ORIGINAL RESEARCH

Southern Chinese populations harbour non-nucleatum Fusobacteria possessing homologues of the colorectal cancer-associated FadA virulence factor

Yun Kit Yeoh [ID],[1,2] Zigui Chen [ID],[1,2] Martin C S Wong [ID],[1,3] Mamie Hui,[1,2] Jun Yu [ID],[1,4] Siew C Ng [ID],[1,4] Joseph J Y Sung,[4] Francis K L Chan [ID],[1,4] Paul K S Chan[1,2]

ABSTRACT

Objective Fusobacteria are not common nor relatively abundant in non-colorectal cancer (CRC) populations, however, we identified multiple *Fusobacterium* taxa nearly absent in western and rural populations to be comparatively more prevalent and relatively abundant in southern Chinese populations. We investigated whether these represented known or novel lineages in the *Fusobacterium* genus, and assessed their genomes for features implicated in development of cancer.

Methods Prevalence and relative abundances of fusobacterial species were calculated from 3157 CRC and non-CRC gut metagenomes representing 16 populations from various biogeographies. Microbial genomes were assembled and compared with existing reference genomes to assess novel fusobacterial diversity. Phylogenetic distribution of virulence genes implicated in CRC was investigated.

Results Irrespective of CRC disease status, southern Chinese populations harboured increased prevalence (maximum 39% vs 7%) and relative abundances (average 0.4% vs 0.04% of gut community) of multiple recognised and novel fusobacterial taxa phylogenetically distinct from *Fusobacterium nucleatum*. Genomes assembled from southern Chinese gut metagenomes increased existing fusobacterial diversity by 14.3%. Homologues of the FadA adhesin linked to CRC were consistently detected in several monophyletic lineages sister to and inclusive of *F. varium* and *F. ulcerans*, but not *F. mortiferum*. We also detected increased prevalence and relative abundances of *F. varium* in CRC compared with non-CRC cohorts, which together with distribution of FadA homologues supports a possible association with gut disease.

Conclusion The proportion of fusobacteria in guts of southern Chinese populations are higher compared with several western and rural populations in line with the notion of environment/biogeography driving human gut microbiome composition. Several non-nucleatum taxa possess FadA homologues and were enriched in CRC cohorts; whether this imposes a risk in developing CRC and other gut diseases deserves further investigation.

Significance of this study

What is already known about this subject?
► *Fusobacterium nucleatum* are specifically enriched in gut microbiomes of individuals with colorectal cancer (CRC).
► The FadA adhesin and Fap2 lectin are implicated in the association between *F. nucleatum* and CRC.

What are the new findings?
► Non-CRC southern Chinese populations carry multiple known and novel fusobacterial taxa phylogenetically distinct from *F. nucleatum* in their guts; these taxa are nearly absent in other surveyed populations.
► Several fusobacterial taxa other than *F. nucleatum* are enriched in CRC cohorts relative to non-CRC controls.
► Homologues of the FadA adhesin were detected in several species of *Fusobacterium* including *F. varium* and *F. ulcerans*, suggesting potential associations with CRC and/or disease.

How might it impact on clinical practice in the foreseeable future?
► These findings indicate that CRC in southern Chinese populations may be linked to *F. varium* and other fusobacterial species in addition to *F. nucleatum*.
► Use of microorganisms as disease biomarkers or targets for therapeutic intervention needs to be tailored according to discrepancies in gut microbiome composition among human populations.

INTRODUCTION

Fusobacterium nucleatum is a bacterial pathogen most well-known for its association with colorectal cancer (CRC) in humans. Irrespective of biogeography, multiple studies have consistently reported enrichment of *F. nucleatum* in the guts[1–7] and tumour tissue[8–10] of CRC subjects compared with non-CRC cohorts. Furthermore, the association between *F. nucleatum* and CRC has been demonstrated through cell model studies, implicating two proteins FadA[11 12] and Fap2[13 14] in facilitating adherence, invasion and induction of oncogenic and inflammatory responses in CRC cells by *F. nucleatum*.

In contrast, relatively less is known about the biology of fusobacterial species other than *F. nucleatum* and their roles in human health, if any. According to the List of Prokaryotic names with Standing in Nomenclature (LPSN), there are 21 recognised species in the *Fusobacterium* genus at the time of writing. Apart from *F. nucleatum*, a few other species such as *F. necrophorum*,[15] *F. gonidiaformans*,[16] *F. periodonticum*,[17] *F. mortiferum*, *F. ulcerans* and *F. varium*[18] have been reported in human-associated samples. For example, *F. necrophorum* are often associated with thrombophlebitis of the internal jugular vein (termed Lemierre's syndrome), *F. gonidiaformans* found in urogenital and intestinal tracts,[16] *F. periodonticum* in oral cavities associated with squamous cell carcinoma,[19] *F. ulcerans* in skin ulcers[20] and *F. varium* in human guts associated with ulcerative colitis (UC).[21 22] Apart from cases of disease, their prevalence and relative abundances in guts of healthy individuals are relatively low, often below detection thresholds[23–28] consistent with the notion that the presence of *Fusobacterium* in human guts is specifically associated with CRC.[26]

We initially produced shotgun metagenomes using stools collected from 556 self-reported healthy individuals recruited for establishing a gut microbiota databank in Hong Kong (HKGutMicMap project). These data were analysed together with publicly-available stool metagenomes of other healthy subjects from Hong Kong,[2 29] Austria,[4] China,[30 31] Denmark,[32] France, Germany,[5] Israel,[33] Spain,[32] Sweden[34 35] and the USA,[3 27] as well as several rural populations from El Salvador, Peru,[36] Fiji,[37] Mongolia[38] and Tanzania[39 40] to assess variation in gut microbial community composition across different biogeographies. We serendipitously observed a consistent increase in prevalence and relative abundance of multiple fusobacterial species in the Hong Kong, Chinese and Spanish but not American, European and other rural cohorts, concordant with the idea of variation in human gut microbiomes primarily driven by environment/geography.[41] Here, we reconstructed fusobacterial genomes from the Hong Kong gut metagenomes and showed that *F. varium*, *F. ulcerans*, *F. mortiferum* and other as yet uncharacterised fusobacterial taxa are prevalent in this population. We then investigated whether these genomes contained characteristics that could indicate potential associations with cancer and/or disease. Findings reported here suggest that the fusobacterial lineages prevalent in the Chinese gut possess genomic potential to facilitate development of CRC and possibly other diseases.

MATERIALS AND METHODS

HKGutMicMap cohort sample collection and DNA sequencing

Subjects were recruited from the Hong Kong public as part of the HKGutMicMap study to generate gut microbiome profiles representative of the local, non-disease population. A research associate measured parameters such as body weight, waist circumference, body height and blood pressure and subjects were provided stool collection kits for self-collection. They were asked to deliver fresh stools to the laboratory within 2 hours of defecation. Faecal specimens were stored at −80°C until further processing. DNA was extracted from 0.1 g homogenised fractions of stool using the DNeasy PowerSoil Kit (QIAGEN, Hilden, Germany) following manufacturer's instructions. The concentration of extracted DNA was determined using the Qubit dsDNA BR Assay Kit (Thermo Fisher Scientific, Waltham, Massachusetts) and normalised to 20 ng/µL with 10 mM Tris-HCl. Normalised DNA samples were sent to a sequencing service provider (Novogene HK Company Limited, Wan Chai, Hong Kong) for library preparation and paired-end shotgun metagenomic sequencing

(Illumina NovaSeq 6000). A mock community sequencing control (ZymoBIOMICS Microbial Community DNA Standard, catalogue number D6305, Zymo Research, Irvine, California) was included.

