Journal of Cancer Science and Research

ISSN: 2576-1447

OPEN ACCESS Freely available online

Research Article

Colorectal cancer biomarkers remain consistent across geography and ethnicity despite differing baseline gut microbiome compositions

Adith Reddi

Phillips Exeter Academy, New Hampshire, USA

Abstract

We focused on 4 shotgun sequenced taxonomic abundance datasets, 2 from East Asia (Japan and China) and 2 from Europe (Austria and France). Using 2 distinct classifiers, we were able to identify the specific bacteria that are important in diagnosing CRC across geography and ethnicity. It has previously been documented that gut microbiome composition varies across geography and ethnicity. This makes it difficult to determine what a "healthy" gut microbiome composition looks like since there is a large amount of innate variation. Gut microbiome composition has also been shown to vary across individuals based on disease and, as a result, specific gut microbes have been implicated in the development of major GI disease. However, without a "healthy" baseline composition to compare to, the task of diagnosing disease through the gut microbiome becomes challenging. Our findings suggest that CRC biomarkers are limited to a handful of recurring bacteria. Our results indicate that Gemella morbillorum, Peptostreptococcus stomatis, Parvimonas micra, Fusobacterium nucleatum, Clostridium hathewayi, Solobacterium moorei, an unclassified bacterium in the genus Oscillibacter, and an unclassified bacterium in the genus Parvimonas play important roles in CRC diagnosis. These bacteria act independently from other variation in the gut microbiome. We conclude that CRC diagnosis can focus on a specific subset of bacteria and is not dependent on geography and ethnicity.

Keywords: colorectal cancer; CRC; gut microbiome; machine learning; geography; ethnicity; genomics and bacteria

Introduction

Over the last decade, the human gut microbiome has grown in significance as its applications to medicine, anthropology, and evolution have become more apparent. Now often treated as an organ in of itself, the microbiome continues to increase in importance as new correlations are drawn between it and various aspects of our lives [1]. An important topic in gut microbiology is how diet affects gut microbiome composition. Rasnik K. Singh mentions the following in his study about the topic, "consumption of particular types of food produces predictable shifts in existing host bacterial genera" [2]. Near the end of his study, Singh examines the Mediterranean diet, which is highly regarded as a healthy balanced diet, and comes to the conclusion that an increase in a certain set of bacteria (Lactobacillus, Bifidobacterium, and Prevotella) caused by following the Mediterranean diet reduces the risk of obesity and inflammation [2]. Other studies have explored the topic in depth as well, including one by Connie M Weaver in which a link between prebiotics use and increased calcium absorption in adolescents was identified [3]. Realizing the importance of diet, scientists began to consider the various factors in human life that could affect one's diet, and thus gut microbiome composition. This eventually brought them to two key influencers: geography and ethnicity. An association between geography and the gut microbiome is supported by several famous studies. Most notably, a paper by Tanya Yatsuneko and others explored gut microbial differences in populations from the Amazonas of Venezuela, rural Malawi, and US metropolitan areas [4]. The result was a noticeable difference in bacterial assemblages between the US population and the other two, which remained consistent from infancy to adulthood [4]. Another study compared Japanese gut microbiomes to those from 11 other nations [5]. Japanese microbiomes were considerably different from the other nations, and the authors condensed much of this variation down to one genus: Bifidobacterium [5]. Another point worth making is that "geography" does not necessarily have to only include countries.

Correspondence to: Phillips Exeter Academy, New Hampshire, E-mail: areddi@exeter.edu

Received: 11- Apr-2022, Manuscript No. JCSR-22-6546; **Editor assigned:** 13- Apr-2022, PreQC No. JCSR-22-6546 (PQ); **Reviewed:** 26-Apr-2022, QC No. JCSR-22-6546; **Revised:** 03-May-2022, Manuscript No. JCSR-22-5646 (R); **Published:** 11-May-2022, DOI: 10.35248/2576-1447.22.7.517

Citation: Reddi A (2022), Colorectal cancer biomarkers remain consistent across geography and ethnicity despite differing baseline gut microbiome compositions J Can Sci Res 7:517.

A study by Aashish R Jha examined the differences between four Himalayan populations (Tharu, Raute, Raji, and Chepang), and found that their gut microbiome compositions greatly differed [6]. The findings emphasize the importance of considering geography in conjunction with other aspects of a region when studying the gut microbiome [6]. The findings also suggest that there is a significant amount of variation within countries, sometimes more so than between countries. This phenomenon can be better understood by moving on to our next key influencer: ethnicity [6].

