Are we any closer to screening for colorectal cancer using microbial markers?
A critical review
Lubomir Starya, Kristyna Mezerovab, Pavel Skalickyb, Pavel Zborila, Vladislav Raclavskyc

The role of gut microbiota in the development of sporadic colorectal cancer (CRC) is supported by a number of studies, however, the conclusiveness of published metagenomic studies is questioned by technical pitfalls and limited by small cohort sizes. In this review, we evaluate the current knowledge critically and outline practical solutions. We also list candidate CRC risk markers that are – in our opinion – well supported by available data and thus deserve clinical validation. Last but not least, we summarise available knowledge useful for improving care for patients immediately.

Key words: colorectal cancer, microbiome, culturomics, screening markers

INTRODUCTION

Sporadic colorectal cancer (CRC) represents 80-85% of all CRC cases. It originates from healthy mucosal cells that develop mutations in tumour-suppressor genes (APC, CTNNB1 and p53 in particular) and oncogenes (KRAS in particular) (ref.1). Gut microbiome and diet are widely accepted as playing important roles in the process of mutagenesis and provide an attractive target for CRC prevention. Whereas definition and quantification of dietary habits are not always easy, identification and quantitation of gut microbes should be feasible. This way, individuals at increased risk of CRC could be identified and offered existing options of secondary prevention. Here we review critically the available data on potential microbial markers of CRC with special emphasis on perspectives of routine detection including the pitfalls of available technologies.

HOW MAY BACTERIA BE INVOLVED IN CRC DEVELOPMENT?

The known CRC risk factors potentially related to microbial action can be summarised as follows:

Inflammatory bowel disease (IBD) – important role of microbes in the progression of IBD is widely accepted and patients suffering from either of IBD entities (Crohn’s disease and ulcerative colitis) are at high risk of developing CRC (ref.2).4.

Chronic inflammation – is proved to participate in cancer development; several bacterial species that induce inflammation are suspected to facilitate CRC development1-4.

Genotoxins – several bacterial species also produce toxins that can damage genes in colon epithelial cells by different mechanisms7-8. Recently, production of carcinogenic acetaldehyde from ethanol by anaerobic bacteria has been described which most probably explains the dose-dependent increased incidence of CRC in drinkers9,10.

Immune response modulation – both negative11, and positive12 influence of different microbes on the anti-tumour immunity has been described.

Metabolism – among others, bacteria convert bile acids into carcinogenic deoxycholic acid, particularly if combined with a high-fat diet13. By contrast, with high-fibre diet, butyrate is generated by bacterial action, which most probably protects from CRC through histone deacetylase inhibition14,15.

CHALLENGING COMPLEXITY

Detailed studies on the mechanisms outlined above should enable us to identify the pro-and anti-oncogenic actions of particular microbial species and summarize them into a quantitative CRC risk in individual patients. Unfortunately, the complexity of microbiome action turns this effort into a real challenge. Any microbe can participate in more than one pro- or anti-oncogenic processes, which is also true for products of microbial metabolism. Furthermore, the intensity of such influence will always depend upon the relative abundance of each species and upon the real spatial exposure of epithelial cells to the microbial bodies or toxins or metabolites. Also, human tissue processes will influence each other in a complex fashion, e.g. chronic inflammation will always modulate immune response. All these factors make any CRC risk prediction synthesis based on fragmented analytical knowledge from in vitro studies hardly possible.

Therefore, any pro- or anti-oncogenic microbial action needs to be carefully verified in clinical settings, even if
confirmed in vitro previously. Unfortunately, such studies are uneasy to be performed prospectively. Instead, baseline data in CRC-affected individuals versus matched controls are collected typically. Although such approach does not enable us to confirm causative relationship between microbe or toxin and CRC, it can still be extremely useful to identify potential markers of CRC risk. The possible causative role of these markers in carcinogenesis can then be confirmed or refuted during future use of such markers in CRC screening. Dozens of such studies have been performed and revealed numerous associations. Unfortunately, many of these associations either were not confirmed by others later or were even refuted. For brief overview, we therefore summarise only those species as associations that have been reported by at least two independent research groups (Table 1).