Prevalence and relative abundances of fusobacterial species based on gut metagenomes

To examine prevalence and relative abundances of fusobacterial species in guts of human populations from different geographical backgrounds, we included non-CRC gut metagenome sequence data generated by previous studies in Hong Kong,[2 29] China,[30 31] USA,[3 27] Austria,[4] Denmark,[32] France, Germany,[5] Spain,[32] Israel,[33] Sweden,[34 35] El Salvador, Peru,[36] Fiji,[37] Mongolia[38] and Tanzania[37 40] (online supplementary table S1). The Hong Kong, Austria, France, Germany and one USA cohort[3] were comprised of CRC and non-CRC subjects. These gut metagenome data sets were chosen because they have been binned for microbial genomes.[42–44] Together with data generated from the HKGutMicMap cohort (present study), raw sequences were quality-filtered using Trimmomatic V.0.38 to remove adapter and low quality regions. Next, microbial community compositional profiles were inferred from quality-filtered sequences (forward reads) using MetaPhlAn2[45] V.2.6 with the v20 database. For each fusobacterial species identified by MetaPhlAn2, their prevalence rates were calculated based on the number of samples each species was detected in (ie, relative abundance >0%) divided by the total number of samples in the respective cohorts.

Binning fusobacterial population genomes from metagenomes

To explore genomic diversity of fusobacterial species in the Hong Kong population, we assembled metagenomes from the Hong Kong cohorts and binned population genomes (termed metagenome-assembled genomes (MAGs)) from each de novo assembly. Overlapping sequence pairs in the quality-filtered data were first merged to produce longer sequences, and then assembled together with unmerged pairs using MEGAHIT[46] V.1.1.1. Sequence coverage profiles were then obtained by mapping quality-filtered reads to their respective assemblies using BWA-MEM[47] V.0.7.17. With these coverage information, MAGs were binned from each of the metagenomes using MetaBAT[48] V.2.10.2, MetaBAT V.2.12.1 and MaxBin[49] V.2.2.5. A non-redundant set of MAGs were calculated by merging output from the three sets of bins using DASTool[50] V.1.1.0. The resulting non-redundant MAGs were quality-checked using the lineage workflow in CheckM[51] V.1.0.13. MAGs with >90% completeness and <5% contamination were retained, and their taxonomy was inferred using Genome Taxonomy Database[52] (GTDB) toolkit (GTDB-Tk) V.0.2.2 database release 86_2.

Construction of fusobacterial phylogenetic trees

We downloaded fusobacterial reference genomes from the National Centre for Biotechnology Information (NCBI) RefSeq database (release 89), and MAGs from recent publications that have assembled microbial genomes from the human metagenome data sets used above.[42–44] These genomes were checked for completeness and contamination using CheckM, and only those with >90% completeness and <5% contamination were retained. We constructed two phylogenetic trees of the *Fusobacterium* genus—one using a dereplicated set of fusobacterial reference genomes and MAGs to highlight existing genomic diversity, and the other using all genomes to explore

distribution of putative virulence protein homologues in this genus (described in section below). We included *Cetobacterium*, a member of the *Fusobacteriaceae* family as outgroup taxa in both trees. For the first tree, genomes were dereplicated using dRep V.1.4.3 based on genome distances and average nucleotide identities (ANI),[53] and a concatenated amino acid alignment of 120 phylogenetically informative single-copy bacterial marker genes was generated using GTDB-Tk. A maximum-likelihood tree was constructed based on this alignment using RAxML[54] V.8.2.11, and node support estimated from 100 bootstraps. For the second tree without genome dereplication, a concatenated amino acid alignment consisting of all genomes was generated and used to infer a bootstrapped phylogeny as per the first tree.

Annotating genes in fusobacterial genomes

First, protein-coding sequences in all MAGs and reference genomes were translated into amino acid sequences using Prodigal[55] V.2.6.3. Amino acid sequences were aligned against the UniRef100 database (March 2018) using DIAMOND[56] V.0.9.24 (≥30% sequence identity and ≥70% alignment length between query and reference) to identify gene families, and aligned counts were collated according to Kyoto Encyclopaedia of Genes and Genomes orthology to infer presence/absence of gene families. ANI comparisons were performed using FastANI[57] V.1.1.

To explore presence/absence of two known CRC-associated fusobacterial genes FadA and Fap2, we annotated fusobacterial genomes using eggNOG-mapper[58] V.2.0.1 with reference to eggNOG database V.5.0. The presence/absence of FadA and Fap2 homologues was visualised on a phylogenetic tree consisting of all 663 fusobacterial genomes. Amino acid gene trees were constructed for both FadA and Fap2 by aligning the putative homologues using MAFFT[59] V.7.407 and inferring a maximum likelihood tree using RAxML. Both trees were midpoint-rooted using GenomeTreeTk V.0.0.53 (https://github.com/dparks1134/GenomeTreeTk).

Isolation of Fusobacterium from stools

Frozen stools were thawed and diluted in brain heart infusion culture medium. Dilutions were inoculated onto blood agar plates and anaerobically cultured at 37°C for 2 days. Colonies were identified using a MALDI Biotyper (Bruker, Billerica, Massachusetts). Colonies identified as *Fusobacterium* were subcultured onto fresh blood agar plates. Genomic DNA was extracted from pure cultures using a Gentra Puregene Yeast/Bact. DNA isolation Kit (QIAGEN, Hilden, Germany), and sent to Novogene HK for library preparation and paired-end shotgun metagenomic sequencing (Illumina NovaSeq 6000). Trimmomatic quality-filtered reads were assembled using MEGAHIT V.1.1.1 and annotated using eggNOG-mapper V.2.0.1 with reference to eggNOG database V.5.0.

Data availability

Raw sequence data generated for this study are available in the Sequence Read Archive under BioProject accession PRJNA557323.

RESULTS

The HKGutMicMap cohort

At the time of analysis, the HKGutMicMap cohort representing the general population in Hong Kong consisted of 556 subjects with shotgun metagenome data. These subjects were self-reported as healthy with no chronic disease. There were 294 females to 262 males, and their median age at time of sample collection was 51 years (SD 16.3 years). The median body mass index was 22.7 kg m^{-2} (SD 3.4). These and other parameters such as body weight, blood pressure and waist circumference are listed in online supplementary table S2.

F. mortiferum, *F. ulcerans* and *F. varium* are prevalent in Chinese populations irrespective of CRC disease status

In total, 3157 stool metagenomes comprising non-CRC and CRC subjects were included in this study. These metagenomes represent populations from China (Hong Kong, Shenzhen and Zhejiang), USA, Austria, Denmark, France, Germany, Spain, Sweden, Israel, El Salvador, Peru, Fiji, Mongolia and Tanzania. To assess the distribution of fusobacterial species across biogeography of these populations, quality-filtered sequences from each metagenome were mapped to lineage-specific marker genes using MetaPhlAn2 to produce prevalence and relative abundance estimates.

In non-CRC subjects (n=2515), overall gut microbial community composition significantly differed among cohorts (p<0.05, permutational multivariate analysis of variance; figure 1A, online supplementary figure S1). At the phylum level, the three Chinese and USA cohorts had higher relative abundances of *Bacteroidetes* compared with *Firmicutes* (63% vs 30%), whereas *Firmicutes* were more relatively abundant in the other cohorts compared with *Bacteroidetes* (55% vs 29%). Peru was the exception with *Actinobacteria* being the most dominant phylum (60%) (online supplementary table S3). In addition, *Spirochaetes* were detected at >1% relative abundance only in El Salvador, Fiji and Tanzania cohorts, compared with an average 0.004% in the Western and Chinese cohorts. *Fusobacterium* were also more relatively abundant in the Chinese and Spanish (average 0.47%) compared with other cohorts (0.01%). We were interested in the comparatively higher relative abundances of *Fusobacterium* since *F. nucleatum* has been widely implicated in CRC. Within the fusobacterial genus, *F. mortiferum*, *F. nucleatum*, *F. ulcerans* and *F. varium* were more prevalent and relatively abundant in the Chinese cohorts relative to others including Spain (p<0.001, Kruskal-Wallis test adjusted for false discovery rate) (table 1, figure 1B).