Ethnicity is distinct from geography in that within a geography there can be multiple ethnicities. Between ethnicities there are a myriad of lifestyle and genetic differences that contribute to the gut microbiome. A study by Deschasaux investigated this idea and came to the conclusion that individuals living in the same city had similar gut microbiomes to others of their own ethnicity [7]. They successfully were able to classify some of these ethnicities into 3 poles: OTUs classified as Prevotella tended to be Moroccan, Turkish, or Ghanaian, OTUs classified as Bacteroides tended to be African Surinamese and South Asian Surinamese, and OTUs classified as Clostridiales tended to be Dutch [7]. A separate study, focusing specifically on the Hadza hunter-gatherers, compared their gut microbiomes to those from an urban Italian control cohort and found clear differences between the two [8]. Hadza hunter-gatherers tended to have a more diverse microbiome overall when compared to the urban Italian cohort [8]. Another study generalized some of these findings by clumping ethnicities together based on whether they followed a hunter-gatherer diet or an urban diet [9]. The findings were similar, showing that ethnicities which followed a hunter-gatherer diet tended to have a more diverse microbiome than ethnicities following an urban diet [9]. Beyond geography and ethnicity, there are many more factors which influence microbiome composition. One that has received a considerable amount of attention from the medical community in recent years is disease. A study by Le Chatelier investigated obesity and its relationship to the gut microbiome in a cohort of 123 non-obese and 169 obese Danish individuals [10]. The results showed that a lack of bacterial richness in the gut microbiome is an adequate marker for identifying obesity [10]. Le Chatelier's paper paved the way for future work in diagnosing major disease through bacterial composition. A more recent paper by Jonas Halfvarson and others explored the role of the gut microbiome in inflammatory bowel disease (IBD) over a period of time [11]. It was concluded that IBD patients' gut microbiomes fluctuated more often than healthy patients' gut microbiomes [11]. This fluctuation was measured from a base point which the study referred to as the "healthy plane" (HP). Interestingly enough, during the entire period of fluctuation, IBD microbiomes periodically visited the HP before deviating away from it [11]. These unique behaviors are all "markers" for diagnosing disease.

Colorectal cancer (CRC) is another major disease that is associated with the gut microbiome. CRC is the second most deadly cancer worldwide, with 881,000 estimated deaths in 2018 [12]. Traditionally, it has been important to identify CRC earlier rather than later because after a certain point it becomes difficult to treat [12]. Thus, there is an urgent need to improve the current tools used to diagnose and screen for CRC. Luckily, studying the gut microbiome has allowed for the innovation of novel diagnostic methods to identify CRC in the very early stages. One of the earlier studies that looked into this topic was conducted by Jun Yu and others. In the study, Yu's team built a classifier that was trained on healthy controls and CRC patients from China and was validated on a Danish cohort [13]. As a result of the study, associations between the bacteria Fusobacterium nucleatum and Peptostreptococcus stomatis and CRC were reconfirmed, while new bacteria such as Parvimonas micra and Solobacterium moorei were shown to have links to CRC as well [13]. Future studies expanded on this idea of finding biomarkers for CRC to include even more geographies. In a study done by Andrew Maltez Thomas, multiple highly predictive CRC biomarkers were identified across various geographies [14]. Thomas' findings suggested that these biomarkers could be generalized to encapsulate multiple geographies, which would then increase the validity of CRC classifiers when predicting across geographies [14]. This notion of global CRC biomarkers was also been investigated separately by Jakob Wirbel and others [15]. Wirbel managed to identify a core set of 29 highly influential bacteria which seemed to show up regardless of geography, further supporting the notion of global CRC biomarkers [15]. The important question now is "how do CRC classifiers compare to the current CRC screening tests?" Georg Zeller conducted a study where he compared classifiers that were trained to diagnose CRC from gut microbiome data to the current standard fecal occult blood test (FOBT) [16]. He found that the classifiers performed at the same accuracy as the FOBT, and when the two were combined the sensitivity of the diagnosis increased by > 45% [16]. Identifying CRC biomarkers in the gut microbiome might improve non-invasive methods to screen for CRC. It is then critical that the validity and generalizability of these biomarkers is investigated further. In this study, we perform an analysis using multiple classifiers and feature selection techniques to replicate the identification of well-known CRC biomarkers, as well as to show how they remain consistent across geographies in East Asia and Europe.

Methods

Data acquisition

In order to train proper machine learning models, one first needs to have access to high quality data. To acquire good data, we started by doing a survey of the literature on CRC on PubMed. We used keywords such as "gut microbiome," "CRC," "colorectal cancer," "shotgun," and "whole genome sequencing." We recorded each relevant paper in a document. We eventually narrowed down our potential data sources to three papers: one by Wirbel, one by Thomas, and one by Zeller. These meta-analyses provided links to smaller studies with high-quality data. After aggregating the relevant data, in our initial dataset we had an Italian cohort (n=80), a Chinese cohort (n=128), an American cohort (n=110), an Austrian cohort (n=154), a French cohort (n=156), a German cohort (n=43), and a Japanese cohort (n=80). For the final analysis, we only used the Japanese cohort, Chinese cohort, Austrian cohort, and the French cohort. This is because we only had access to 2 East Asian datasets, and so to balance out the number of samples we paired them up with 2 European datasets, instead of all 4. The reason we specifically chose the French and Austrian datasets was because both the Italian and German cohorts were significantly smaller than the French and Austrian ones.

It is also important to note that the French and German cohorts initially came together, but we programmatically separated them later on.

Data preprocessing

We acquired all of the datasets, except for the Japanese one, from curatedMetagenomicData. The Japanese dataset was not available in the curatedMetagenomicData package, and so to process it we used a pipeline which has the ability to scrape any desired dataset from the NCBI website and process it through the MetaPhlAn2 algorithm.