### Table 1. Bacterial species positively or negatively associated with colorectal cancer.

<table>
<thead>
<tr>
<th>Species</th>
<th>Positive CRC association</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fusobacterium nucleatum</td>
<td></td>
<td>11,16</td>
</tr>
<tr>
<td>Colibactin-producing Escherichia coli</td>
<td></td>
<td>7,17</td>
</tr>
<tr>
<td>Enterotoxigenic Bacteroides fragilis (ETBF)</td>
<td></td>
<td>8,18</td>
</tr>
<tr>
<td>Enterococcus faecalis</td>
<td></td>
<td>19,20</td>
</tr>
<tr>
<td>Streptococcus galaloyticus</td>
<td></td>
<td>21-23</td>
</tr>
<tr>
<td>(Streptococcus bovis biotype I)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Porphyromonas sp.</td>
<td></td>
<td>24-26</td>
</tr>
<tr>
<td>Salmonella sp.</td>
<td></td>
<td>27,28</td>
</tr>
<tr>
<td>Prevotella sp.</td>
<td></td>
<td>29,30</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>Negative association with CRC</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Faecalibacterium sp.</td>
<td></td>
<td>26,31</td>
</tr>
<tr>
<td>Bifidobacterium sp.</td>
<td></td>
<td>26,32</td>
</tr>
</tbody>
</table>

Genotoxins from E. coli and B. fragilis have also been linked to CRC, however, not all strains of the particular species produce a particular toxin. Thus, it may be necessary to detect toxin production for screening purposes as well. Furthermore, genomic islands coding for toxin production are shared between species; unfortunately, our knowledge on their prevalence in particular species is still limited (Table 2).

### Table 2. Carcinogenic toxins present in enterobacterial intestinal flora.

<table>
<thead>
<tr>
<th>Species</th>
<th>colibactin (pkS)</th>
<th>References</th>
<th>cytotoxic necrotizing factor (cnf1, cnf2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citrobacter freundii</td>
<td>33</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Citrobacter koseri</td>
<td>33</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Enterobacter aerogenes</td>
<td>33</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Escherichia coli</td>
<td>33</td>
<td>34,35</td>
<td>34,35</td>
</tr>
<tr>
<td>Klebsiella oxytoca</td>
<td>33</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
<tr>
<td>Klebsiella pneumoniae</td>
<td>33</td>
<td>n.a.</td>
<td>36,37</td>
</tr>
<tr>
<td>Proteus sp.</td>
<td>38</td>
<td>n.a.</td>
<td>n.a.</td>
</tr>
</tbody>
</table>

n.a. – data not available

### TECHNICAL LIMITATIONS AND PITFALLS

Although particular bacterial species (e.g. S. gallolyticus, formerly S. bovis biotype I) and genotoxins (e.g. colibactin, B. fragilis enterotoxin) have been suspected for their link to CRC based on results of conventional culture techniques earlier, a real boom of studies that revealed lots of other candidates came with the advent of high-throughput next-generation sequencing technology. This technology, however, is still rather expensive and typically yields papers with underpowered cohort sizes. Not surprisingly, the described associations are often not reproduced in follow up studies. In addition, the results of sequencing have been found to be significantly affected by the site and way of sample collection and processing. E.g. Chen et al. found Lactobacillales enriched in CRC tissue, Fusobacterium, Porphyromonas, Peptostreptococcus, Gemella, Mogibacterium and Klebsiella enriched in CRC mucosal adherent flora, and Erysipelotrichaceae, Prevotellaceae and Coriobacteriaceae enriched in the lumen of CRC patients. The choice of DNA extraction technique and primers has also been demonstrated to influence the results of sequencing profoundly, with Veillonellaceae comprising 40-50% of the total bacteria detected following extraction by one method, whereas being a minor component (<5%) in samples extracted using another method. Similarly, severe discrepancies were found when using different primer sets for the same 16S rRNA gene. Limited reproducibility was also observed when results of 16S rRNA sequencing and whole-genome shotgun sequencing were compared on the same set of samples. Furthermore, the issue of reference sequence database quality is also central to these type of studies because with the current suboptimal database coverage, up to 80% of gene sequences obtained are left unannotated.

