and 0.79) (Extended Data Fig. 8).

To explore changes in the metabolic capacity of gut microbiomes from CRC patients more broadly, we quantified gut metabolic modules (defined in Vieira et al.[31]) and subjected these to the same differential abundance analysis developed for microbial species. Gut metabolic modules with significantly higher abundance in CRC metagenomes (FDR < 0.01, Wilcoxon test blocked for 'study' and 'colonoscopy') predominantly belonged to pathways for the degradation of amino acids, mucins (glycoproteins), and organic acids. This clear trend was accompanied by a depletion of genes from carbohydrate degradation modules (Fig. 4a,b). The differences in all four high-level categories were highly significant ($P < 1 \times 10^{-6}$ in all cases, blocked Wilcoxon tests) and consistent across studies (Fig. 4b). Overall, these results establish a clear shift from dietary carbohydrate utilization in a healthy gut microbiome to amino acid degradation in CRC that is consistent with an earlier report based on a subset of the data[8]. Correlation analysis suggests that increased capacity for amino acid degradation is mostly contributed by CRC-associated Clostridiales (compare with cluster 4 in Fig. 2 and Supplementary Fig. 4). Approximately one half of these metagenomic pathway enrichments are also in agreement with independent metabolomics data, suggesting increased availability of amino acids in the epithelial cells or feces of CRC patients (Supplementary Table 5)[32–36]. While the observed pathway enrichments could potentially result from many factors, including unmeasured ones[13], they are consistent with established dietary risk factors for CRC, which include red and processed meat consumption[37] and low fiber intake[38].