Once all of the datasets were downloaded, we had to place them in a standardized format so that we could process them through our machine learning algorithms. We settled on assigning each sample a file which contained all of the taxonomic abundance information for that sample.

As a final step, we removed all of the non-CRC and non-control samples from the datasets.

Our data processing workflow: Downloaded files from curatedMetagenomicData/MetaPhlAn2 pipeline → Reformatted files so that each sample had its own taxonomic abundance file → Processed samples and arranged them into a Pandas DataFrame → Removed non-CRC and non-control samples from the DataFrame

Showing that CRC biomarkers exist

To show that CRC biomarkers exist, we used several data visualization techniques coupled with multiple classifier trials, each trial being performed with 2 distinct classifiers. Here are the exact steps we took:

(1) Created scree plots to determine which principal components (PC) were sufficient to capture the most relevant variation in a principal component analysis.

(2) Performed principal component analysis (PCA) where we assigned unique colors to the CRC and healthy control samples (target names were "CRC" and "control").

(3) Performed t-distributed stochastic neighbor embedding (t-SNE) where we assigned unique colors to the CRC and healthy control samples (target names were "CRC" and "control").

(4) Created a logistic regression classifier (lambda=0.1) to diagnose CRC based on the data (target names were "CRC" and "control"). Area under the receiver operating characteristic (AUROC) was our metric.

(5) Created a random forest classifier to diagnose CRC based on the data (target names were "CRC" and "control"). AUROC was our metric. Since random forest classifiers are innately random, they will produce a slightly different AUROC each time you use them. To counteract this, each time we used the random forest classifier we ran the prediction 10 times and recorded the range.

We mainly used PCA and t-SNE as a way to identify strong signals that could be easily visualized. They were used as surface-level analyses to see how strong the signal in the data was.

The classifiers acted as a deeper set of analyses since they are more sensitive and tend to pick up on subtler signals. Using the classifiers, we did the following analyses on the 4 datasets (Japanese, Chinese, Austrian, French):

(1) Performed a "round-robin" set of tests where we trained on each dataset and tested on every other dataset.

(2) Cross-validated within each of the datasets (train set=2/3 of total, test set=1/3 of total). Since cross validation selects a random portion of the dataset, each cross-validation trial was run 10 times and we created a range of the AUROCs.

We did a similar set of analyses on a European aggregate dataset (Austrian, French) and an East Asian aggregate dataset (Japanese, Chinese).

Overall, if the classifiers produced high AUROCs, that would confirm the existence of CRC biomarkers.

Showing that CRC biomarkers remain consistent across geographies

To show that CRC biomarkers remain consistent across geographies, we compared the AUROCs produced by the classifier trials in which the classifier was trained on data from one continent and used to predict on data from another continent. If the AUROCs were high and similar across the board, then the biomarkers had to be consistent across geographies.

Identifying the specific CRC biomarkers that remain consistent

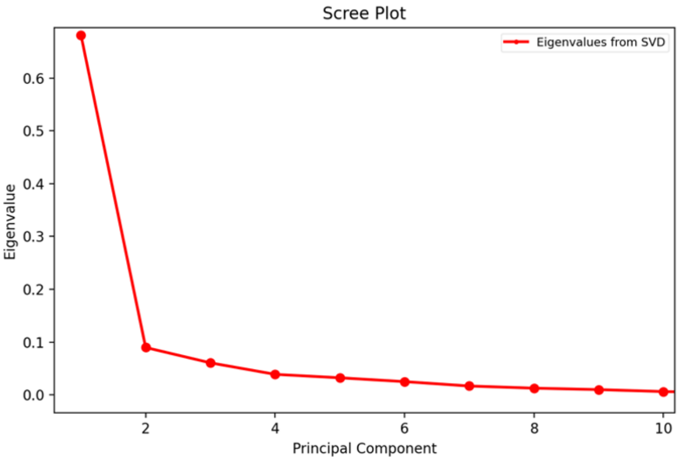
To identify the specific CRC biomarkers that remain consistent, we created a "weights table." The rows of this table were built from a superset of the top 20 most influential bacteria in CRC diagnosis in each of the datasets. The columns of the table were the names of all the geographies we considered. The cells of this table had a numerical weight representing the influence that a given bacterium had on CRC diagnosis in a given geography.

To create this table, we first created a "select from model" (sfm) model using the scikit-learn library in python. Using sfm, we were easily able to identify the top 20 features in each of our classifiers. Since our random forest classifier performed better than our logistic regression classifier overall, we decided to use the top 20 features from that model. Since the random forest classifier is random, we wrote a script that ran the model 10 times and then took a superset of the top 20 features from each run. The weight placed on a given feature by a geography (the values that were going into the cells of the table) was calculated by averaging the weights placed on the feature by each sample in that geography. We then wrote a script to construct a table which used the bacterial superset as the rows and the geographies as the columns, with all of the corresponding weights as the cells.