In CRC subjects (n=642), average fusobacterial relative abundances in subjects from Hong Kong were higher than the USA, German and Austrian, but not the French cohort (online supplementary figure S2). *F. nucleatum* was detected in all six cohorts as expected for CRC. *F. varium*, *F. ulcerans* and *F. mortiferum* were more prevalent in Hong Kong (online supplementary table S4). In addition, the French were enriched in *F. gonidiaformans* and *F. necrophorum* relative to the others. *F. ulcerans* was also present in the Austrian cohort, however, its prevalence was still six fold higher in Hong Kong. These findings indicate that *F. mortiferum*, *F. ulcerans* and *F. varium* are typically more common and detected at higher relative abundances in the guts of Hong Kong populations compared with some North Americans and Europeans, irrespective of CRC disease status.

Several fusobacterial species other than *F. nucleatum* are enriched in CRC

F. nucleatum and more broadly the *Fusobacterium* genus have been shown to be enriched in the guts of CRC patients compared with non-CRC controls,[6 7] although associations between CRC and other fusobacterial species have not been specifically mentioned. Since *F. mortiferum*, *F. ulcerans* and *F. varium* were more prevalent and relatively abundant in the guts of Chinese populations, we wanted to know whether their distributions and abundances were changed in association with CRC akin to
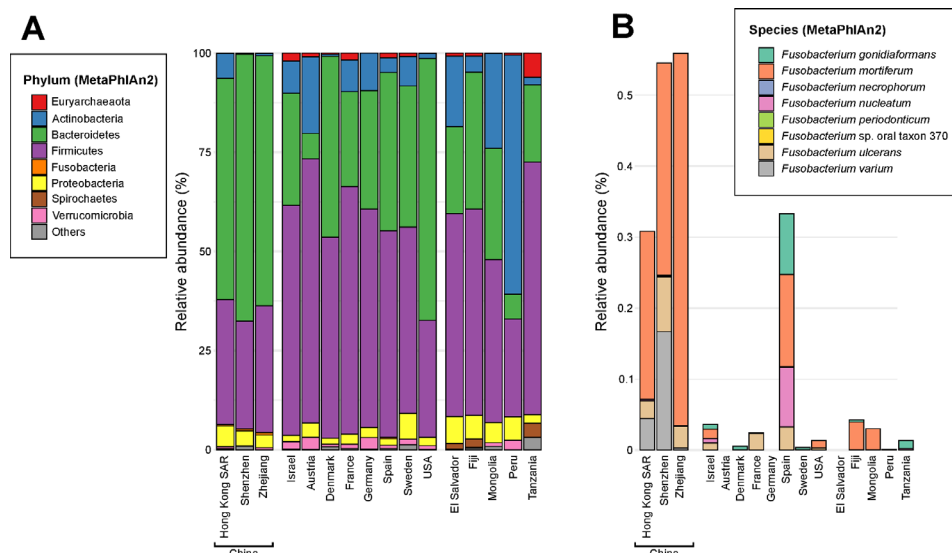
Figure 1 Microbial community composition of the typical human gut. (A) Average relative abundances of microbial phyla detected in human stool metagenomes from the HKGutMicMap cohort (this study) and previously described non-colorectal cancer (CRC) individuals from various geographical backgrounds. (B) Average relative abundances of fusobacterial species. The stacked bars represent cohorts from: Hong Kong (HKGutMicMap and two others),[2 29] Austria,[4] China,[30 31] Denmark,[32] France, Germany,[5] Israel,[33] Spain,[32] Sweden[34 35] and the USA,[3 27] as well as several rural populations from El Salvador, Peru,[36] Fiji,[37] Mongolia[38] and Tanzania.[39 40] Relative abundances were calculated using MetaPhlAn2 on quality-filtered metagenome sequences. Values shown in (B) for fusobacterial species are percentages of the total community. For case-control studies with CRC cohorts,[2–5 29] only non-CRC individuals were included in the calculation of relative abundances.

F. nucleatum. We compared the prevalence and relative abundances of the fusobacterial species between CRC and non-CRC subjects in studies with case-control cohorts, and found that values for *F. gonidiaformans* and *F. nucleatum* were increased in all six CRC cohorts, and *F. periodonticum* and *F. varium* in five of six CRC cohorts compared with non-CRC cohorts (online supplementary tables S4,S5). A generalised linear model taking into account cohorts indicated that relative abundances of *F. nucleatum* and *F. varium* were significantly associated with CRC disease ($p < 0.05$), although the prevalence of *F. varium* was not as striking as *F. nucleatum*.

Population genomes from Chinese gut metagenomes reveal expanded diversity in the *Fusobacterium* genus

At the time of writing, there were 157 fusobacterial genomes in the NCBI RefSeq database (release 89), of which 65 (41.4%), 36 (22.9%) and 17 (10.8%) are classified as *F. nucleatum*, *F. necrophorum* and *F. periodonticum*, respectively (online supplementary table S6). The other 18 recognised fusobacterial species according to the LPSN and any novel taxa yet to be classified are represented by the remaining 39 genomes. Since MetaPhlAn2 indicated that fusobacterial species such as *F. mortiferum*, *F. ulcerans* and *F. varium* were more prevalent in the guts of Chinese populations, we wanted to explore and expand known genomic diversity of these less characterised fusobacterial lineages. Using metagenomes from the Hong Kong cohort (inclusive of non-CRC and CRC subjects), we binned 171 high quality fusobacterial MAGs (>90% complete, <5% contamination based on CheckM's lineage workflow) (online supplementary table S7). Another four high quality fusobacterial MAGs we previously binned from gut metagenomes of patients from a clinic were also included in this study (annotated in online supplementary table S7). In addition, recent efforts in characterising genomic diversity of human microbiomes[42–44] have yielded an additional 336 high quality fusobacterial MAGs (online supplementary table S7). Together with 152 high quality fusobacterial genomes

from RefSeq R89, we first dereplicated these 663 genomes and inferred a genome tree to establish their phylogenetic relationships to one another. Dereplication was performed to highlight existing genome diversity in the *Fusobacterium* genus, resulting in a phylogenetic tree comprising 218 unique fusobacterial genomes. Taxonomic information for all MAGs and reference genomes inferred according to the GTDB release 86_v2 was then appended onto the phylogenetic tree (figure 2).