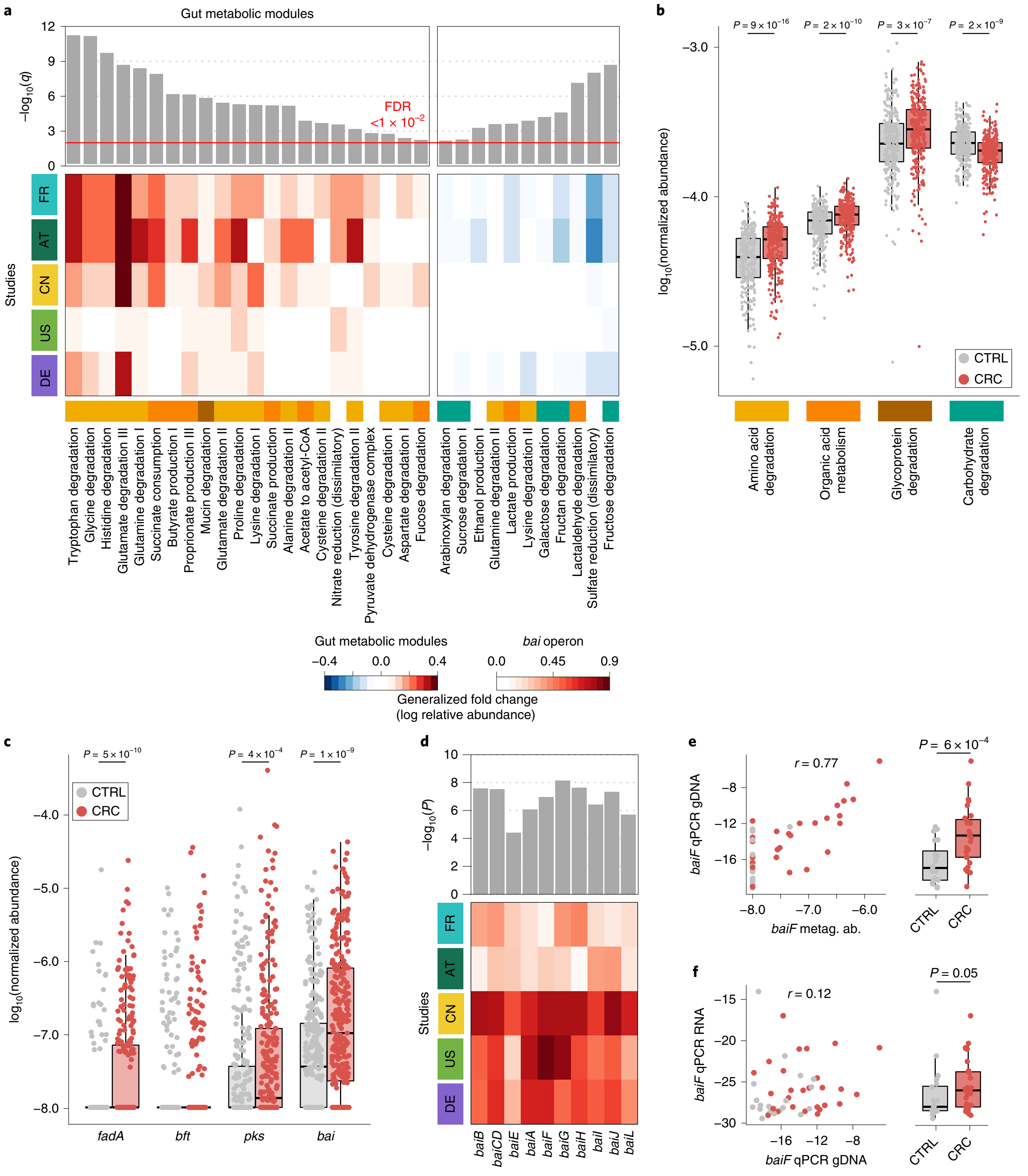
The large metagenomic data set analyzed in this study allowed us to quantify the prevalence of the gut microbial virulence and toxicity mechanisms thought to play a role in colorectal carcinogenesis. Prominent examples include the *Fusobacterium nucleatum* adhesion protein A (encoded by the *fadA* gene), the *Bacteroides fragilis* enterotoxin (*bft* gene) and colibactin produced by some *Escherichia coli* strains (from the *pks* genomic island)[39,40]. Moreover, intestinal *Clostridium* species are known to contribute to the conversion of primary to secondary bile acids using several metabolic pathways including 7α-dehydroxylation, encoded in the *bai* operon[41]. The products of this 7α-dehydroxylation pathway, deoxycholate and lithocholate, are known hepatotoxins associated with liver cancer[42] and hypothesized to also promote CRC[43]. Although intensely studied at a mechanistic level, these factors are not (well)-represented in general databases that can be used for metagenome annotation (Supplementary Fig. 5). Thus, we built a targeted metagenome annotation workflow based on Hidden Markov Models (HMMs) to identify and quantify the virulence factors and toxicity pathways of interest in CRC. Additionally, we used co-abundance clustering to infer operon completeness for factors encoded by multiple genes (see Methods, Extended Data Fig. 9, and Supplementary Fig. 5). While *fadA*, *bft*, the *pks* island, and the *bai* operon were clearly detectable in deeply sequenced fecal metagenomes, they varied broadly with respect to abundance, significance, and cross-study consistency of enrichment (Fig. 4c). *fadA* and *pks* were significantly enriched in CRC metagenomes ($P = 5.3\times10^{-10}$ and 4.1×10^{-4}, respectively), whereas no significant abundance difference could be detected for *bft* in fecal metagenomes, despite reports on its enrichment in the mucosa of CRC patients[44], its carcinogenic effect in mouse models[45], and synergistic action with *pks*[46]. Our quantification of the *bai* operon showed a highly significant enrichment in CRC metagenomes ($P = 1.6\times10^{-9}$) observed across all five studies (Fig. 4d) at an average abundance that exceeded *fadA* and *pks* copy numbers (Fig. 4c). Metagenome analysis indicated that at least four Clostridiales species (including the well characterized *Clostridium scindens* and *Clostridium hylemonae*)[47,48] have a (near)-complete 7α-dehydroxylation pathway contributing to the observed enrichment of *bai* operon copies (Extended Data Fig. 9). To validate this finding and further explore its value toward diagnostic application, we developed a targeted quantification assay for the *baiF* gene based on quantitative PCR (qPCR; see Methods). Quantification of *baiF* by qPCR using genomic DNA (gDNA) from 47 fecal samples of the German study population was found to be similar to, yet more sensitive than by metagenomics (Fig. 4e). Gut microbial *baiF* copy numbers clearly distinguished CRC patients from CTRLs ($P = 0.001$) at an AUROC of 0.77, which in this subset of samples is surpassed by only a single-species marker for CRC (Extended Data Fig. 9). Although consistent with the increased deoxycholate metabolite levels reported for serum and stool samples of CRC patients[49], this finding does not imply 7α-dehydroxylation pathway activity. Therefore, we quantified *baiF* expression using RNA extracts from the same set of fecal samples, and found transcript levels to be elevated in CRC patients also (Fig. 4f). The observed weak correlation of *baiF* expression with genomic abundance (Fig. 4f) might be explained by dynamic transcriptional regulation[47] and therefore *bai* expression in feces might not accurately reflect the tumor environment. Taken together, these data suggest gut microbial metabolic markers to be meaningful and highly predictive of CRC status.