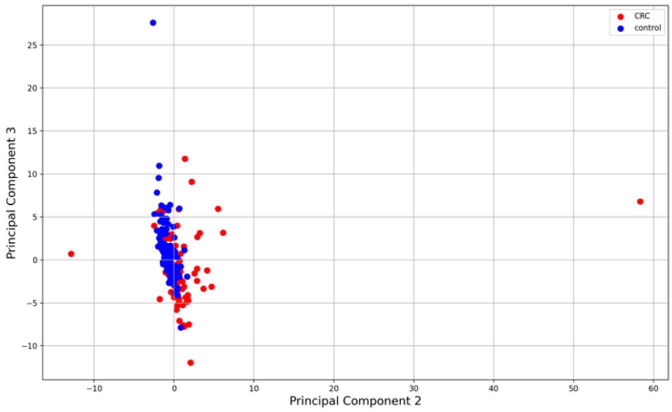
Results

Biomarkers for CRC exist

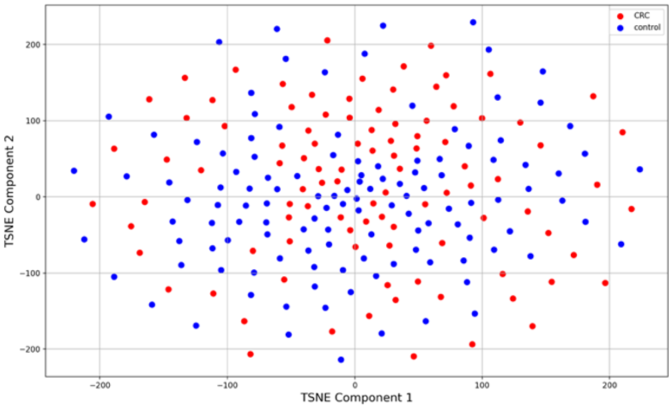
Scree: Europe (Austria and France)



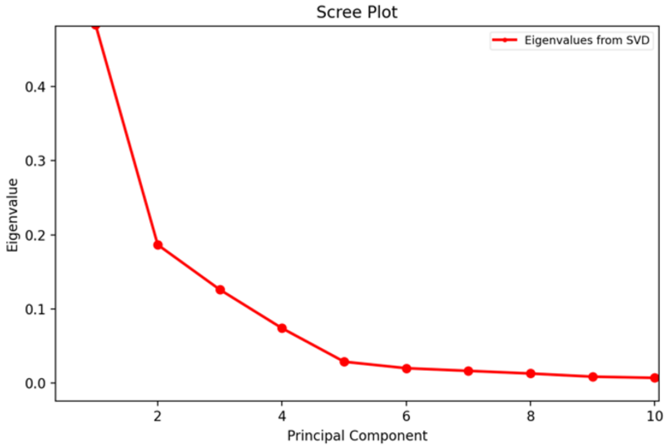
PCA: CRC vs control in Europe (Austria and France)



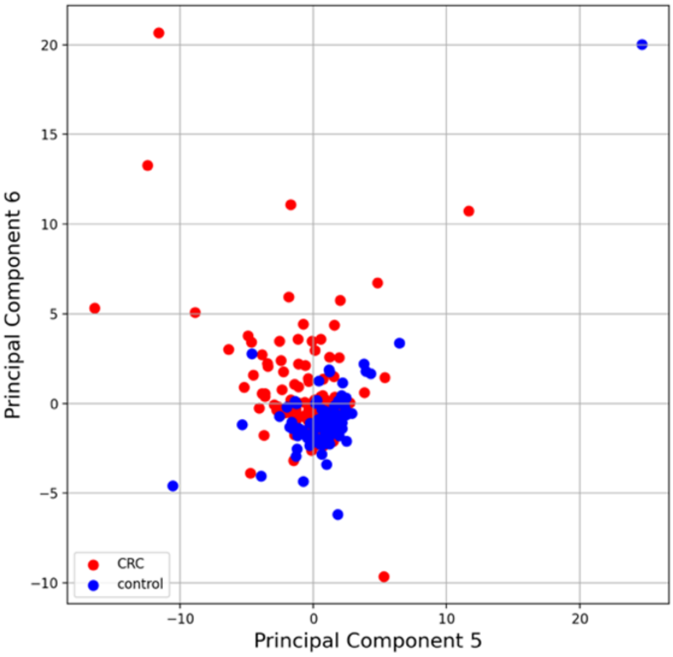
TSNE: CRC vs control in Europe (Austria and France)



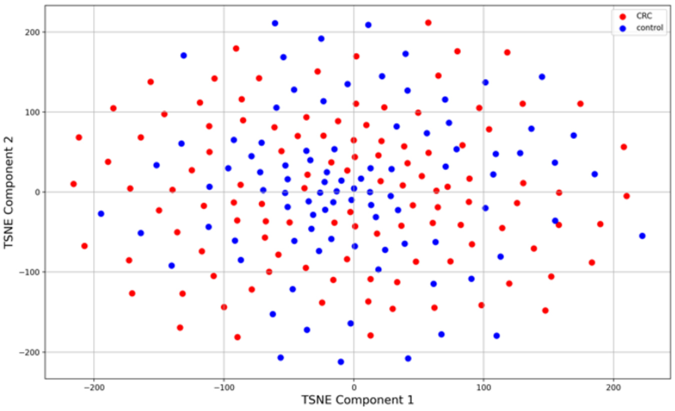
Scree: East Asia (Japan and China)