Four major monophyletic lineages (termed clades) were resolved within the *Fusobacterium* genus congruent with taxonomic inferences produced by GTDB (denoted with suffixes *Fusobacterium*, *Fusobacterium*_A, *Fusobacterium*_B and *Fusobacterium*_C) (figure 2). The clade denoted as *Fusobacterium* was comprised of *F. nucleatum* including its traditional subspecies *animalis*, *vincentii*, *nucleatum* and *polymorphum*, *F. hwasookii*, *F. periodonticum*, *F. massiliense* and *F. russii*. *Fusobacterium*_A was comprised of *F. ulcerans*, *F. varium*, *F. mortiferum* and various unclassified fusobacterial genomes; *Fusobacterium*_B comprised of *F. perfoetens* and other unclassified genomes; *Fusobacterium*_C of *F. gonidiaformans* and *F. necrophorum*. Lineages within the *Fusobacterium*_A and *Fusobacterium*_B clades were highly represented by genomes derived from Hong Kong and Chinese metagenomes (48 of 67 genomes; many genomes from RefSeq do not have accompanying geographic information and were assumed to be of non-Chinese origin) (figure 2). In contrast, the *Fusobacterium* and *Fusobacterium*_C clades were more represented by genomes from other regions (only 7 of 151 genomes were from Chinese sources). MAGs from Chinese populations collectively increased phylogenetic diversity of the overall tree by 14.3% based on branch lengths, indicating that the Chinese gut harbours novel fusobacterial diversity not yet represented by reference genomes. To demonstrate that these novel fusobacteria were indeed more prevalent in Chinese populations, we mapped sequences from all non-CRC samples to the dereplicated set of 218 fusobacterial genomes and counted the proportion of aligned sequences in each cohort. The Chinese had 10–100-fold higher proportions of

Table 1 Prevalence and average relative abundances of fusobacterial species in non-CRC subjects

Country	Fusobacterium gonidiaformans	F. mortiferum	F. necrophorum	F. nucleatum	F. periodonticum	F. oraltaxon 370	F. ulcerans	F. varium
Hong Kong (China)								
Prevalence (%)	0.384	15.493	0.128	4.609	0.640	0.128	9.475	13.828
Relative abundance (%)	0.000	0.236	0.000	0.002	0.000	0.000	0.025	0.044
Shenzhen (China)								
Prevalence (%)	0.000	10.526	0.000	13.158	7.895	0.000	23.684	39.474
Relative abundance (%)	0.000	0.299	0.000	0.001	0.001	0.000	0.077	0.167
Zhejiang (China)								
Prevalence (%)	0.000	30.345	0.000	0.690	0.690	0.000	10.345	6.207
Relative abundance (%)	0.000	0.525	0.000	0.001	0.000	0.000	0.030	0.003
Israel								
Prevalence (%)	2.667	0.667	0.000	3.333	0.667	0.000	2.000	0.000
Relative abundance (%)	0.007	0.013	0.000	0.006	0.001	0.000	0.009	0.000
Austria								
Prevalence (%)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Relative abundance (%)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Denmark								
Prevalence (%)	0.565	0.000	0.000	0.565	0.565	0.000	0.000	0.000
Relative abundance (%)	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000
France								
Prevalence (%)	1.639	1.639	0.000	1.639	1.639	0.000	1.639	0.000
Relative abundance (%)	0.000	0.000	0.000	0.000	0.000	0.000	0.024	0.000
Germany								
Prevalence (%)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Relative abundance (%)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Spain								
Prevalence (%)	4.110	1.826	1.370	6.849	3.196	0.000	1.370	0.000
Relative abundance (%)	0.085	0.130	0.000	0.085	0.000	0.000	0.032	0.000
Sweden								
Prevalence (%)	2.098	0.000	0.000	1.399	0.000	0.000	0.699	0.000
Relative abundance (%)	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000
USA								
Prevalence (%)	0.980	1.471	0.000	2.451	2.941	0.000	2.941	0.490
Relative abundance (%)	0.000	0.010	0.000	0.000	0.000	0.000	0.003	0.000
El Salvador								
Prevalence (%)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Relative abundance (%)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Fiji								
Prevalence (%)	4.545	6.818	0.000	1.705	2.273	0.000	0.000	0.000
Relative abundance (%)	0.003	0.039	0.000	0.000	0.000	0.000	0.000	0.000
Mongolia								
Prevalence (%)	0.000	2.727	0.000	0.909	0.000	0.000	0.000	0.000
Relative abundance (%)	0.000	0.030	0.000	0.000	0.000	0.000	0.000	0.000
Peru								
Prevalence (%)	1.299	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Relative abundance (%)	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Tanzania								
Prevalence (%)	1.493	0.000	0.000	1.493	0.000	0.000	0.000	0.000
Relative abundance (%)	0.011	0.000	0.000	0.002	0.000	0.000	0.000	0.000

CRC, colorectal cancer.

reads mapped to *Fusobacterium_A* genomes compared with other cohorts (online supplementary table S8, figure S3), consistent with MetaPhlAn2 estimates of higher relative abundances of *Fusobacterium_A* lineages in Chinese samples. Similarly, in CRC samples the Hong Kong cohort generally showed 10–100-fold higher proportions of reads mapped to *Fusobacterium_A* genomes compared

with Austrian, French, German and USA samples (online supplementary table S9, figure S4).

Circumscribing new species in the *Fusobacterium* genus
Using the 218 dereplicated fusobacterial genomes, we performed pairwise ANI comparisons to establish species boundaries
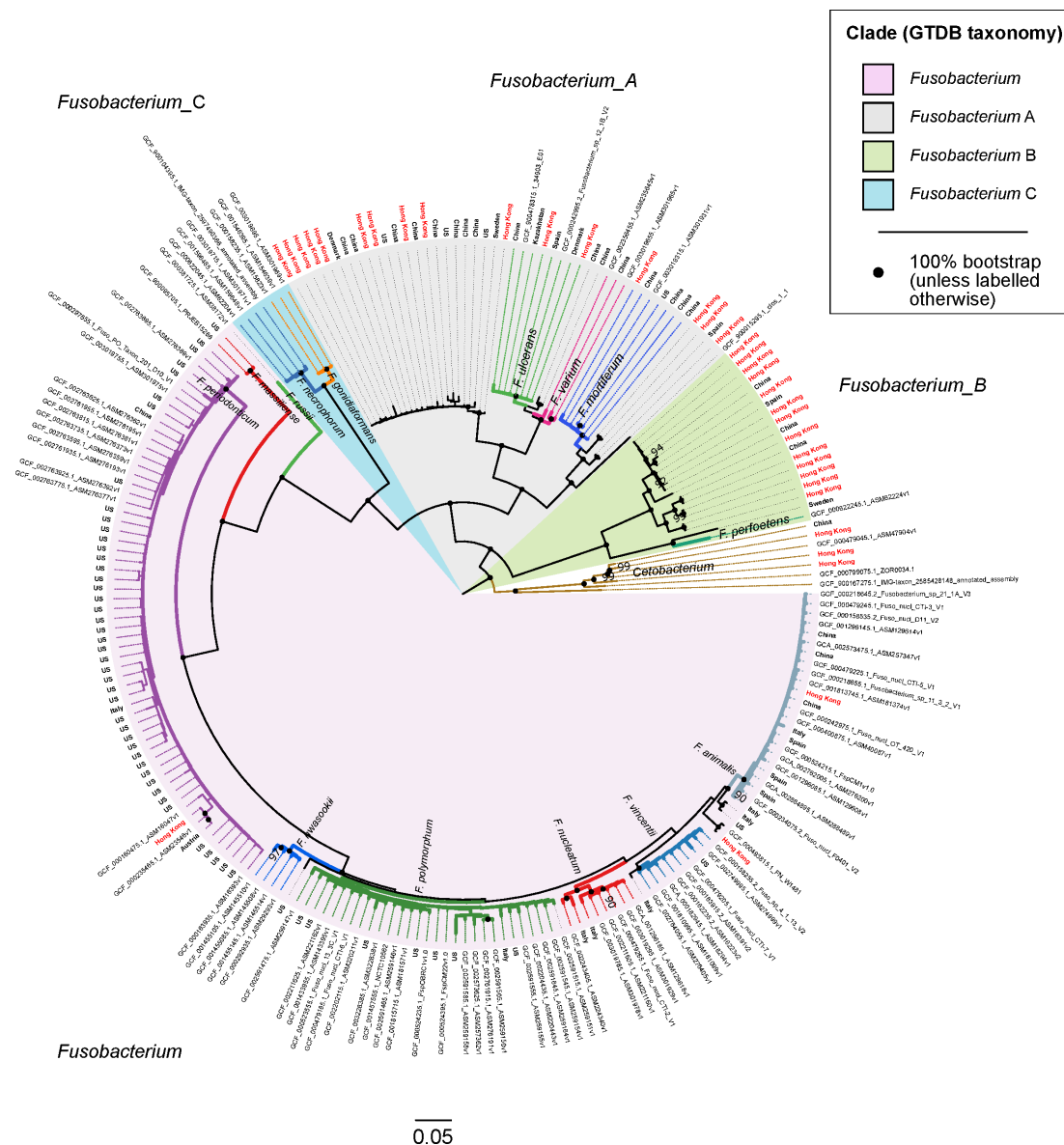
Figure 2 Phylogenetic tree showing evolutionary relationships among 218 fusobacterial genomes. Seven *Cetobacterium* genomes serve as outgroup to root the tree. Genomes in this figure are from a dereplicated set of 676 fusobacterial and *Cetobacterium* genomes assembled from gut metagenomes from Hong Kong (HKGutMicMap, Yu *et al*[2] and Coker *et al*[29]) and other regions,[42–44] and reference genomes downloaded from RefSeq (release 89). Reference genomes obtained from RefSeq are labelled with their corresponding accession numbers, while metagenome-assembled genomes have branch labels showing their country of origin (those from Hong Kong are in red text). All 676 genomes were >90% complete and had <5% contamination based on the lineage workflow in CheckM,[51] and were dereplicated using dRep[53] to highlight existing genome diversity of the *Fusobacterium* genus in this figure. A concatenated amino acid alignment was produced to infer taxonomy of the genomes according to the genome taxonomy database (GTDB),[52] and subsequently used to construct maximum likelihood trees using RAxML.[54] Four major monophyletic clades in the *Fusobacterium* genus are shaded and denoted with suffixes according to the GTDB. Branch colours are intended to delineate species boundaries (indicated by labels) and do not represent any taxa in particular; genomes without species designations have black branches. Black circles at nodes represent 100% bootstrap support unless otherwise indicated (no less than 90% bootstrap). Scale bar indicates number of amino acid substitutions per site.