Validation of CRC signatures in independent study populations. Even though CRC classification accuracy for both species and functions were evaluated on independent data, we nonetheless sought to confirm it using two additional study populations from Italy (Italy 1 and Italy 2, combined $N = 61$ CRC, $N = 62$ CTRLs; see Methods and Table 1) and one from Japan ($N = 40$ CRC, $N = 40$ CTRLs; see Methods and Table 1). The overlap of single- species associations detected in the Italy 2 study and those from the meta-analysis was found to vary within the range seen for the other studies, whereas for Italy 1 and Japan, the overlap was slightly lower (compare study precision in Supplementary Fig. 2 and Extended Data Fig. 10). Nonetheless, the AUROC of LOSO classification models based on species ranged between 0.79 and 0.81; that for the classifiers based on eggNOG ranged from 0.71 to 0.92 (Fig. 5a,b). We also validated CRC enrichment of the *fadA*, *pks*, and *bai* genes in these three study populations (Fig. 5c). Altogether, these results highlight consistent alterations in the gut microbiome of CRC patients across eight study populations from seven countries in three continents.

Fig. 4 | Meta-analysis identifies consistent functional changes in CRC metagenomes. a, The meta-analysis significance of gut metabolic modules derived from blocked Wilcoxon tests ($n = 574$ independent samples) is indicated by the bar height (top panel, FDR = 0.01). Underneath, the generalized fold change (see Methods) for gut metabolic modules[31] within individual studies is displayed as a heatmap (see color key in **b**). Metabolic modules are ordered by significance and direction of change. A higher-level classification of the modules is color-coded below the heatmap for the four most common categories (colors as in **b**; white indicates other classes). **b**, Normalized log abundances for these selected functional categories is compared between CTRLs and CRC cases. Abundances are summarized as the geometric mean of all modules in the respective category and statistical significance determined using blocked Wilcoxon tests ($n = 574$ independent samples, see Methods). **c**, Normalized log abundances for virulence factors and toxins compared between metagenomes of CTRLs and CRC cases (significant differences, $P < 0.05$ was determined by blocked Wilcoxon test, $n = 574$ independent samples; see Methods for gene identification and quantification in the metagenomes). *fadA*, gene encoding *F. nucleatum* adhesion protein A; *bft*, gene encoding *B. fragilis* enterotoxin; *pks*, genomic island in *E. coli* encoding enzymes for the production of genotoxic colibactin; *bai*, bile acid-inducible operon present in some Clostridiales species encoding bile acid-converting enzymes. **d**, The meta-analysis significance (uncorrected *P* value), as determined by blocked Wilcoxon tests ($n = 574$ independent samples), and generalized fold change within individual studies are displayed as bars and heatmap, respectively, for the genes contained in the *bai* operon. Due to high sequence similarity to *baiF*, *baiK* was not independently detectable with our approach. **e**, Metagenomic quantification of *baiF* (metagenomic abundance-normalized relative abundance) is plotted against qPCR quantification in gDNA extracted from a subset of German study samples ($n = 47$), with Pearson correlation (r) indicated (see Methods). **f**, Expression of *baiF* determined via qPCR on reverse-transcribed RNA from the same samples in contrast to gDNA (as in **e**). The boxplots on the right of **e** and **f** show the difference between CRC and CTRL samples in the respective qPCR quantification (the *P* values on top were calculated using a one-sided Wilcoxon test). All boxplots show the interquartile ranges (IQRs) as boxes, with the median as a black horizontal line and the whiskers extending up to the most extreme points within 1.5-fold IQR. Country codes as in Fig. 1b.