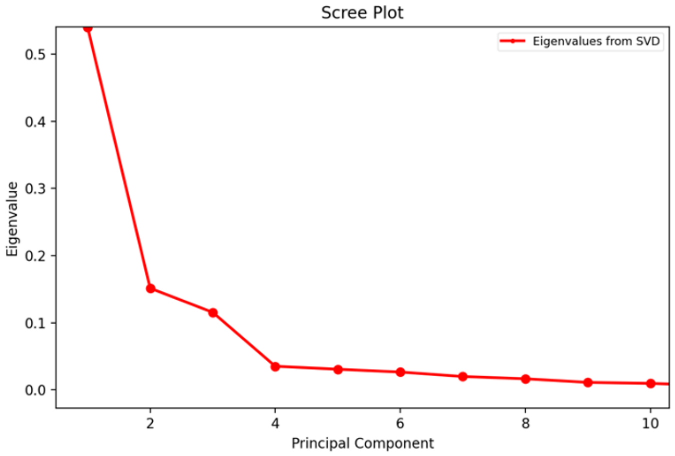
PCA: CRC vs control in East Asia (Japan and China)



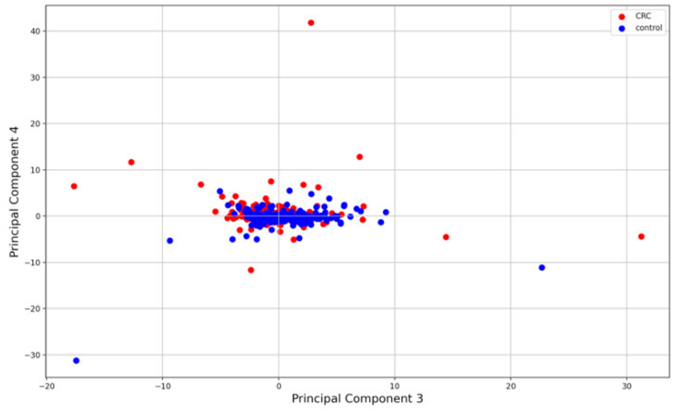
TSNE: CRC vs control in East Asia (Japan and China)



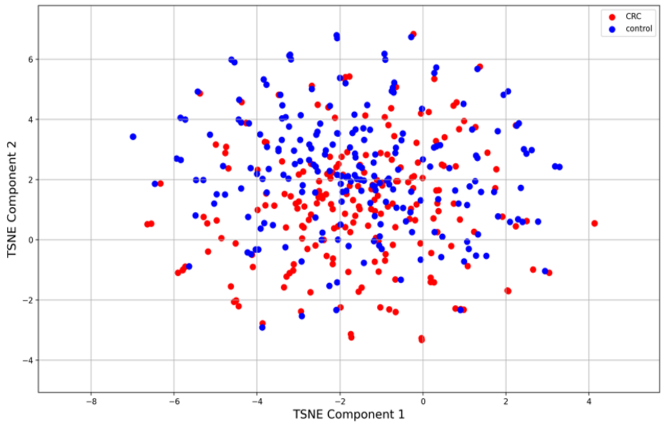
Scree: Europe (Austria and France) and East Asia (Japan and China)



PCA: CRC vs control in Europe (Austria and France) and East Asia (Japan and China)



TSNE: CRC vs control in Europe (Austria and France) and East Asia (Japan and China)



The above graphs indicate that there is no strong signal that helps differentiate between CRC and control samples in the various datasets. We conclude that the signal is too subtle to be captured by simple data visualization techniques like PCA and t-SNE.

Training set	Test set	AUROC (logistic regression)	AUROC (random forest)
Random 2/3 of Chinese	Remaining 1/3 of Chinese	0.71-0.89	0.76-0.89
Random 2/3 of Japanese	Remaining 1/3 of Japanese	0.64-0.81	0.68-0.88
Random 2/3 of Austrian	Remaining 1/3 of Austrian	0.79-0.86	0.83-1.0
Random 2/3 of French	Remaining 1/3 of French	0.8-0.92	0.73-0.92
Random 2/3 of European (French, Austrian)	Remaining 1/3 of European (French, Austrian)	0.79-0.87	0.8-0.94
Random 2/3 of East Asian (Chinese, Japanese)	Remaining 1/3 of East Asian (Chinese, Japanese)	0.69-0.78	0.75-0.88
Random 2/3 of European and East Asian	Remaining 1/3 of European and East Asian	0.72-0.81	0.79-0.88

Above are the cross-validation trials we performed to show that CRC biomarkers exist. Unlike PCA and t-SNE, the classifiers were able to pick up a signal. The Japanese cross-validation trial produced a slightly worse AUROC range than the other trials because it was a smaller dataset (n=80). Regardless, the high AUROCs across the board provide sufficient support for the existence of CRC biomarkers.