with reference to intraspecies and interspecies cutoffs derived from published studies (intraspecies >95% ANI; interspecies 78%–95%).[59 60] With these cutoffs, we identified (i) six putative species in the *Fusobacterium_B* clade not including *F. perfoetens*, (ii) three species basal to *F. mortiferum*, (iii) one species sister to *F. ulcerans*, (iv) one species sister to the *F. ulcerans* and *F. varium* lineage, (v) one species basal to the lineage containing *F. polymorphum, nucleatum, vincentii* and *animalis* and (vi) two species basal/sister to *F. animalis* (online supplementary figure

S5, table S10). These genomes share <95% ANI to any circumscribed fusobacterial taxa, and could represent novel or one of the 18 recognised species in the LPSN yet to have genome representation. In addition to drawing species boundaries, we could infer the degree of intraspecies genome similarity by comparing the initial number of 663 genomes to the resulting number of dereplicated genome clusters. For example, we observed that *F. mortiferum* was highly clonal despite its high prevalence in Chinese populations, whereas *F. periodonticum* genomes were
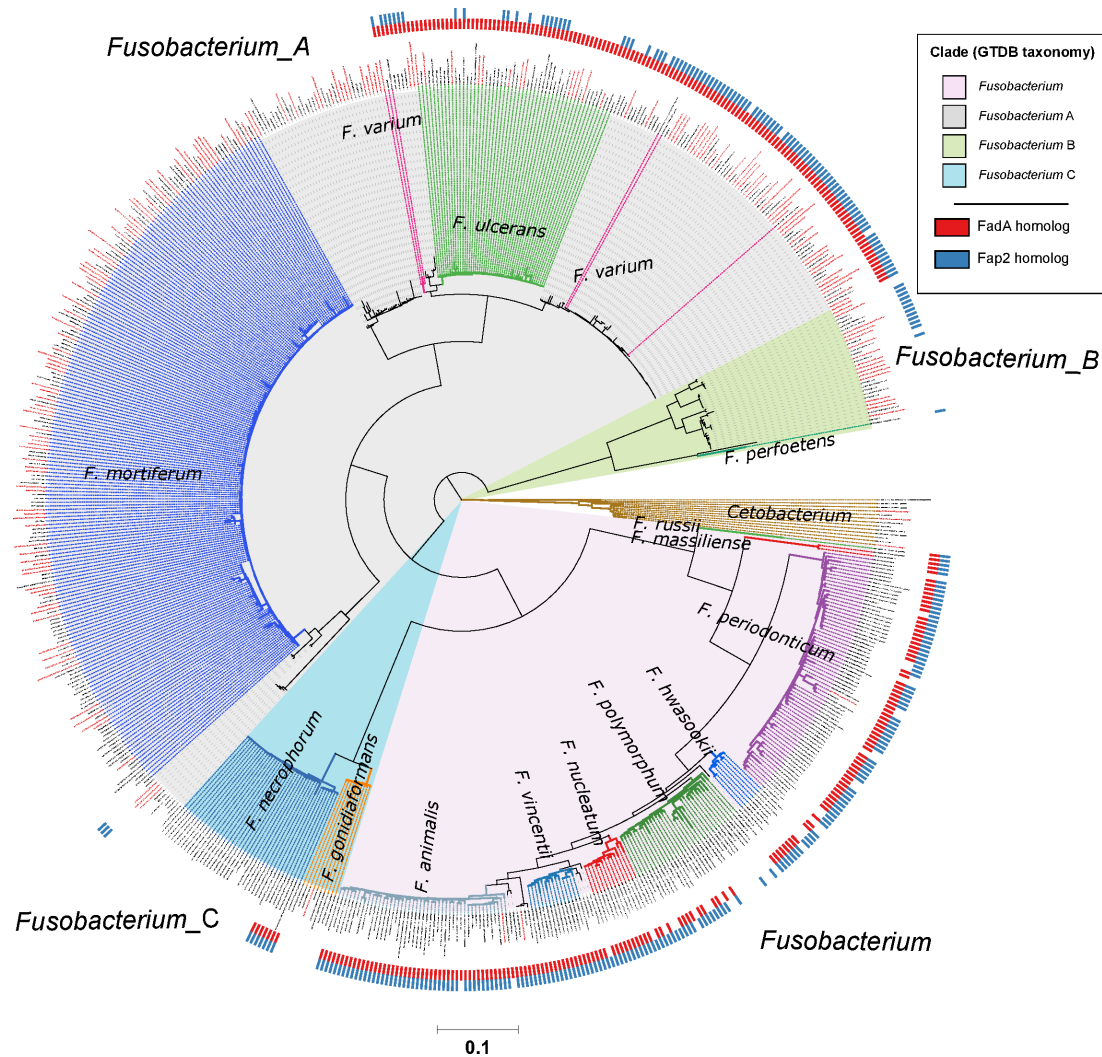
Figure 3 Distribution of FadA and Fap2 homologues in the *Fusobacterium* genus. Red and blue ticks next to branch tips indicate detection of FadA and Fap2 homologues, respectively, in the corresponding genomes. Homologues were identified using the eggNOG-mapper[58] with reference to the eggNOG database V.5.0. This phylogenetic tree consists of 663 fusobacterial and 13 *Cetobacterium* genomes assembled from gut metagenomes from Hong Kong (HKGutMicMap cohort from this study, Yu *et al*[2] and Coker *et al*[29] cohorts) and other regions,[42–44] and reference genomes downloaded from RefSeq (release 89). Reference genomes obtained from RefSeq are labelled with their corresponding accession numbers, while metagenome-assembled genomes are labelled with bin IDs. Genomes from Hong Kong have labels in red. All genomes were >90% complete and had <5% contamination based on the lineage workflow in CheckM.[51] A concatenated amino acid alignment was produced to infer taxonomy of the genomes according to the genome taxonomy database (GTDB),[52] and subsequently used to construct maximum likelihood trees using RAxML.[54] Four major monophyletic clades in the *Fusobacterium* genus are shaded and denoted with suffixes according to the GTDB. Branch colours are intended to delineate species boundaries (indicated by labels) and do not represent any taxa in particular; genomes without species designations have black branches. Scale bar indicates number of amino acid substitutions per site.

more variable and formed more clusters of unique genomes compared with *F. mortiferum* (figures 2 and 3, online supplementary table S11).