Discussion

Through extensive and statistically rigorous validation, where data from studies used for training is strictly separated from that for testing, our meta-analysis firmly establishes that gut microbial signatures are highly predictive of CRC (see also Thomas et al.[27]). In particular, metagenomic classifiers trained on species profiles from multiple studies maintained an AUROC of at least 0.8 in seven out of eight data sets and achieved an accuracy similar to the fecal occult blood test, a standard non-invasive clinical test for CRC (Supplementary Fig. 6; see Zeller et al.[8]). Thus, these results suggest that polymicrobial CRC classifiers are globally applicable and can overcome technical and geographical study differences, which we found to generally impact observed microbiome composition more than the disease itself (Fig. 1c and Extended Data Figs. 1 and 2).



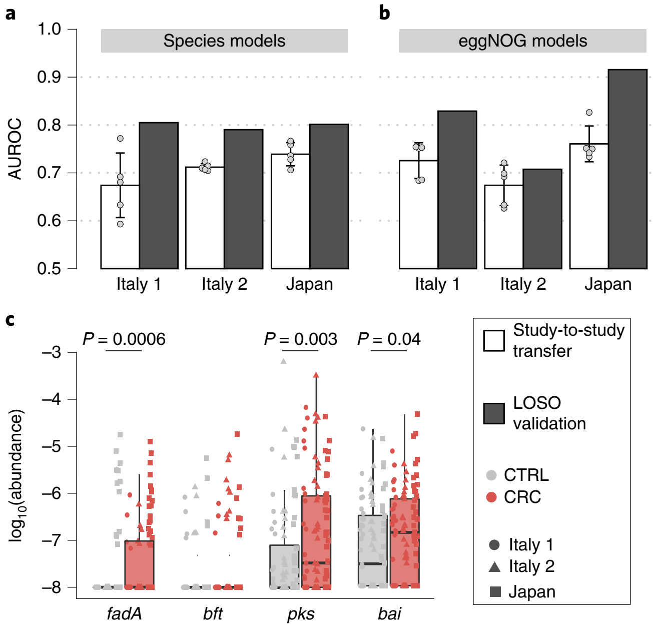


Fig. 5 | Meta-analysis results are validated in three independent study populations. a,**b**, CRC classification accuracy for independent data sets, two from Italy and one from Japan (see Table 1 and Supplementary Table 2), is indicated by the bar height for single-study (white) and LOSO (gray) models using either species (**a**) or eggNOG gene family (**b**) abundance profiles (see Fig. 3). Bar height for single-study models corresponds to the average of five classifiers (the error bars indicate the s.d., $n = 5$). **c**, Normalized log abundances for virulence factors and toxins (see Fig. 4c) compared between CTRLs and CRC cases. *P* values were determined by one-sided blocked Wilcoxon tests ($n = 193$ independent samples). The boxes represent the IQRs with the median as a black horizontal line and the whiskers extending up to the most extreme points within 1.5-fold IQR.

The generalization accuracy of classifiers across studies seen in this study is higher than that reported in 16S rRNA gene amplicon sequencing studies, which are characterized by even larger heterogeneity across studies[16,18] (Supplementary Fig. 7).

Previous microbiome meta-analyses suggested that the majority of gut microbial taxa differing in any given case-control study reflect general dysbiosis rather than disease-specific alterations, thereby illustrating the difficulty of establishing disease-specific microbiome signatures[17,19]. In the current study, by combining data across studies for training (LOSO), we developed disease-specific signatures that maintained false positive control on diabetes and inflammatory bowel disease metagenomes at a very similar level as for CRC (Fig. 3c), despite these diseases having shared effects on the gut microbiome[17,50] and an increased comorbidity risk[51].