Biomarkers for CRC remain consistent across geographies

Training set	Test set	AUROC (logistic regression)	AUROC (random forest)
European (French, Austrian)	Chinese	0.78	0.83-0.87
European (French, Austrian)	Japanese	0.7	0.78-0.84
East Asian (Chinese, Japanese)	Austrian	0.66	0.81-0.87
East Asian (Chinese, Japanese)	French	0.8	0.74-0.8

Above are the trials we performed to show that CRC biomarkers are consistent across geographies. As one can see, regardless of which geographies are being trained on and which are being tested on, the AUROCs are roughly the same (and also high). This means that there is some signal in the training dataset that is also present in the testing dataset that is predictive of CRC. These results imply that there are specific CRC predictive bacteria which are present (or not present) in CRC patients in both European and East Asian geographies.

There are a handful of specific CRC biomarkers that remain consistent across geographies

Please refer to the supplementary table. The table shows the weight that various bacteria have on predicting CRC in geographies in East Asia and Europe. The bacteria that appear in every geography have been highlighted. These represent predictive CRC biomarkers that remain consistent across geographies.

Discussion

Our findings imply that CRC biomarkers remain consistent across geographies in Eastern Asia and Europe. We trained our models on countries from one of those regions and predicted on countries from the other. Similar and high AUROCs across the board, as well as the weights table, act as sufficient evidence.

The specific biomarkers that were found to remain consistent across all geographies were Gemella morbillorum, Peptostreptococcus stomatis, Parvimonas micra, Fusobacterium nucleatum, Clostridium hathewayi, Solobacterium moorei, an unclassified bacterium in the genus Oscillibacter, and an unclassified bacterium in the genus Parvimonas. Several of these bacteria have already been implicated in the development of CRC [17]. Most notably Fusobacterium nucleatum, which is widely regarded as one of the most predictive CRC biomarkers [18]. At this point, several questions still need to be answered concerning where some of these bacteria come from, what their relationships to other diseases are, and what role they play in the human diet.

The classifiers identified multiple non-infectious bacteria found in both the oral and gut microbiomes as candidate CRC biomarkers. The genus Gemella has frequently been cited as having bacteria associated with CRC [19]. However, the species Gemella morbillorum has rarely been a direct cause of disease in humans (Bench 17). It is typically found in the oropharyngeal area and is reported to be one of the most common bacteria present in teeth with cysts that do not resolve after several repeated root canal treatments (Bench 17). Besides an association with CRC, Gemella morbillorum does not seem to be a culprit in many other diseases. Another bacterium that lives in the oral cavity is Solobacterium moorei, which on its own, much like Gemella morbillorum, does not have any harmful effect on the body [20]. However, it is a key contributor to halitosis (commonly known as bad breath) [20]. Both of these bacteria are basically harmless, yet they have such a strong association with CRC. The only trait that these two bacteria have in common, besides the fact that they are gram positive and anaerobic, is that they reside in the oral cavity, and contribute in some way to oral hygiene (teeth with cysts and bad breath).

On the more infectious end of the spectrum of bacteria found in the oral and gut microbiomes, there is the genus Peptostreptococcus, which is known to be clinically significant [21]. Bacteria that are members of the genus Peptostreptococcus tend to reside in the mouth, skin, gastrointestinal tract, vaginal tract, and urinary tract [21]. Under immunosuppressed or traumatic conditions, members of Peptostreptococcus are known to become pathogenic, causing abscesses to form in the brain, liver, lungs, and breasts, as well as leading to soft tissue necrosis [21]. A likely reason for the genus Peptostreptococcus being an effective biomarker for CRC could be because cancer tends to weaken the immune system, which is one of the factors that causes Peptostreptococcus to become harmful and change in abundance [22]. The bacterium Parvimonas micra is similar to bacteria from the genus Peptostreptococcus in that it also causes an infection: chronic periodontitis [23]. It can be found in dental plaque, which is a hotbed of oral bacteria, some of which are pathogenic [23]. There is not much to say about the unclassified species in the genus Parvimonas since not much research has been done into them yet. One final bacterium in this category to consider is Fusobacterium nucleatum [24]. Outside of associations with CRC, Fusobacterium nucleatum has been cited as a key contributor to periodontal plaque and disease due to its ability to aggregate with other bacterial species in the oral microbiome [24]. In terms of CRC, Fusobacterium nucleatum is said to accelerate cancer development by creating a pro-flammatory environment which promotes tumor growth, which as a direct result causes carcinogenesis [24]. However, in addition to having strong ties to CRC, it is said that the disease caused by Fusobacterium nucleatum, periodontal disease, also correlates with preterm birth [25]. This is particularly interesting since periodontal disease has also been shown to correlate with CRC [26]. In addition to this, excessive colonization of the vaginal microbiome by Fusobacterium nucleatum is said to lead to preterm births, much like excessive colonization of the gut/oral microbiome by Fusobacterium nucleatum is said to lead to CRC [25]. Perhaps there is a link between these two conditions that might be worth exploring.

The findings discussed so far, which are all regarding the oral microbiome, line up with a paper authored by Burkhardt Flemer, where it was found that the oral microbiota in CRC is highly predictive [27]. Most of the bacteria we have discussed so far were explicitly mentioned in that paper as well.