Fusobacterial genome features possibly associated with disease

Previous functional analyses of CRC gut metagenomes have revealed features such as shifts towards amino acid degradation and trimethylamine (TMA) production via choline metabolism.[6 7 61] We annotated the fusobacterial MAGs and observed that while they did not contain key genes involved in TMA production (TMA-lyase (*cut*C, K20038), and L-carnitine/ gamma-butyrobetaine antiporter (*cai*T, K05245)), several orthologues such as proline iminopeptidase (K01259), glutamate formiminotransferase (K00603) and tryptophanase (K01667)

were prevalent in genomes from the *Fusobacterium* clade (online supplementary table S12). Moreover, *Fusobacterium* clade genomes possess genes that may be involved in the catabolism of amino acids and production of glucose (phosphoenolpyruvate carboxykinase K01610, fructose-bisphosphate aldolase K01623, oxaloacetate decarboxylase K01571), as well as several other features that could be linked to cancer such as iron scavenging (K07230, K07243, K11707, K11708, K11709, K11710),[62] ceramide glucosyltransferase (K00720) involved in production of glycosylated sphingolipids[63] and para-aminobenzoate synthetase (K01664, K01665) in the production of folate.[64] Likewise, urease (K01428–K01430)[65] was detected in *Fusobacterium_A* and *Fusobacterium_B* clades but not *Fusobacterium*. Some of these features are consistent with those identified in CRC gut microbiota metagenomes, though it is important to point out
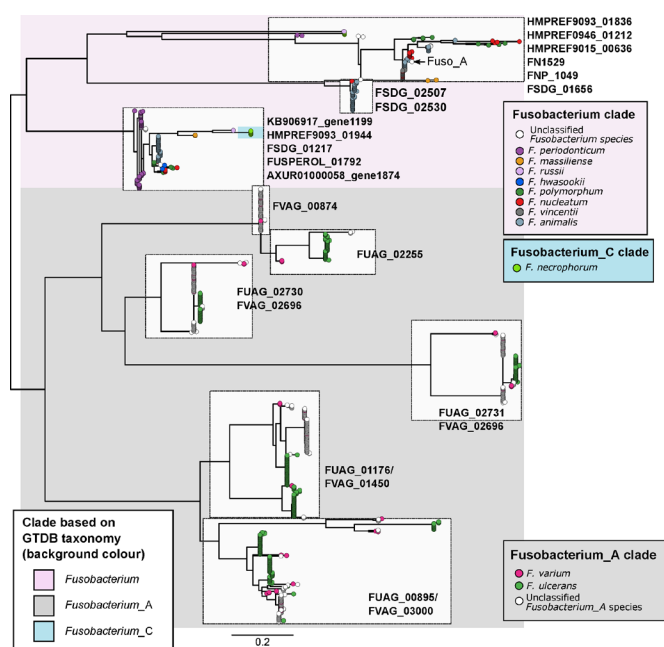
Figure 4 Phylogenetic relationships of FadA protein homologues identified in fusobacterial genomes. Figure shows a maximum-likelihood tree of aligned amino acid sequences of FadA homologues rooted at the midpoint. Each tip represents a homologue and is coloured according to species of the genome homologues were found in. Text labels next to tree tips indicate the corresponding seed orthologues in the eggNOG database. Background shading is according to the four major monophyletic clades identified in the genome-based phylogenetic tree in figure 2. Scale bars indicate amino acid substitutions per site.

that these findings do not imply that fusobacteria contribute wholly to the altered microbial functional signature in CRC guts[6 7] as they are typically <1% relative abundance. In addition, the distribution of features by clades suggest that disease associations, if any, likely vary among the fusobacterial lineages.

Homologues of colorectal cancer-associated fadA and Fap2 are present in several fusobacterial species

Previous cell model studies have identified two proteins in *F. nucleatum* that allow the bacterium to potentiate CRC, namely the FadA adhesin[9 10] and Fap2 lectin.[11 12] To identify whether fusobacterial species other than *F. nucleatum* also possess similar genes that may allow them to interact with CRC cells, we annotated all 663 fusobacterial genomes with reference to the eggNOG database and searched for putative homologues. A phylogenetic tree incorporating all 663 genomes was constructed to visualise distribution of these homologues within the *Fusobacterium* genus. For FadA, a total of 999 homologues (online supplementary table S13) were identified in 311 genomes including all lineages belonging to the *Fusobacterium* clade, a monophyletic subset of *F. necrophorum*, and in *F. varium*, *F. ulcerans* and several uncharacterised monophyletic taxa in the *Fusobacterium*_A clade (figure 3). These FadA homologues possibly comprise three or more protein families as shown by a protein tree constructed from amino acid alignments. Sequences from the *Fusobacterium*_A clade were distinct compared with those from the *Fusobacterium* clade, while homologues from *F. necrophorum* were placed together with the *Fusobacterium* homologues (figure 4). These observations indicate that FadA homologues from *F. varium*, *F. ulcerans* and uncharacterised *Fusobacterium*_A lineages could have distinct roles compared

with homologues found in the *Fusobacterium* clade. For Fap2, we identified 754 putative homologues in 288 genomes (online supplementary table S14). Lineages in which Fap2 homologues were identified largely overlapped with FadA, encompassing members of the *Fusobacterium* clade, a monophyletic subset of *F. necrophorum* genomes, several *F. varium* and *F. ulcerans*, and in a subset of *Fusobacterium*_B clade genomes (figure 3, online supplementary figure S6). Overall distribution of the FadA and Fap2 homologues in the *Fusobacterium* genus suggests that the potential association with CRC may be present in several distinct fusobacterial lineages. Since some of these fusobacterial species were increased in relative abundance and prevalence in CRC compared with non-CRC subjects, the detection of FadA and Fap2 homologues suggest that species such as *F. varium* may have the ability to potentiate disease akin to *F. nucleatum*.

To check that the fusobacterial MAGs recovered from metagenome data are representative of actual genomes, we isolated and sequenced genomes of eight fusobacteria obtained from five stool samples. These genomes were classified as *Fusobacterium*_A (seven genomes) and *F. ulcerans* (one), and scored >99% ANI to MAGs recovered from Hong Kong gut metagenomes (online supplementary table S15). Moreover, they contained FadA and Fap2 homologues consistent with their phylogeny, providing confidence that MAGs recovered here indeed represent real microbial genomes. Nevertheless, we recognise that MAG validation with only eight isolates is inadequate, and more work is needed to verify MAGs representing other fusobacterial lineages.