Although for diagnostic purposes, unresolved causality between microbial and host processes during CRC development are not a central concern, elucidating the underlying mechanisms would greatly enhance our understanding of colorectal tumorigenesis. Toward this goal, we developed both broad and targeted annotation workflows for functional metagenome analysis. First, we found functional signatures based on the abundances of orthologous groups of microbial genes to yield accuracies as high as taxonomic signatures (Fig. 3), which raises the hope for future improvements in metagenome annotation that can be translated into microbiome signature refinements. Second, by investigating potentially carcinogenic bacterial virulence and toxicity mechanisms using a targeted metagenome annotation approach, we confirmed highly significant enrichments of the colibactin-producing *pks* gene cluster and the *F. nucleatum* adhesin *FadA* in CRC metagenomes (Fig. 4c). Our results support the clinical relevance of these factors and add to the experimental evidence for their carcinogenic potential[46,52–54]. We further examined the *bai* operon, which encodes enzymes that produce secondary bile acids via 7α-dehydroxylation, as an example of toxic host-microbe co-metabolism (see Thomas et al.[27] for another intriguing example). While α-dehydroxylated bile acids are established liver carcinogens[42], their contribution to CRC is less clear[43]. In the current study, we have, for the first time, shown *bai* to be highly enriched in stool from CRC patients (Fig. 4c,d) and confirmed this finding at both the genomic and transcriptomic level using qPCR (Fig. 4e,f). Since *bai* enrichment (and expression) is probably a consequence of a diet rich in fat and meat[55], it is intriguing to explore whether *bai* could be used as a surrogate microbiome marker for such difficult-to-measure dietary CRC risk factors.

To further unravel the molecular underpinning of dietary CRC risk factors, molecular pathological epidemiology studies that investigate the mucosal microbiome as part of the tumor microenvironment hold great potential[56,57]. However, they will require more comprehensive diet questionnaires, medical records, and molecular tumor characterizations than are available for the study populations analyzed in the current study. In this context, carcinogens possibly contained in the virome also warrant further investigation[58,59]; however, for this goal, metagenomic data need to be generated with protocols optimized for virus enrichment[60].

Taken together, our results and those by Thomas et al.[27], strongly support the promise of microbiome-based CRC diagnostics. Both the taxonomic and metabolic gut microbial marker genes established in these meta-analyses could form the basis of future diagnostic assays that are sufficiently robust, sensitive, and cost-effective for clinical application. The targeted qPCR-based quantification of the *baiF* gene is a first step in this direction. Our metagenomic analysis of this and other virulence and toxicity markers bridge to existing mechanistic work in preclinical models and could enable future work that aims to precisely determine the contribution of gut microbiota to CRC development.