From this point on, we would like to discuss all of the microbes which are found frequently in the gut microbiome but not as often in the oral microbiome that appeared in our list of biomarkers. The Clostridium genus is a set of highly infectious bacteria, some of which play a role in diseases as severe as botulism and tetanus [28]. Although there is not much information about Clostridium hathewayi in particular, the Clostridium genus as a whole is known to have the ability to selectively target cancer cells, and some strains actually replicate within solid tumors [28]. Due to this property, there is some discussion around using specific strains of Clostridium to treat cancer [28]. This property might also provide a scientific explanation as to how Clostridium hathewayi contributes to CRC. The next set of bacteria to consider are in the Oscillibacter genus. Even though this genus does not have many previously cited associations with CRC, one species in the genus is linked to Crohn's disease, which is most definitely relevant to CRC [29]. The bacterium in question is Oscillibacter valericigenes. It can most commonly be found in the alimentary canal of Japanese Corbicula clams [28]. Perhaps this is can be linked back to diet. This is peculiar because China is the world's largest seafood consumer, and they also happened to place the lowest weight on the Oscillibacter genus when predicting CRC [30]. We speculate that the abundance of the Oscillibacter genus varies based on seafood consumption. If this is true, then the ubiquitous consumption of seafood in China will keep the abundances of the genus roughly the same in control and CRC patients, explaining why the marker was less effective in China than in the other geographies.

The final microbe to consider, which was in our top 20 biomarkers list but did not show up in every geography, is the Dasheen mosaic virus. The Dasheen mosaic virus typically infects the plant C. esculenta (commonly known as Taro), causing harmful symptoms and reducing crop yield [31]. This microbe appeared in the top 20 features for the Austrian dataset, but it did not count as a generalizable CRC biomarker since it was only relevant in the Austrian dataset. Regardless, it is strange that it appeared there in the first place. It is also strange how in the Austrian cohort the virus had 10x more weight placed on it than is typically placed on a top 20 microbe in that cohort. In addition, after looking at the raw abundance numbers, the Dasheen mosaic virus is 10-100x more abundant in the Austrian dataset than it is in the other datasets. We speculate that diets involving a high consumption of infected Taro will cause this virus to appear in the gut. There has not been a major sighting of the Dasheen mosaic virus in Austria in recent times, however, it is worth noting that the Dasheen mosaic virus has been spotted in Bosnia and Herzegovina, which is not far from Austria and likely exports crops to Austria as well [32]. The virus is negatively correlated with CRC, so it could be a candidate biomarker in populations where it is in high abundance and where a decrease in abundance will be noticeable. The global predictability of the Dasheen mosaic virus in CRC diagnosis should be tested in a future study.

In conclusion, we developed a workflow to show that CRC biomarkers exist, are consistent across East Asian and European geographies, and can be identified. We have also discussed how the key biomarkers that appear across all of the geographies we examined can be linked to other diseases and some cultural/ societal conditions such as diet. In future studies, the validity of currently known CRC biomarkers can be further refined so that classifiers built to diagnose CRC will outperform the current standards for screening.

Acknowledgements

I would like to thank Dr. David Shin (biotech expert), Mr. Eric Proffitt (Machine Learning Scientist at Prime Discoveries), and Dr. Arun Manoharan (Co-founder and Chief Science Officer at Prime Discoveries) for their support and insightful feedback regarding this report. The pipeline I used to process the Japanese dataset was provided by Prime Discoveries.

Data availability statement

Below are the project IDs for the datasets used in this study:

PRJEB10878 [13]

PRJEB12449 [33]

ERP008729 [34]

ERP005534 [16]

SRP136711 [14]

DRA006684

The code for our analyses can be found here: https://github.com/recko10/Identifying-CRC-biomarkers

References

1. Baquero F, Nombela C (2012) The microbiome as a human organ. Clin Microbiol Infect 18: 2-4.
2. Singh RK, Chang HW, Yan D, Lee KM, Ucmak D, et al. (2017) Influence of diet on the gut microbiome and implications for human health. J Transl Med 15: 73.
3. Weaver CM (2015) Diet, gut microbiome, and bone health. Curr Osteoporos Rep 13: 125-130.
4. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, et al. (2012) Human gut microbiome viewed across age and geography. Nature 486: 222-227.
5. Nishijima S, Suda W, Oshima K, Kim SW, Hirose Y, et al. (2016) The gut microbiome of healthy Japanese and its microbial and functional uniqueness. DNA Res 23: 125-133.
6. Jha AR, Davenport ER, Gautam Y, Bhandari D, Tandukar S, et al. (2018) Gut microbiome transition across a lifestyle gradient in Himalaya. PLoS Biol 16: e2005396.
7. Deschasaux M, Bouter KE, Prodan A, Levin E, Groen AK, et al. (2018) Depicting the composition of gut microbiota in a population with varied ethnic origins but shared geography. Nat Med 24: 1526-1531.
8. Schnorr SL, Candela M, Rampelli S, Centanni M, Consolandi C, et al. (2014) Gut microbiome of the Hadza hunter-gatherers. Nat Commun 5: 3654.
9. Gupta VK, Paul S, Dutta C (2017) Geography, ethnicity or subsistence-specific variations in human microbiome composition and diversity. Front Microbiol 8: 1162.