All of the above mentioned difficulties turn the focus of microbiome research back to a more balanced approach that would include the old-fashioned culture and appropriate verification of the associations suggested by sequencing studies. In fact, sequencing and culture are not competitors here; they rather complement and support each other. Successful culture followed by whole-genome sequencing of those species underrepresented in sequence databases are the essential prerequisites for increasing the quality of these databases. Similarly, coverage of a broader...
species spectrum by next-generation sequencing does encourage improvements in conventional culture techniques to achieve cultivation of more species present in the gut. Actually, culture is performing far better than generally believed. Thus, ongoing culture studies are very useful and critical both to verify findings of sequencing studies, and to improve their depth because current metagenome sequencing typically reveals the relative abundance of bacterial phyla which encompass a staggering range of diversity. If this criterion were used to characterize animal communities, an aviary of 100 birds and 25 snails would be considered identical to an aquarium with 8 fish and 2 squid, because each has four times as many vertebrates as molluscs. Although culture is still comparatively more laborious and time-consuming, MALDI-TOF MS has brought about a breakthrough in routine species identification by reducing the time to bacterial identification by c. 55-fold as compared to conventional techniques. It also performs excellently in routine identification of anaerobic bacteria and its running costs are competitive even if compared to single-gene sequencing. When focusing on microbiome, many authors also typically limit the approach to bacterial species simply because of targeting bacterial sequences. However, our gut also harbours many fungal species that have been studied in relation to health and disease, including CRC, recently. Conventional culture followed by MALDI-TOF MS identification is able to cover both bacteria and fungi in one assay.

OTHER IMPORTANT CONSIDERATIONS

In addition to microbes, which surely play a rather important role in health and disease, another strong player is diet. Actually, both appear to be in fact interconnected: mice supplied with a microbiome from not-obese human, lost weight in a carefully controlled experiment, but this effect was diet-dependent. Also, different geographical locations and different racial/ethnic groups were shown to influence the gut microbiome. Such differences can be plausibly linked both to different diet/lifestyle and genetic background. Obviously, microbiome associations with cancer may also differ across many more host factors, including sex, age, smoking, alcohol consumption, physical inactivity, etc. Some factors on the side of the microbiome may also be variable and important, e.g. colibactin produced by some E. coli strains was considered to play a role in CRC development, however, its production is actually linked to CRC only when produced by diffusely adherent E. coli, making the intimate contact with mucosa as important as the toxin production itself. Last but not least, case-control studies that revealed some epidemiological association should be interpreted with caution when it comes to any cause-effect considerations because the carcinogenic process itself can create an environment that may favour some microbes over others.

ANY WAY OUT?

Obviously, the early excitement and “overselling of the microbiome” should be replaced by painstaking further work. Our recommendations can be summarised as follows:

First, any studies should be prepared carefully, taking into account recommendations and limitations of particular techniques published previously. Further studies with underpowered cohort sizes, using low quality sequence databases, or lacking sufficient depth of sequencing should be avoided, because they can hardly enrich our knowledge in the field. Long-term prospective studies should be preferred over baseline studies whenever feasible.

Second, validation of CRC screening markers on particular populations should be encouraged, because of differences in the gut microbiome observed in different geographical locations or different racial/ethnic groups. Such targeting on more homogeneous populations should also reduce the potential bias coming from different dietary habits.

Third, whenever possible, conventional culture should be used to confirm the findings from sequencing studies. Particularly, candidate marker status of species or genes revealed by molecular techniques should be confirmed or refuted by culture and vice versa. Sequence databases should be improved by whole-genome sequencing of species of interest, rather than by producing more and more metagenomic data of limited information value.

Fourth, the time is ripe for the synthesis of available knowledge. Obviously, there is no single microbial species or toxin that would facilitate development of all sporadic CRC cases. Rather, different species or genotoxins may do so in different individuals and populations. Then, screening techniques of CRC risk should follow all known candidate markers, combining particular species, genotoxin production and possibly further strain characteristics whenever relevant. When recruiting patients for any type of study, as many candidate markers as possible should be evaluated and detection techniques should ideally be standardised across research groups to ensure the possibility of pooling the data in future, because limited cohort sizes are critical in most studies.