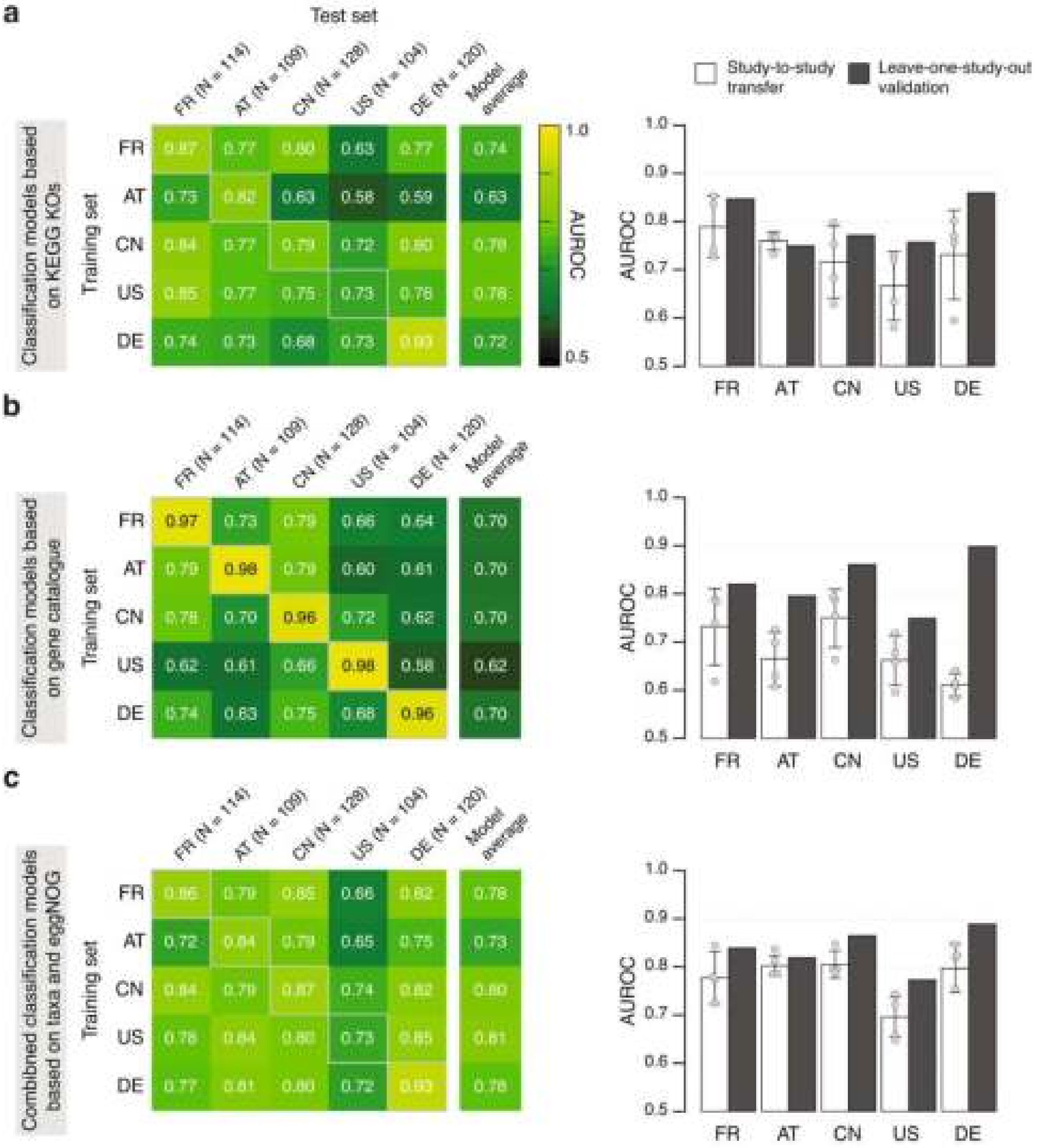
Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41591-019-0406-6.

Received: 30 July 2018; Accepted: 20 February 2019;
Published online: 1 April 2019

References

1. Tringe, S. G. & Rubin, E. M. Metagenomics: DNA sequencing of environmental samples. *Nat. Rev. Genet.* **6**, 805–814 (2005).
2. Tremaroli, V. & Bäckhed, F. Functional interactions between the gut microbiota and host metabolism. *Nature* **489**, 242–249 (2012).
3. Lynch, S. V. & Pedersen, O. The human intestinal microbiome in health and disease. *N. Engl. J. Med.* **375**, 2369–2379 (2016).
4. Qin, J. et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
5. Karlsson, F. H. et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498**, 99–103 (2013).
6. Qin, J. et al. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* **464**, 59–65 (2010).
7. Schirmer, M. et al. Dynamics of metatranscription in the inflammatory bowel disease gut microbiome. *Nat. Microbiol.* **3**, 337–346 (2018).
8. Zeller, G. et al. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol. Syst. Biol.* **10**, 766 (2014).
9. Feng, Q. et al. Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat. Commun.* **6**, 6528 (2015).
10. Vogtmann, E. et al. Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. *PLoS ONE* **11**, e0155362 (2016).
11. Yu, J. et al. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* **66**, 70–78 (2017).



Extended Data Fig. 8 | Cross-study performance of statistical models based on KEGG KO abundances, single-gene abundances from the metagenomic gene catalog (IGC), and the combination of taxonomic and eggNOG database abundance profiles. a–c, CRC classification accuracy resulting from cross-validation within each study (gray boxed along the diagonal) and study-to-study model transfer (external validations off the diagonal) as measured by the AUROC for the classification models trained on KEGG KOs (**a**), models based on the gene catalog (**b**), and models based on the combination of taxonomic and eggNOG database abundance profiles (**c**) (see Methods for the details on the statistical modeling workflows). The last column depicts the average AUROC across external validations. The barplots on the right show that the classification accuracy on a hold-out study improves if the data from all other studies are combined for training (LOSO validation) relative to models trained on data from a single study (study-to-study transfer, indicated by the bar color) consistently across different types of input data. Country codes as in Fig. 1b.