10. Le Chatelier E, Nielsen T, Qin J, Prifti E, Hildebrand F, et al. (2013) Richness of human gut microbiome correlates with metabolic markers. Nature 500: 541-546.

11. Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, et al. (2017) Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol 2: 17004.

12. Rawla P, Sunkara T, Barsouk A (2019) Epidemiology of colorectal cancer: incidence, mortality, survival, and risk factors. Prz Gastroenterol 14: 89-103.

13. Yu J, Feng Q, Wong QH, Zhang D, Liang Q, et al. (2017) Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut 66: 70-78.

14. Thomas AM, Manghi P, Asnicar F, Pasolli E, Armanini F, et al. (2019) Metagenomic analysis of colorectal cancer datasets identifies cross-cohort microbial diagnostic signatures and a link with choline degradation. Nat Med 25: 667-678.

15. Wirbel J, Pyl PT, Kartal E, Zych K, Kashani A, et al. (2019) Meta-analysis of fecal metagenomes reveals global microbial signatures that are specific for colorectal cancer. Nat Med 25: 679-689.

16. Zeller G, Tap J, Voigt AY, Sunagawa S, Kultima JR, et al. (2014) Potential of fecal microbiota for early-stage detection of colorectal cancer. Mol Syst Biol 10: 766.

17. Wong SH, Yu J (2019) Gut microbiota in colorectal cancer: mechanisms of action and clinical applications. Nat Rev 16: 690-704.

18. Yang Z, Ji G (2019) Fusobacterium Nucleatum-positive colorectal cancer. Oncol Lett 18: 975-982.

19. Ternes D, Karta J, Tsenkova M, Wilmes P, Haan S, et al. (2020) Microbiome in colorectal cancer: how to get from meta-omics to mechanism? Trends Miscrobiol 28: 401-423.

20. Pedersen RM, Holt HM, Justesen US (2011) Solobacterium Moorei Bacteremia: identification, antimicrobial susceptibility, and clinical characteristics. J Clin Microbiol 49: 2766-2768.

21. Könönen E, Bryk A, Niemi P, Kanervo-Nordström A (2007) Antimicrobial susceptibilities of peptostreptococcus anaerobius and the newly described peptostreptococcus stomatis isolated from various human sources. Antimicrob Agents Chemother 51: 2205-2207.

22. Axelrad JE, Lichtiger S, Yajnik V (2016) Inflammatory bowel disease and cancer: the role of inflammation, immunosuppression, and cancer treatment. World J Gastroenterol 22: 4794-4801.

23. Yoo LJH, Zulkifli MD, O'Connor M, Waldron R (2019) Parvimonas micra spondylodiscitis with psoas abscess. BMJ Case Rep 12: e232040.

24. Brennan CA, Garrett WS (2019) Fusobacterium nucleatum - symbiont, opportunist and oncobacterium. Nat Rev Microbiol 17: 156-166.

25. Han YW, Redline RW, Li M, Yin L, Hill GB, et al. (2004) Fusobacterium nucleatum induces premature and term stillbirths in pregnant mice: implication of oral bacteria in preterm birth. Infect Immun 72: 2272-2279.

26. Kim GW, Kim YS, Lee SH, Park SG, Kim DH, et al. (2019) Periodontitis is associated with an increased risk for proximal colorectal neoplasms. Sci Rep 9: 7528.

27. Flemer B, Warren RD, Barrett MP, Cisek K, Das A, et al. (2018) The oral microbiota in colorectal cancer is distinctive and predictive. Gut 67: 1454-1463.

28. Bush LM, Vazquez-Pertejo MT (2019) Overview of clostridial infections. Merck & Co., Inc., Kenilworth, NJ, USA.

29. Man SM, Kaakoush NO, Mitchell HM (2011) The role of bacteria and pattern-recognition receptors in Crohn's disease. Nat Rev Gastroenterol Hepatol 8: 152-168.

30. Guillen J, Natale F, Carvalho N, Casey J, Hofherr J, et al. (2018) Global seafood consumption footprint. Ambio 48: 111-122.

31. Babu B, Hegde V, Makeshkumar T, Jeeva ML (2011) Detection and identification of Dasheen mosaic virus infecting colocasia esculenta in India. Indian J Virol 22: 59-62.

32. Grausgruber-Gröger S, Richter S, Salapura JM, Joši□ DK, Trkulja V, et al. (2016) First report of Dasheen mosaic virus in Zantedeschia in Bosnia and Herzegovina. New Dis Rep 33: 13.

33. Vogtmann E, Hua X, Zeller G, Sunagawa S, Voigt AY, et al. (2016) Colorectal cancer and the human gut microbiome: reproducibility with whole-genome shotgun sequencing. PloS One 11: p.e0155362.

34. Feng Q, Liang S, Jia H, Stadlmayr A, Tang L, et al. (2015) Gut microbiome development along the colorectal adenoma–carcinoma sequence. Nat Commun 6: 6528.