DISCUSSION

While it has been established in human populations from various geographical backgrounds that *F. nucleatum* is associated with CRC,[1–5] less is known about the distribution of other fusobacterial species in human guts. Here, we showed that fusobacterial lineages such as *F. ulcerans*, *F. varium*, *F. mortiferum* and multiple uncharacterised taxa are more prevalent in the guts of non-CRC Chinese and Spanish cohorts compared with counterparts from several geographical regions. While these non-nucleatum taxa may simply reflect biogeographical differences in human gut microbiome composition, we saw two lines of evidence suggesting that they could possess oncogenic and/or disease-causing potential: (i) increased prevalence and relative abundances in CRC compared with non-CRC cohorts (online supplementary table S5) and (ii) detection of virulence gene homologues in multiple monophyletic lineages (figure 3). Taken together, these evidence suggest that *F. periodonticum*, a subset of *F. necrophorum*, *F. varium* and *F. ulcerans* together with their uncharacterised sister lineages, *F. hwasookii*, *F. massiliense* and *F. russii* might have a role in the development of CRC. These implicated lineages are consistent with a set of 'active versus passive invader' species proposed by Manson McGuire and colleagues based on genomic features such as genome size, presence of FadA-related proteins, expanded number of membrane protein-encoding genes and MORN2 protein domains.[18] In addition, their link to disease is supported by independent microbial community data and cell model studies. For example, a recent microbial community composition survey indicated that *F. periodonticum* in the oral cavity is associated with oral squamous cell carcinoma.[19] Another example is of *F. necrophorum*, in which blood culture-based surveys of fusobacterial infections have indicated that this species was the second most common isolate after *F. nucleatum*.[66 67] As for *F. varium* and *F. ulcerans*, less is known about their distribution and association with cancers or disease. Gut microbiota surveys in Japanese cohorts

have indicated that *F. varium* is associated with UC,[21 22] and a genome sequencing study of *F. varium* strain Fv113-g1 isolated from a UC patient reported expression of FadA homologues in monocultures simulating in vivo conditions within the human gut.[68] Our data on FadA homologues show that sequences from the *Fusobacterium*_A clade (*F. varium*, *F. ulcerans* and other uncharacterised sister taxa) are not identical to sequences from the *Fusobacterium* clade (in which the CRC-associated *F. nucleatum* is located) (figure 4), thereby suggesting distinct functions or targets among these homologues. While the presence/absence of these gene homologues do not directly translate to invasiveness,[69] we postulate that *Fusobacterium*_A taxa and their copies of the FadA homologue can be risk factors for diseases other than CRC.

In light of the inference that *Fusobacterium*_A taxa are prevalent in Chinese populations and could be potential risk factors of disease in humans, a limitation of this study is the lack of published data or cultured isolates to validate our observations. Findings reported here imply that disease-association is possible in this clade, from which we have isolated eight *Fusobacterium*_A members with genomes that match MAGs recovered from metagenomes. The immediate next step is to test these isolates in cell and animal model experiments to determine whether they have the potential to facilitate CRC or other diseases akin to *F. nucleatum*. Specifically, the role of virulence gene homologues such as FadA and Fap2 can be studied via knockout/knockdown experiments to assess their impact on disease outcomes. Following this, further isolation and testing of other uncharacterised fusobacterial lineages will provide a more comprehensive understanding of the biology and disease associations outside the *F. nucleatum* complex.

In conclusion, while fusobacterial species other than *F. nucleatum* have not been identified as risk factors likely owing to their almost non-existence in western populations and ubiquity in non-CRC southern Chinese populations, our findings suggest that there is potential in some of these prevalent but overlooked fusobacterial lineages to facilitate CRC. If any positive associations are confirmed, individuals carrying the corresponding taxa in their guts should be assessed for predispositions to disease. Findings reported here underscore the variability in gut microbiota composition across populations, and support ongoing efforts to characterise microbial diversity of the human microbiome.

Data availability statement Data are available in a public, open access repository. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA557323. Gut metagenome and fusobacterial isolate genome sequence data are available in the Sequence Read Archive (SRA) under BioProject accession PRJNA557323.

ORCID iDs
Yun Kit Yeoh http://orcid.org/0000-0003-0241-6117
Zigui Chen http://orcid.org/0000-0002-8577-1298
Martin C S Wong http://orcid.org/0000-0001-7030-9853
Jun Yu http://orcid.org/0000-0001-9239-2416
Siew C Ng http://orcid.org/0000-0002-6850-4454
Francis K L Chan http://orcid.org/0000-0001-7388-2436

REFERENCES

1. Dai Z, Coker OO, Nakatsu G, *et al*. Multi-Cohort analysis of colorectal cancer metagenome identified altered bacteria across populations and universal bacterial markers. *Microbiome* 2018;6:70.
2. Yu J, Feng Q, Wong SH, *et al*. Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. *Gut* 2017;66:70–8.
3. Vogtmann E, Hua X, Zeller G, *et al*. Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. *PLoS One* 2016;11:e0155362.
4. Feng Q, Liang S, Jia H, *et al*. Gut microbiome development along the colorectal adenoma-carcinoma sequence. *Nat Commun* 2015;6:6528.
5. Zeller G, Tap J, Voigt AY, *et al*. Potential of fecal microbiota for early-stage detection of colorectal cancer. *Mol Syst Biol* 2014;10:766.
6. Thomas AM, Manghi P, Asnicar F, *et al*. Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. *Nat Med* 2019;25:667–78.
7. Wirbel J, Pyl PT, Kartal E, *et al*. Meta-Analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. *Nat Med* 2019;25:679–89.
8. Nakatsu G, Li X, Zhou H, *et al*. Gut mucosal microbiome across stages of colorectal carcinogenesis. *Nat Commun* 2015;6:8727.
9. Castellarin M, Warren RL, Freeman JD, *et al*. Fusobacterium nucleatum infection is prevalent in human colorectal carcinoma. *Genome Res* 2012;22:299–306.
10. Kostic AD, Chun E, Robertson L, *et al*. Fusobacterium nucleatum potentiates intestinal tumorigenesis and modulates the tumor-immune microenvironment. *Cell Host Microbe* 2013;14:207–15.
11. Rubinstein MR, Wang X, Liu W, *et al*. Fusobacterium nucleatum promotes colorectal carcinogenesis by modulating E-cadherin/β-catenin signaling via its FadA adhesin. *Cell Host Microbe* 2013;14:195–206.
12. Han YW, Ikegami A, Rajanna C, *et al*. Identification and characterization of a novel adhesin unique to oral Fusobacteria. *J Bacteriol* 2005;187:5330–40.
13. Abed J, Emgård JEM, Zamir G, *et al*. Fap2 Mediates *Fusobacterium nucleatum* Colorectal Adenocarcinoma Enrichment by Binding to Tumor-Expressed Gal-GalNAc. *Cell Host Microbe* 2016;20:215–25.
14. Gur C, Ibrahim Y, Isaacson B, *et al*. Binding of the Fap2 protein of *Fusobacterium nucleatum* to human inhibitory receptor TIGIT protects tumors from immune cell attack. *Immunity* 2015;42:344–55.
15. Riordan T. Human infection with *Fusobacterium necrophorum* (Necrobacillosis), with a focus on Lemierre's syndrome. *Clin Microbiol Rev* 2007;20:622–59.
16. Citron DM. Update on the taxonomy and clinical aspects of the genus *Fusobacterium*. *Clin Infect Dis* 2002;35:S22–7.
17. Strauss J, White A, Ambrose C, *et al*. Phenotypic and genotypic analyses of clinical *Fusobacterium nucleatum* and *Fusobacterium periodonticum* isolates from the human gut. *Anaerobe* 2008;14:301–9.
18. Manson McGuire A, Cochrane K, Griggs AD, *et al*. Evolution of invasion in a diverse set of Fusobacterium species. *mBio* 2014;5:e01864.
19. Yang C-Y, Yeh Y-M, Yu H-Y, *et al*. Oral microbiota community dynamics associated with oral squamous cell carcinoma staging. *Front Microbiol* 2018;9:862.
20. Adriaans B, Shah H. *Fusobacterium ulcerans* sp. nov. from Tropical Ulcers. *Int J Syst Bacteriol* 1988;38:447–8.
21. Ohkusa T, Sato N, Ogihara T, *et al*. *Fusobacterium varium* localized in the colonic mucosa of patients with ulcerative colitis stimulates species-specific antibody. *J Gastroenterol Hepatol* 2002;17:849–53.
22. Ohkusa T, Okayasu I, Ogihara T, *et al*. Induction of experimental ulcerative colitis by *Fusobacterium varium* isolated from colonic mucosa of patients with ulcerative colitis. *Gut* 2003;52:79–83.
23. Dhakan DB, Maji A, Sharma AK, *et al*. The unique composition of Indian gut microbiome, gene Catalogue, and associated fecal metabolome deciphered using multi-omics approaches. *Gigascience* 2019;8:giz004.