Fifth, not only culture, but systematic high-throughput culturomics should be developed further and introduced widely because cultivation will still represent an approach that is economical and easy to perform, and, therefore, best suitable for long-term studies to establish really large patient cohorts. Such cohorts are critical to detect potential CRC promoting species or factors of low abundance. We have cultured Shewanella putrefaciens in rectal swabs of 4 out of 67 patients with newly diagnosed CRC compared to no culture in 67 controls during a 2 year ongoing study (unpublished data). Thus, number of examined cases needs to be expanded to hundreds to confirm or refute a candidate CRC marker status of this species. Since S. putrefaciens is easy to grow and identify using MALDI-TOF MS, conventional culture is clearly superior to sequencing or PCR-detection for this task at the moment.
Last but not least, the potential of practicable detection should always be considered too when focusing on microbial markers of sporadic CRC. Of course, research studies should carefully delineate the difference between microbes found in the tumour mass, those adherent to the tumour mucosa compared to adjacent healthy mucosa, etc. However, when it comes to studies looking for markers useful for population screening, one should bear in mind that controlled sampling from colonic mucosa would require colonoscopy, which is challenging to the patient, costly, low-throughput and is itself the ultimate tool for CRC screening. Then, to reach wide acceptance of population screening, microbial markers should best be detected in a much more comfortable sample like a rectal swab or a stool sample. Although there are differences in composition of luminal and mucosal microbial flora, a newer study also showed that in luminal samples, distinctive differences can be found between control and CRC cases as well. Any positive result of easily performed microbial screening can thus not only indicate increased CRC risk, but should also increase the patients’ adherence to colonoscopic examination, similar to FOBT – in the Czech Republic, 36 086 primary screening colonoscopies were performed in the period 2006-2015 compared to 154 996 colonoscopies following a FOBT-positive testing.

ANY TO-DATE PATIENT BENEFITS?

Regardless of the plethora of open questions and need for further research, it should be possible – in our opinion – to use the available knowledge to assist patients now. Of course, ethical issues and standards of evidence-based medicine should be kept in mind. However, if S. galalolyticus bacteriemia has been repeatedly demonstrated to be associated with the presence of tumour mass in colon, appropriate examination should follow detection of this species in blood culture. In such case, the patient is not really interested whether the bacterium does cause CRC or just facilitates pre-malignant lesion progression to CRC (ref.55). Similarly, appendicitis and appendectomy have been linked to increased risk of CRC and female genital organ cancer in a relatively short period after surgery, indicating appendicitis to be an early marker of distant tumour proliferation. Although these data come from an Asian population, it would be reasonable to offer primary screening colonoscopy to a patient who has undergone appendectomy. The same may be true for pyogenic liver abscess caused by K. pneumoniae. In this case, a Southeast Asia-specific hypervirulent pks genomic island-harbouring K. pneumoniae clone should be responsible for this link. Unfortunately, this clone appears to have spread to Europe and undergone further evolution. Then, a K. pneumoniae-driven community-acquired liver abscess should also alert us to possible CRC risk. Alongside to the association known from clinical studies, there is an ongoing effort to validate promising markers for screening purposes. A Japanese study demonstrated high positive predictive power for CRC of high-copy number of F. nucleatum DNA detected by species-specific PCR (ref.61) and a Chinese group reported increased levels of anti-F. nucleatum IgA antibodies in serum of CRC patients.

CONCLUSION

Although excitement over the tsunami of early achievements in microbiome research has been dampened by recent criticism, the predictive potential of microbial markers of increased CRC risk cannot be questioned. Currently, more balanced approaches to the field are taking stage, represented mainly by complementary action of sequencing and culture, including evolving techniques of high-throughput culturomics. Furthermore, available knowledge also encourages clinical validation of candidate markers of increased CRC risk, which include both culturable bacteria and bacteria that cannot be cultured easily but can be detected by PCR conveniently. The same is true for bacterial genotoxins or other detectable substances. It should however be stressed that trustworthy conclusions useful for population-wide screening can be reached only when validation studies are verified independently on the same well defined population, representing a challenging but meaningful task for public health authorities and funding agencies.

Search strategy and selection criteria

Data for this article were identified by searches of PubMed using the terms “colorectal cancer”, “gut microbiome”, “culturomics”, “screening” and combinations of these terms, and by following references from relevant articles. We gave preference to publications presenting larger cohorts and using sound methodology. Citations from respectable journals were given special weight. Our own experience was also included.

Acknowledgements: This work was supported by the Faculty of Medicine & Dentistry, Palacky University Olomouc (IGA_LF_2017_031), by RVO (61989592), by MH CZ – DRO (FNOl, 00098952) and by the Ministry of Education, Youth and Sports of the Czech Republic (LO1304).

Author contributions: SL: incited the review, prepared and summarised the clinical related parts of the manuscript, discussed recommendations; KM: searched the literature for CRC-related toxin action and detection; PS and PZ: prepared the clinical related parts of the manuscript, discussed recommendations; VR: searched the literature for CRC-related species associations and draw final conclusions.

Conflict of interest statement: The authors state that there are no conflicts of interest regarding the publication of this article.
REFERENCES


8. Wu S, Morin PJ, Maouyo D, Sears CL. Bacteroides fragilis en-...