24 Zhang W, Li J, Lu S, et al. Gut microbiota community characteristics and disease-related microorganism pattern in a population of healthy Chinese people. *Sci Rep* 2019;9:1594.

25 Brooks AW, Priya S, Blekhman R, et al. Gut microbiota diversity across ethnicities in the United States. *PLoS Biol* 2018;16:e2006842.

26 Falony G, Joossens M, Vieira-Silva S, et al. Population-Level analysis of gut microbiome variation. *Science* 2016;352:560–4.

27 Human Microbiome Project Consortium. Structure, function and diversity of the healthy human microbiome. *Nature* 2012;486:207–14.

28 Qin J, Li R, Raes J, et al. A human gut microbial gene Catalogue established by metagenomic sequencing. *Nature* 2010;464:59–65.

29 Coker OO, Nakatsu G, Dai RZ, et al. Enteric fungal microbiota dysbiosis and ecological alterations in colorectal cancer. *Gut* 2019;68:654–62.

30 Qin J, Li Y, Cai Z, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* 2012;490:55–60.

31 Qin N, Yang F, Li A, et al. Alterations of the human gut microbiome in liver cirrhosis. *Nature* 2014;513:59–64.

32 Li J, Jia H, Cai X, et al. An integrated catalog of reference genes in the human gut microbiome. *Nat Biotechnol* 2014;32:834–41.

33 Zeevi D, Korem T, Zmora N, et al. Personalized nutrition by prediction of glycemic responses. *Cell* 2015;163:1079–94.

34 Karlsson FH, Tremaroli V, Nookaew I, et al. Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* 2013;498:99–103.

35 Bäckhed F, Roswall J, Peng Y, et al. Dynamics and stabilization of the human gut microbiome during the first year of life. *Cell Host Microbe* 2015;17:690–703.

36 Pehrsson EC, Tsukayama P, Patel S, et al. Interconnected microbiomes and resistomes in low-income human habitats. *Nature* 2016;533:212–6.

37 Brito IL, Yilmaz S, Huang K, et al. Mobile genes in the human microbiome are structured from global to individual scales. *Nature* 2016;535:435–9.

38 Liu W, Zhang J, Wu C, et al. Unique features of ethnic Mongolian gut microbiome revealed by metagenomic analysis. *Sci Rep* 2016;6:34826.

39 Rampelli S, Schnorr SL, Consolandi C, et al. Metagenome sequencing of the Hadza hunter-gatherer gut microbiota. *Curr Biol* 2015;25:1682–93.

40 Smits SA, Leach J, Sonnenburg ED, et al. Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. *Science* 2017;357:802–6.

41 Rothschild D, Weissbrod O, Barkan E, et al. Environment dominates over host genetics in shaping human gut microbiota. *Nature* 2018;555:210–5.

42 Almeida A, Mitchell AL, Boland M, et al. A new genomic blueprint of the human gut microbiota. *Nature* 2019;568:499–504.

43 Nayfach S, Shi ZJ, Seshadri R, et al. New insights from uncultivated genomes of the global human gut microbiome. *Nature* 2019;568:505–10.

44 Pasolli E, Asnicar F, Manara S, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. *Cell* 2019;176:649–62.

45 Truong DT, Franzosa EA, Tickle TL, et al. MetaPhlAn2 for enhanced metagenomic taxonomic profiling. *Nat Methods* 2015;12:902–3.

46 Li D, Liu C-M, Luo R, et al. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 2015;31:1674–6.

47 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.

48 Kang DD, Froula J, Egan R, et al. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ* 2015;3:e1165.

49 Wu Y-W, Tang Y-H, Tringe SG, et al. MaxBin: an automated binning method to recover individual genomes from metagenomes using an expectation-maximization algorithm. *Microbiome* 2014;2:26.

50 Sieber CMK, Probst AJ, Sharrar A, et al. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 2018;3:836–43.

51 Parks DH, Imelfort M, Skennerton CT, et al. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 2015;25:1043–55.

52 Parks DH, Chuvochina M, Waite DW, et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 2018;36:996–1004.

53 Olm MR, Brown CT, Brooks B, et al. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *Isme J* 2017;11:2864–8.

54 Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 2014;30:1312–3.

55 Hyatt D, Chen G-L, LoCascio PF, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;11:119.

56 Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using diamond. *Nat Methods* 2015;12:59–60.

57 Jain C, Rodriguez-R LM, Phillippy AM, et al. High throughput ani analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:5114.

58 Huerta-Cepas J, Forslund K, Coelho LP, et al. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. *Mol Biol Evol* 2017;34:2115–22.

59 Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 2013;30:772–80.

60 Parks DH, Chuvochina M, Chaumeil PA, et al. Selection of representative genomes for 24,706 bacterial and archaeal species clusters provide a complete genome-based taxonomy. *bioRxiv* 2019.

61 Yachida S, Mizutani S, Shiroma H, et al. Metagenomic and metabolomic analyses reveal distinct stage-specific phenotypes of the gut microbiota in colorectal cancer. *Nat Med* 2019;25:968–76.

62 Pfeifhofer-Obermair C, Tymoszuk P, Petzer V, et al. Iron in the tumor Microenvironment-Connecting the dots. *Front Oncol* 2018;8:549.

63 Wegner M-S, Schömel N, Gruber L, et al. Udp-Glucose ceramide glucosyltransferase activates Akt, promoted proliferation, and doxorubicin resistance in breast cancer cells. *Cell Mol Life Sci* 2018;75:3393–410.

64 Kim Y-I. Folate: a magic bullet or a double edged sword for colorectal cancer prevention? *Gut* 2006;55:1387–9.

65 Mora D, Arioli S. Microbial urease in health and disease. *PLoS Pathog* 2014;10:e1004472.

66 Garcia-Carretero R, Lopez-Lomba M, Carrasco-Fernandez B, et al. Clinical Features and Outcomes of *Fusobacterium* Species Infections in a Ten-Year Follow-up. *J Crit Care Med* 2017;3:141–7.

67 Afra K, Laupland K, Leal J, et al. Incidence, risk factors, and outcomes of *Fusobacterium* species bacteremia. *BMC Infect Dis* 2013;13:264.

68 Sekizuka T, Ogasawara Y, Ohkusa T, et al. Characterization of *Fusobacterium varium* Fv113-g1 isolated from a patient with ulcerative colitis based on complete genome sequence and transcriptome analysis. *PLoS One* 2017;12:e0189319.

69 Umaña A, Sanders BE, Yoo CC, et al. Utilizing Whole *Fusobacterium* Genomes To Identify, Correct, and Characterize Potential Virulence Protein Families. *J Bacteriol* 2019;201:e00273–19.